

Predict Customer Churning

Lavinia Wang

Agenda



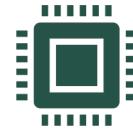
Introduction



Dataset
Overview



Exploratory
Data Analysis



Preprocessing

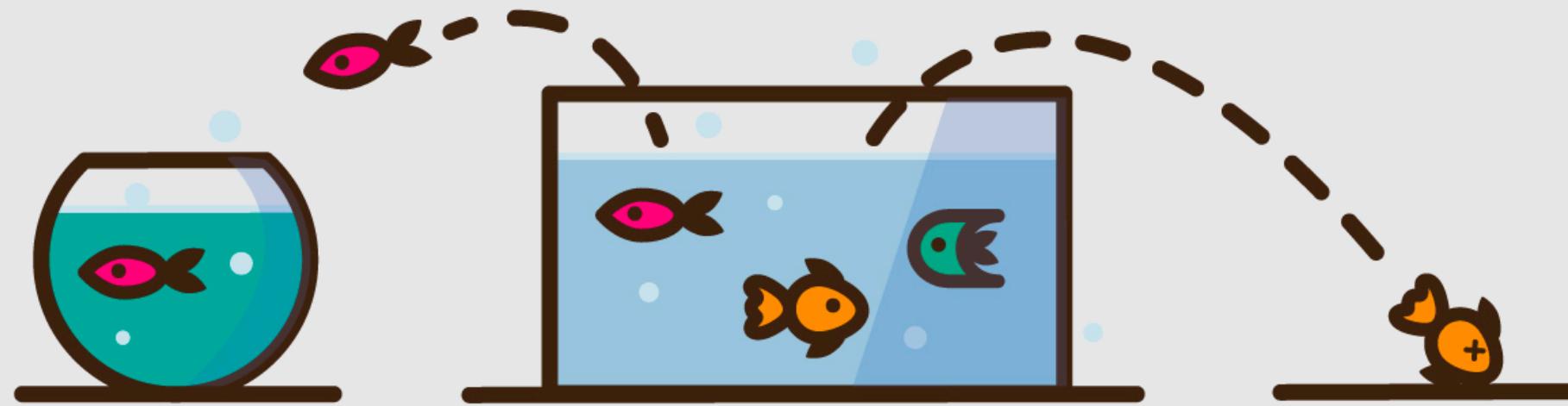


Model Output



Conclusion

CUSTOMER CHURN



Data Overview



10000
Customer
Records



14
Features



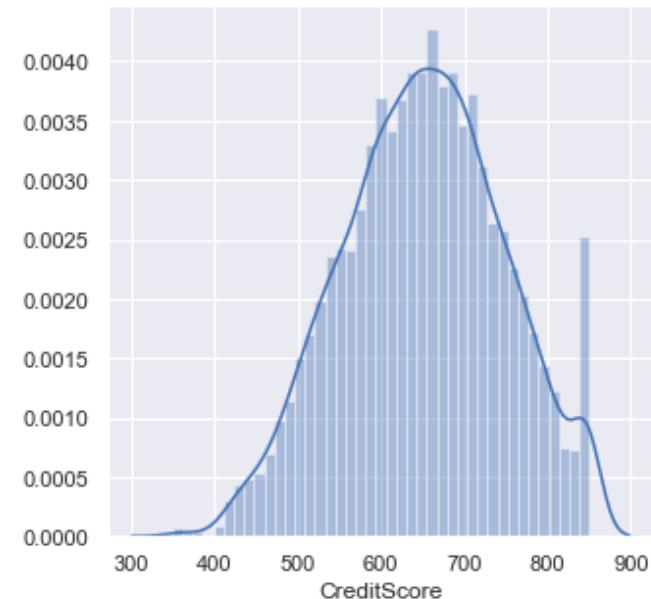
4
Redundant



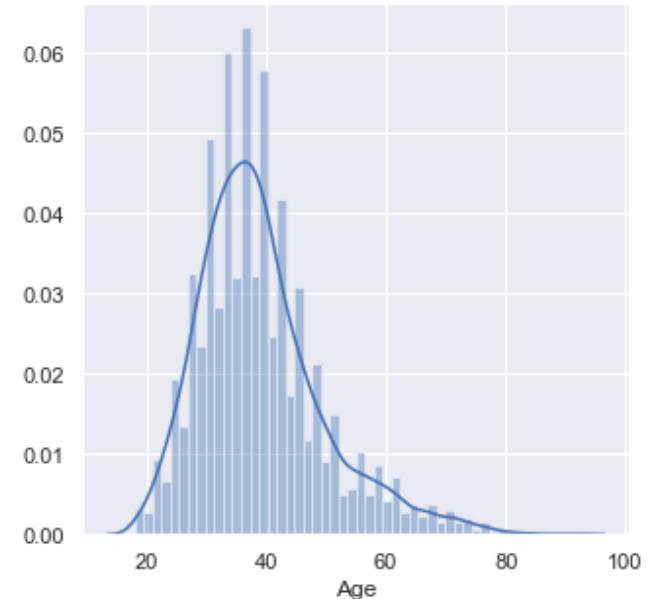
10
Informative

Exploratory Data Analysis(EDA)

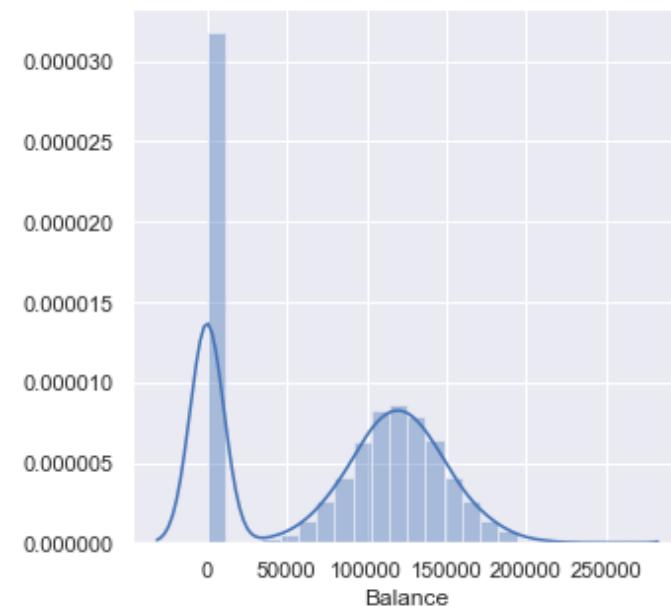
Credit Score



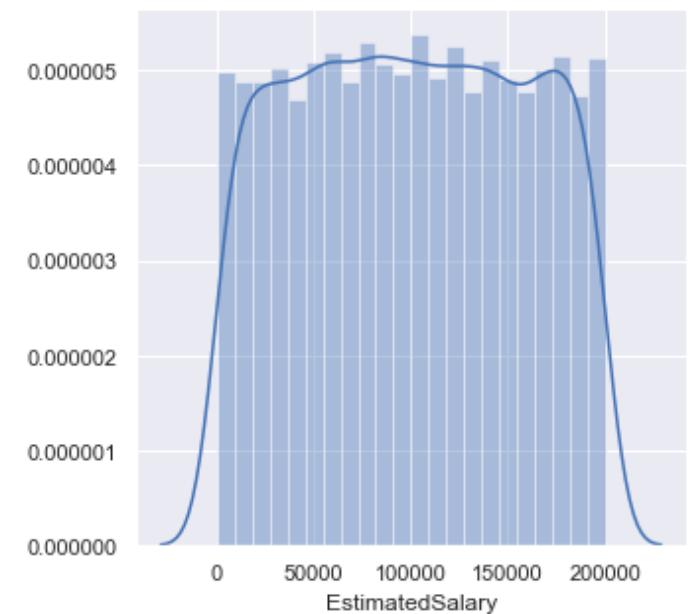
Age



Balance

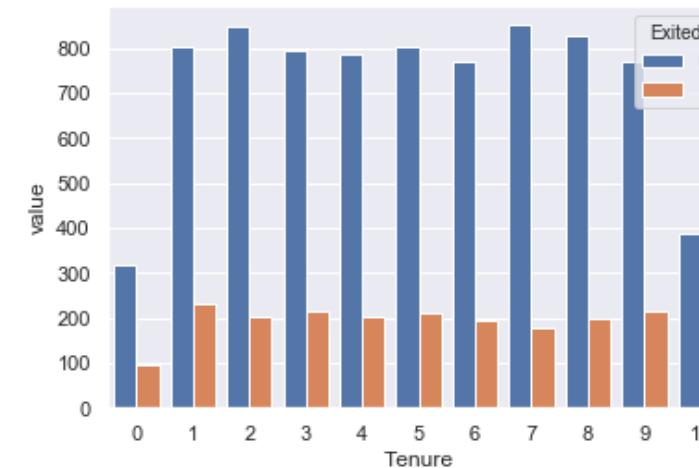


Salary

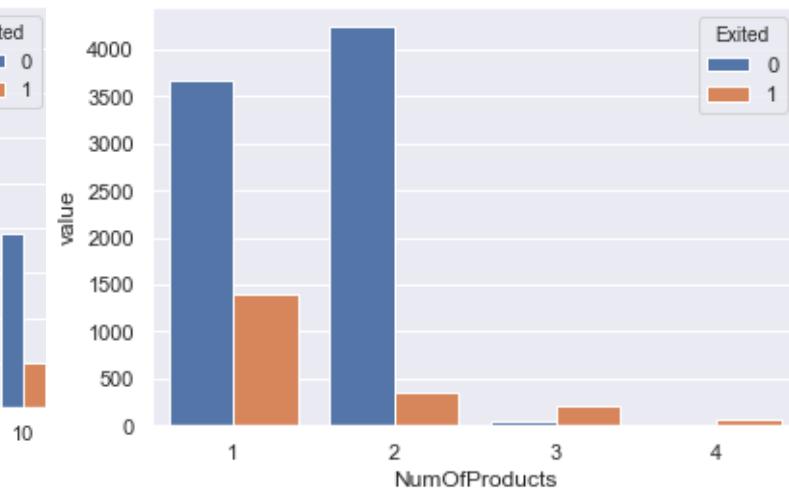


Exploratory Data Analysis(EDA)

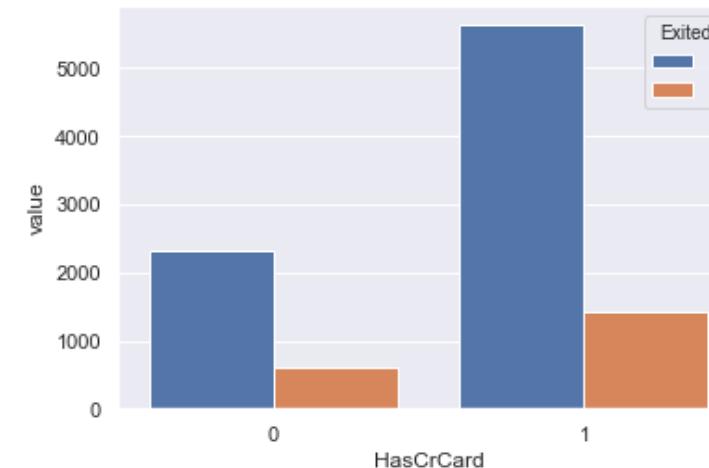
Tenure



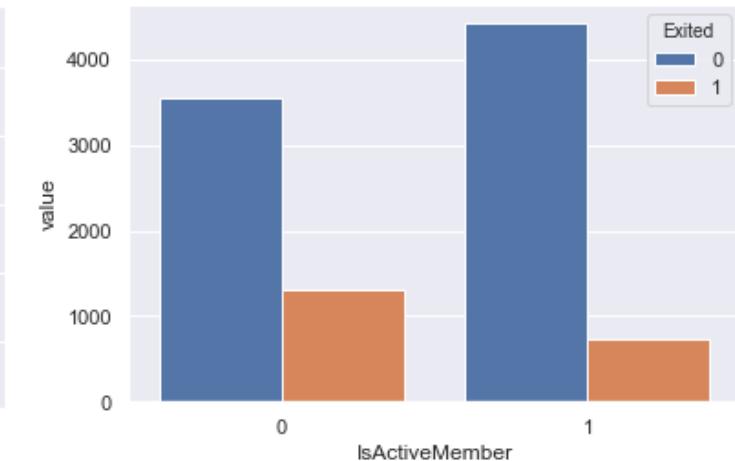
Number of Products



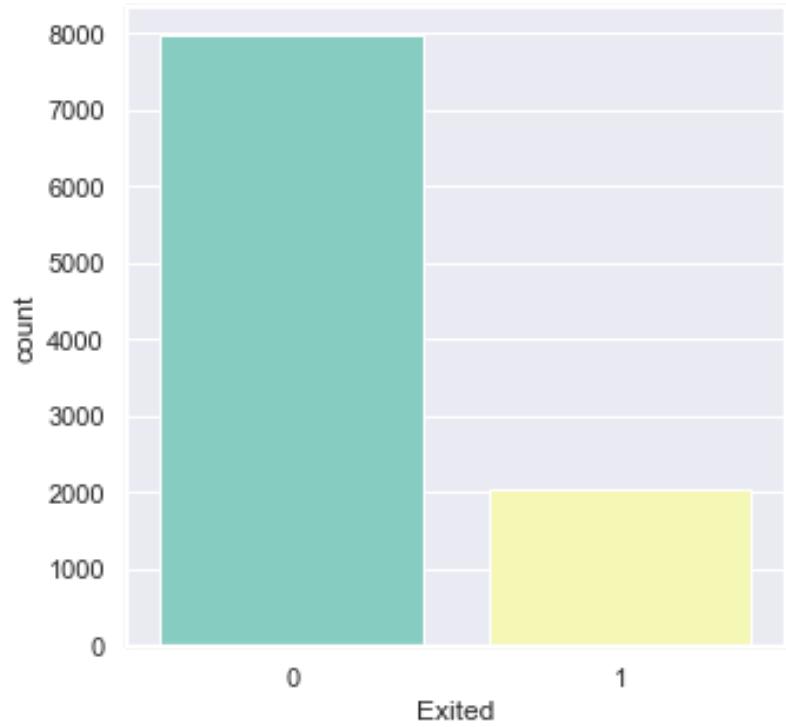
Has Credit Card



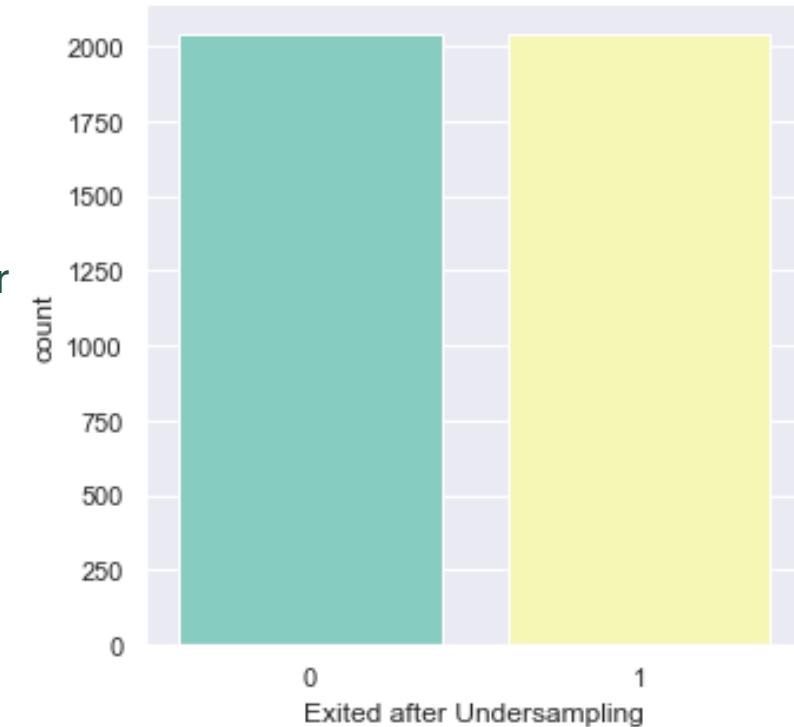
Is Active Member



Preprocessing



Random Under Sampler



Classification Model

Logistic
Regression

Support
Vector
Machine
(SVM)

Naive Bayes

Decision Tree

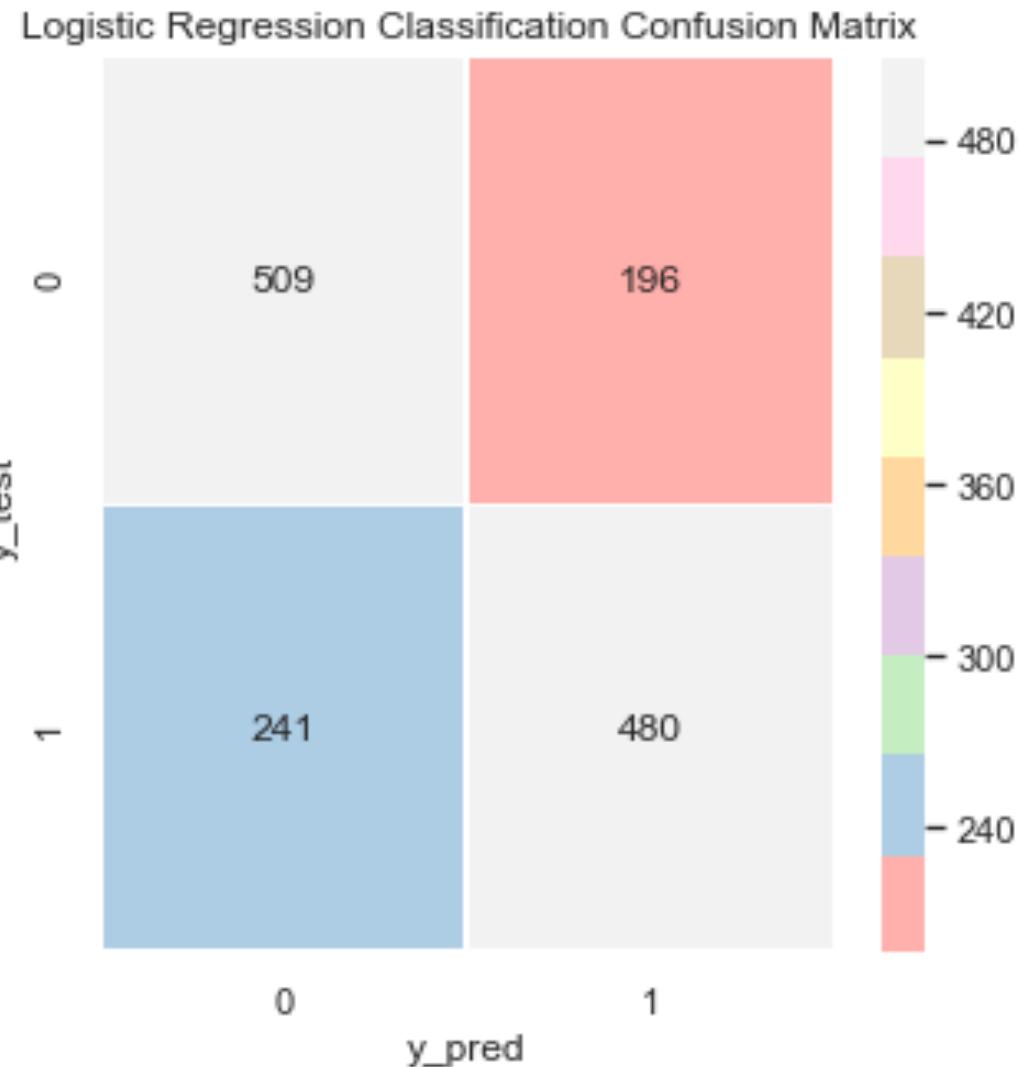
Random
Forest

Gradient
Boost

Ada Boost

Logistic Regression

Acc: 0.69 AUC: 0.77



SVM

Linear

Acc: 0.71 AUC: 0.77

Poly

Acc: 0.74 AUC: 0.82

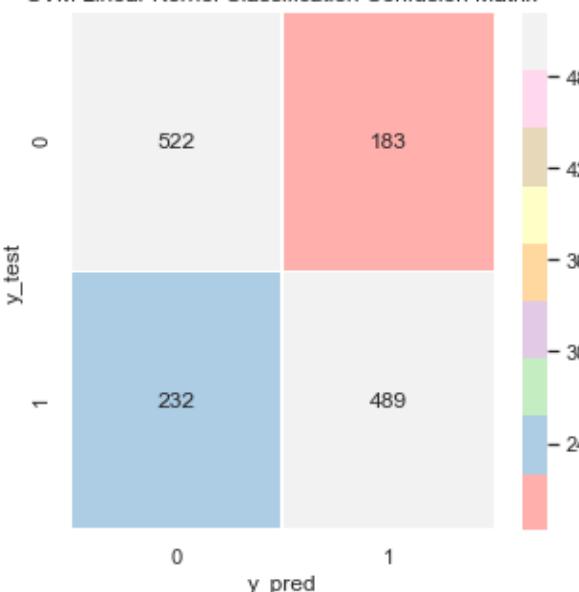
RBF

Acc: 0.73 AUC: 0.81

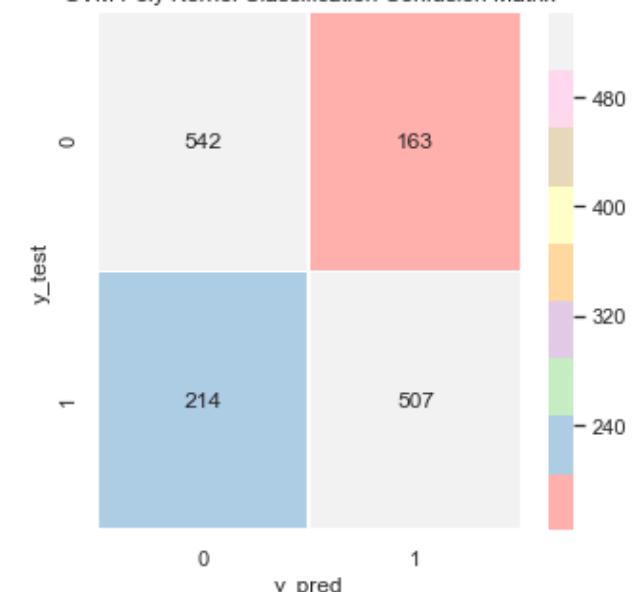
Sigmoid

Acc: 0.54 AUC: 0.55

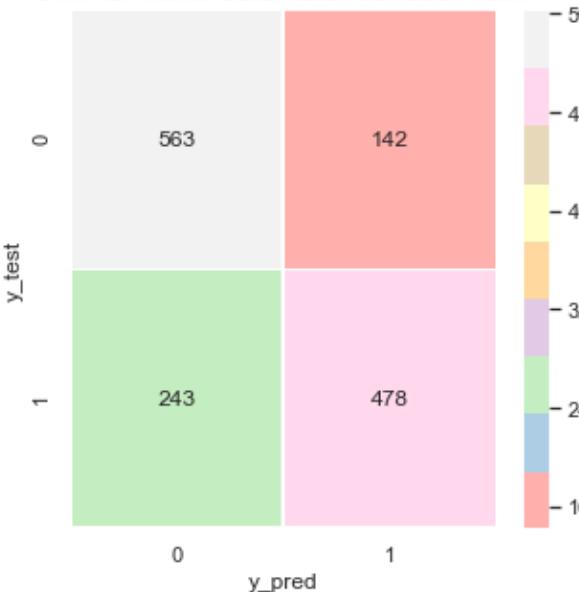
SVM Linear Kernel Classification Confusion Matrix



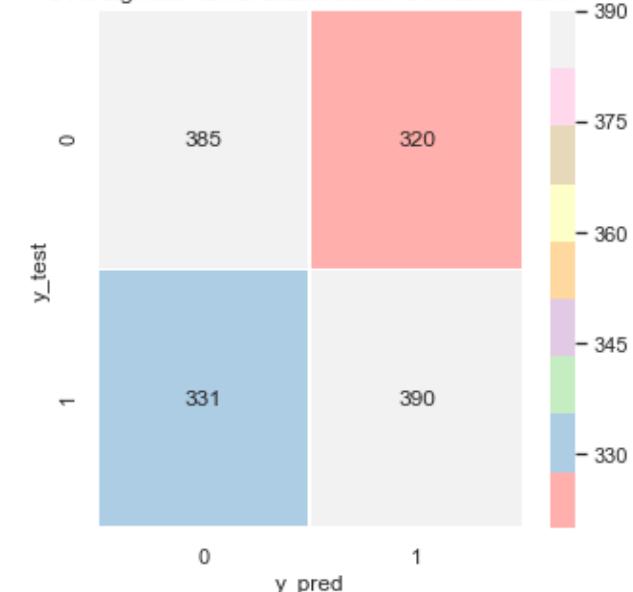
SVM Poly Kernel Classification Confusion Matrix



SVM RBF Kernel Classification Confusion Matrix

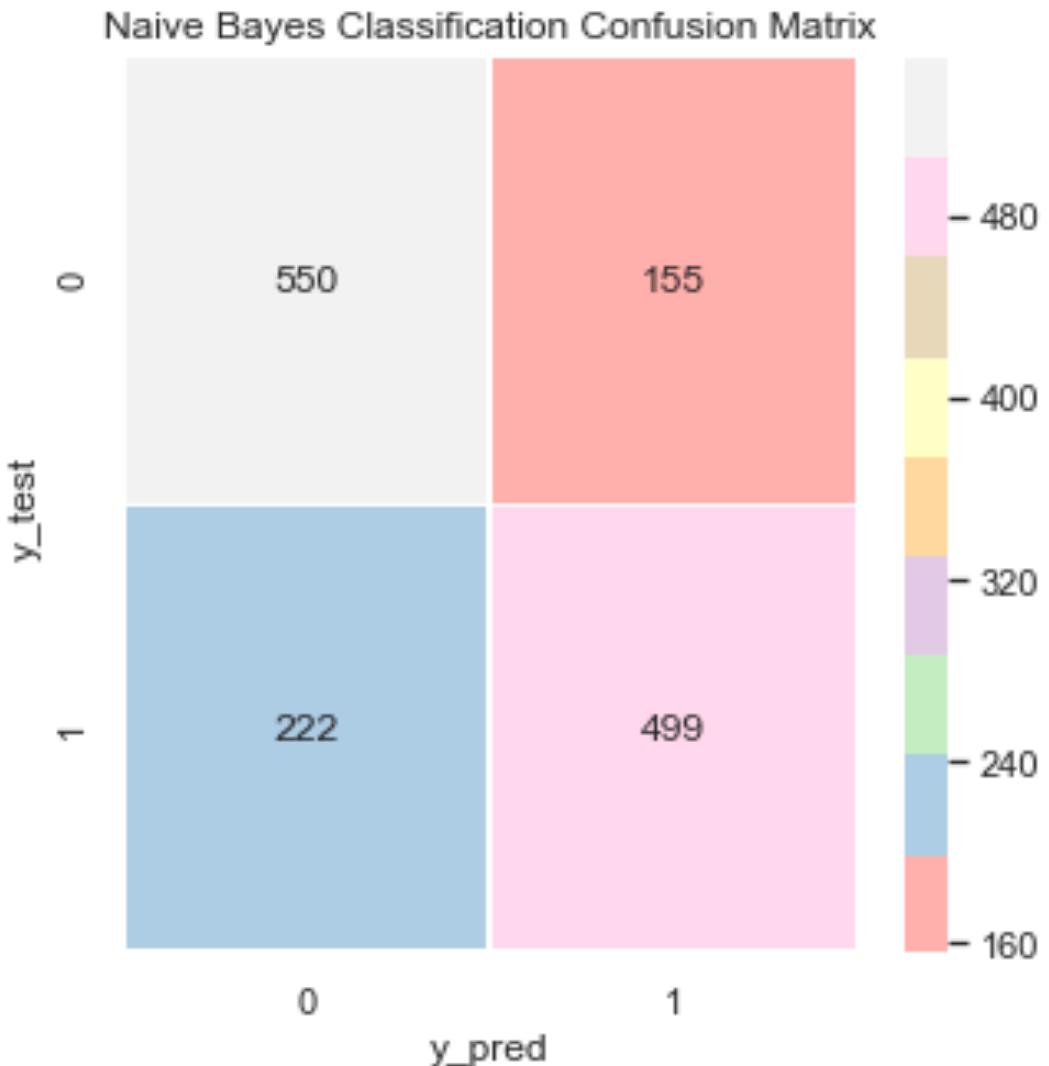


SVM sigmoid Kernel Classification Confusion Matrix



Naïve Bayes

Acc: 0.74 AUC: 0.81



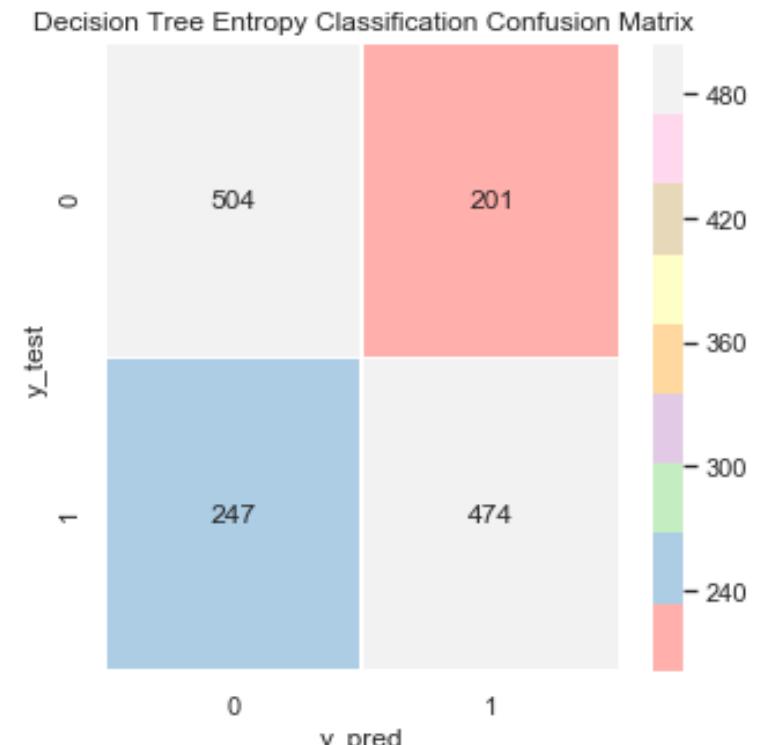
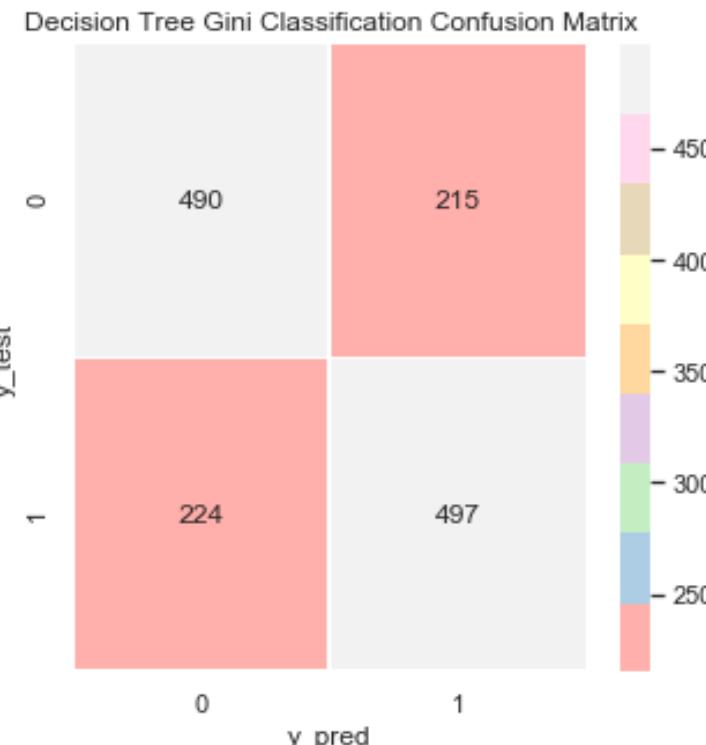
Decision Tree

Gini

Acc: 0.69 AUC: 0.69

Entropy

Acc: 0.69 AUC: 0.69



Random Forest

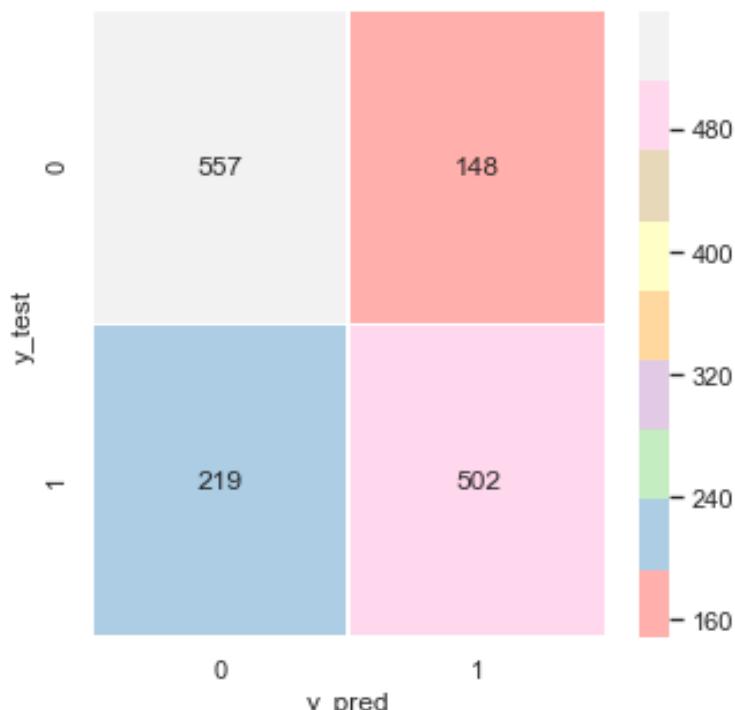
Gini

Acc: 0.74 AUC: 0.83

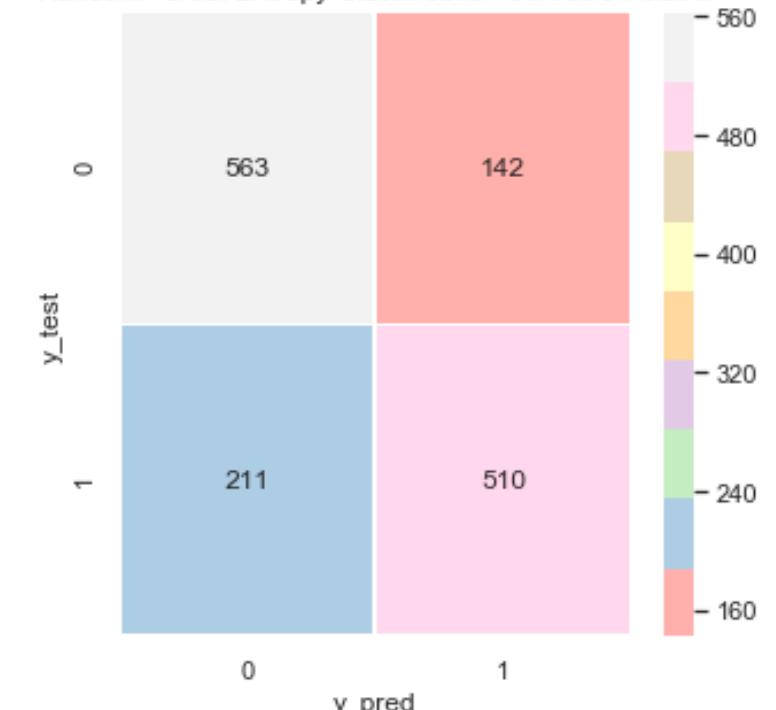
Entropy

Acc: 0.75 AUC: 0.83

Random Forest Gini Classification Confusion Matrix

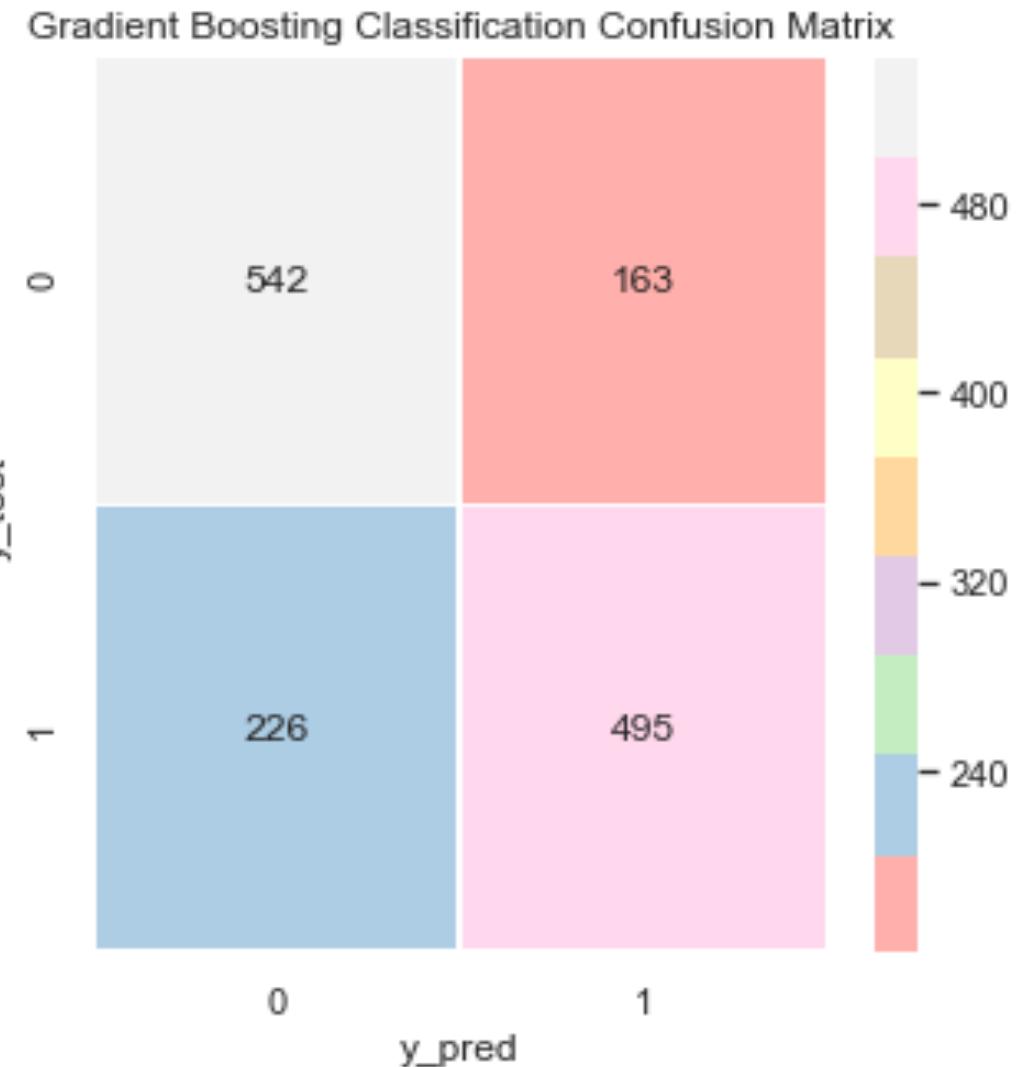


Random Forest Entropy Classification Confusion Matrix



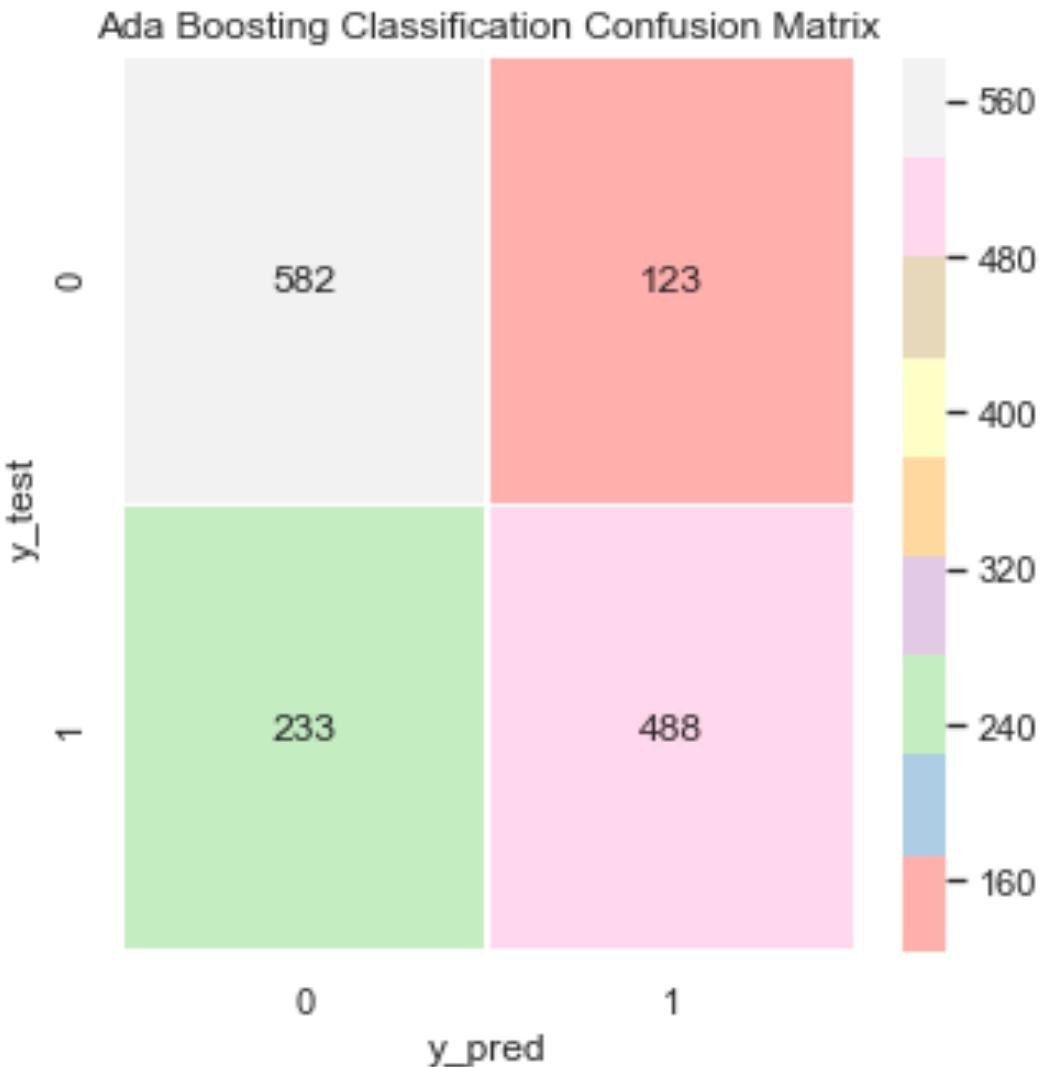
Gradient Boost

Acc: 0.73 AUC: 0.81

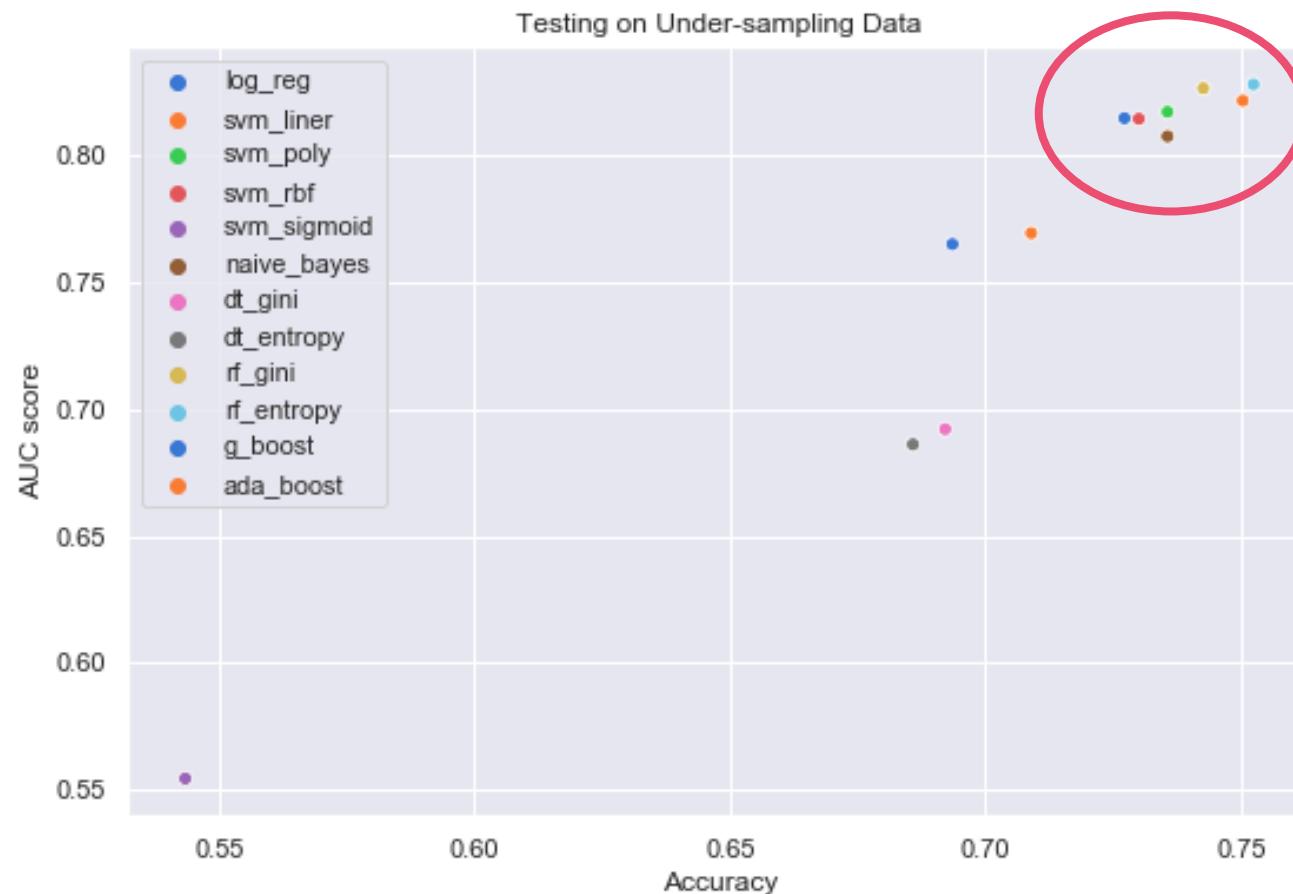


Ada Boost

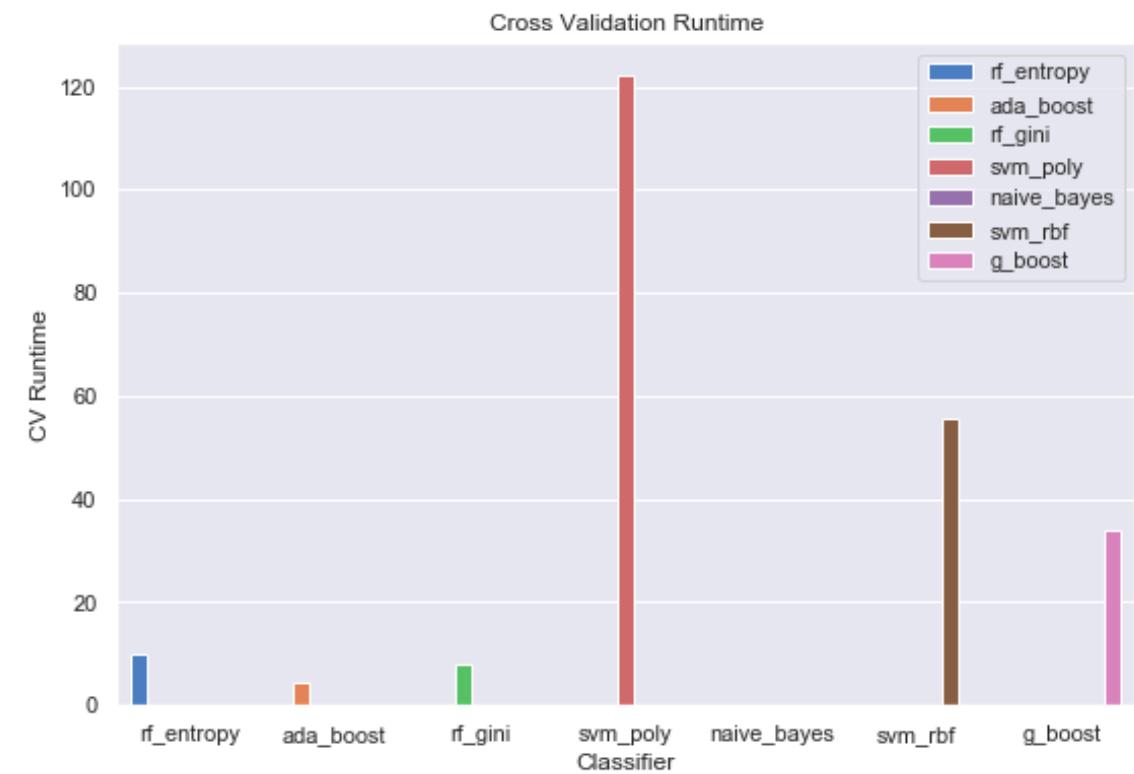
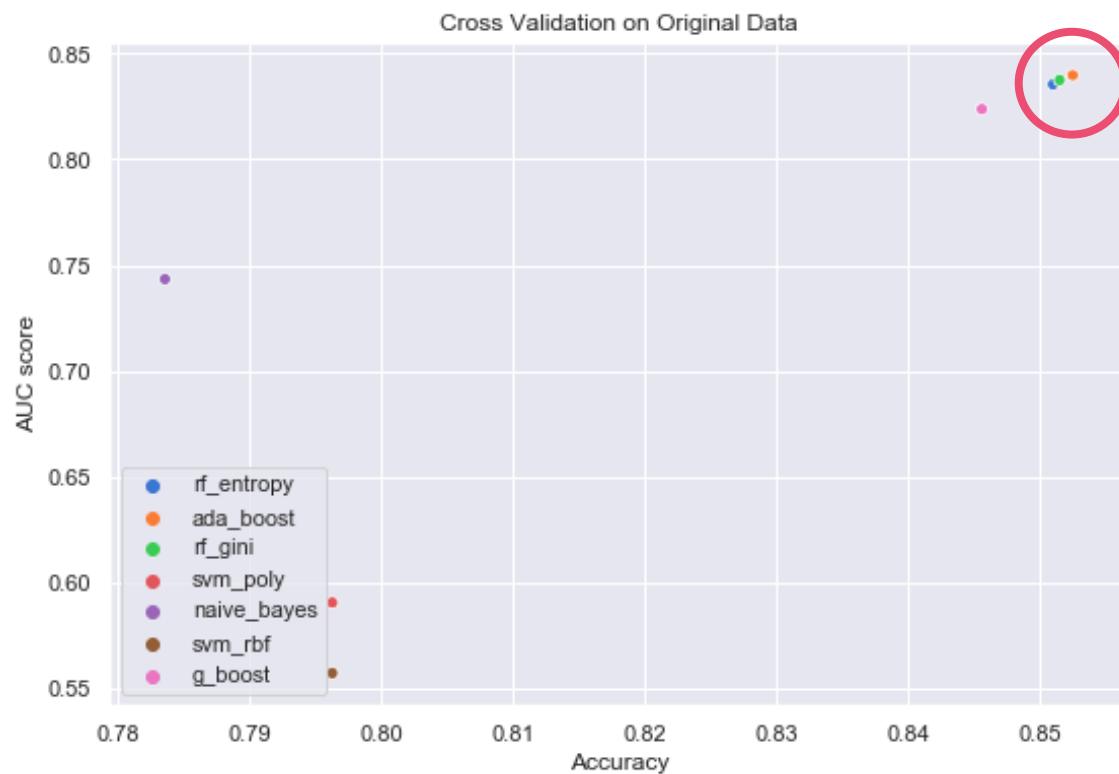
Acc: 0.75 AUC: 0.82

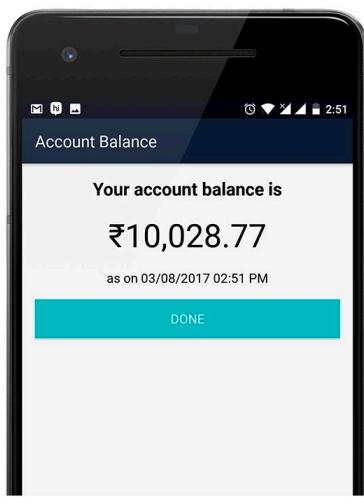


Comparison among Algorithms



Comparison among Algorithms





Feature Selection

Four shared features:

- ❖ Estimated Salary
- ❖ Age
- ❖ Credit Score
- ❖ Balance



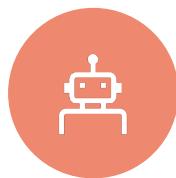
Random Forest Entropy

Conclusion

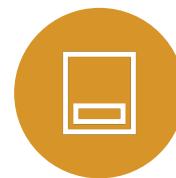
Future Work



Parameter
Tuning



Relative
Features



Rebalance
Methods