# Credit Card Defaults

Derek VanBriesen

# The Business Case

A bank is seeking assistance predicting credit card defaults.

They provide some of their customers' data.

They would like a model that is more accurate than the "baseline model."

# Questions to answer

- What is the baseline model and how accurate is it?
- What specific factors are strong predictors?
- What type of model will be used?

# What Data is Available?

## Non-financial

- Age
- Sex
- Marital Status
- Education

## Bills/Payments

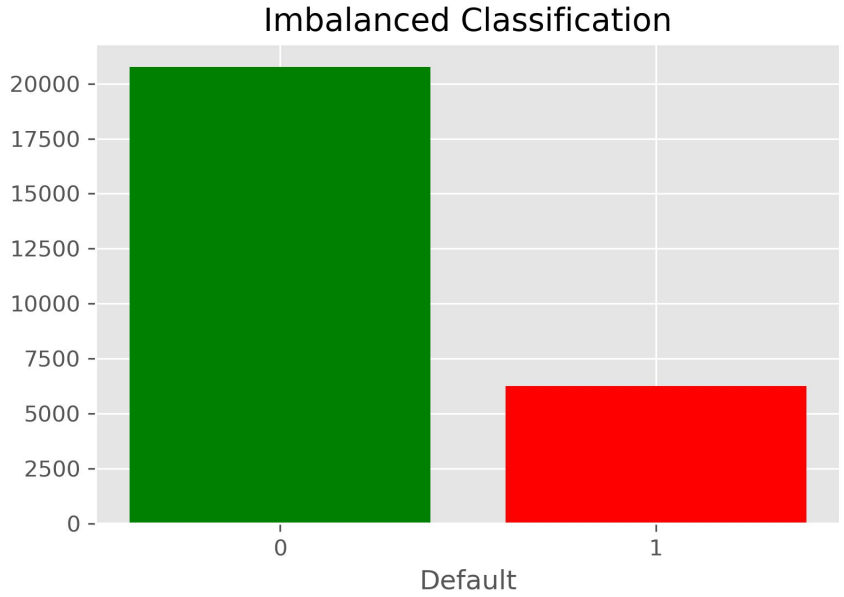
- Six months of bills
- Six months of payment

## Bank “Status”

- Six months of payment status
- Default

**Baseline Model:**
The most basic solution to a problem.  How accurate here?

# 76.8%
# Due to the target being imbalanced.



# Strongest Predictors?

- Most recent payment status
- ~3x stronger than 2nd place
- 2nd place is the 2nd most recent payment status

Bank “Status”

- Six months of payment status
- Default

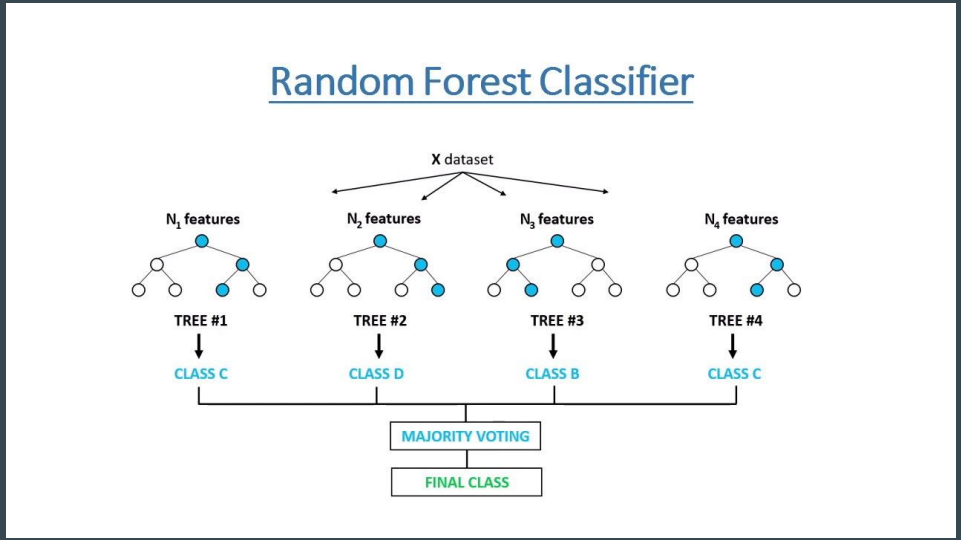
# Model Testing

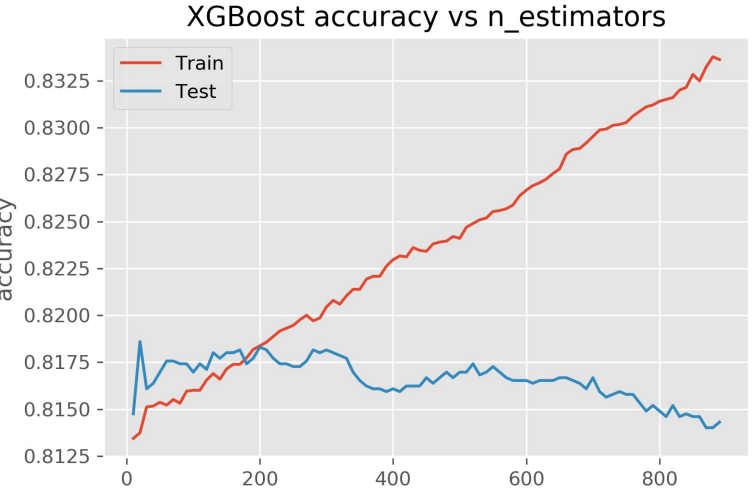
Tested data on basic version of most common models

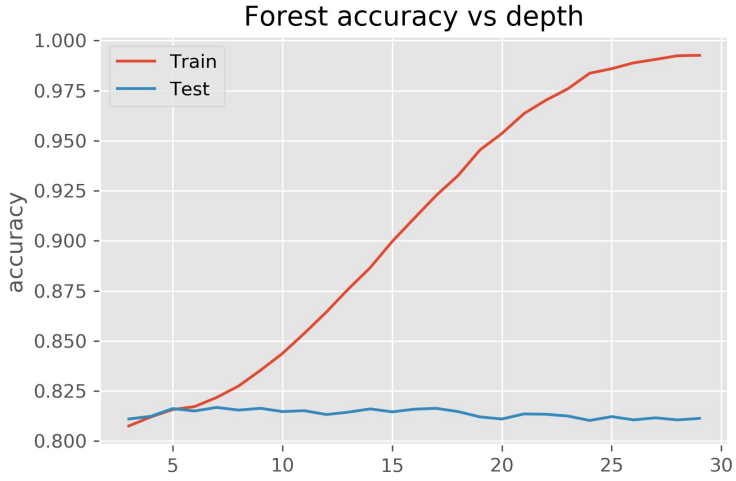
**Two winners :**

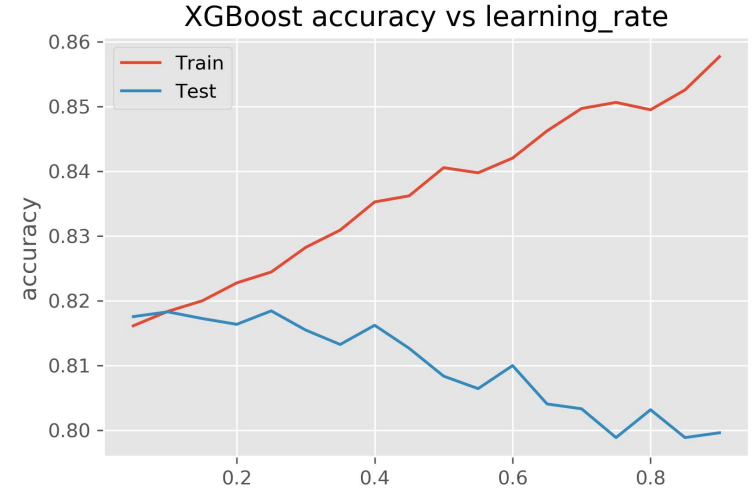
- XGBoost
- RandomForest

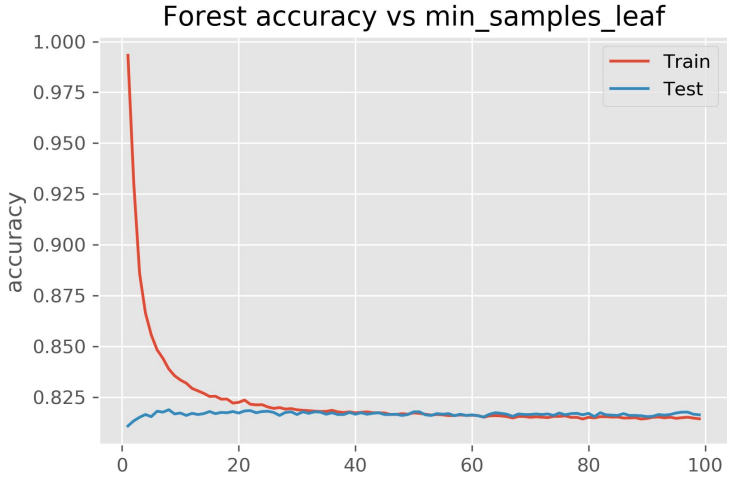
dmlc XGBoost




XGBoost accuracy vs n_estimators
Train
Test
accuracy
0.8325
0.8300
0.8275
0.8250
0.8225
0.8200
0.8175
0.8150
0.8125
0
200
400
600
800


Forest accuracy vs depth
Train
Test
accuracy
1.000
0.975
0.950
0.925
0.900
0.875
0.850
0.825
0.800
5
10
15
20
25
30


XGBoost accuracy vs learning_rate
Train
Test
accuracy
0.86
0.85
0.84
0.83
0.82
0.81
0.80
0.2
0.4
0.6
0.8


Forest accuracy vs min_samples_leaf
Train
Test
accuracy
1.000
0.975
0.950
0.925
0.900
0.875
0.850
0.825
0
20
40
60
80
100

# Final Results

| RandomForest | XGBoost |
| --- | --- |
| 81.59% Accuracy | 81.68% Accuracy |

# Future Work

More time to let models run would likely provide more accurate results

More models could be explored besides the two I went in depth with

I would have liked to do more feature engineering

# Thank you for listening!