# Predicting Credit Card Default

Team Mulberry

# Problem Statement

Can we predict when people will default on their credit cards?

This is helpful to:

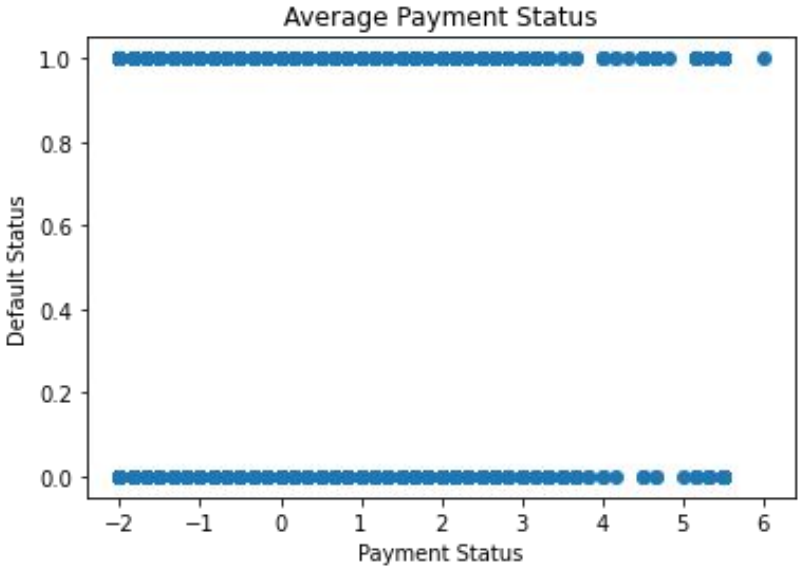
- Credit card companies
- Credit bureaus

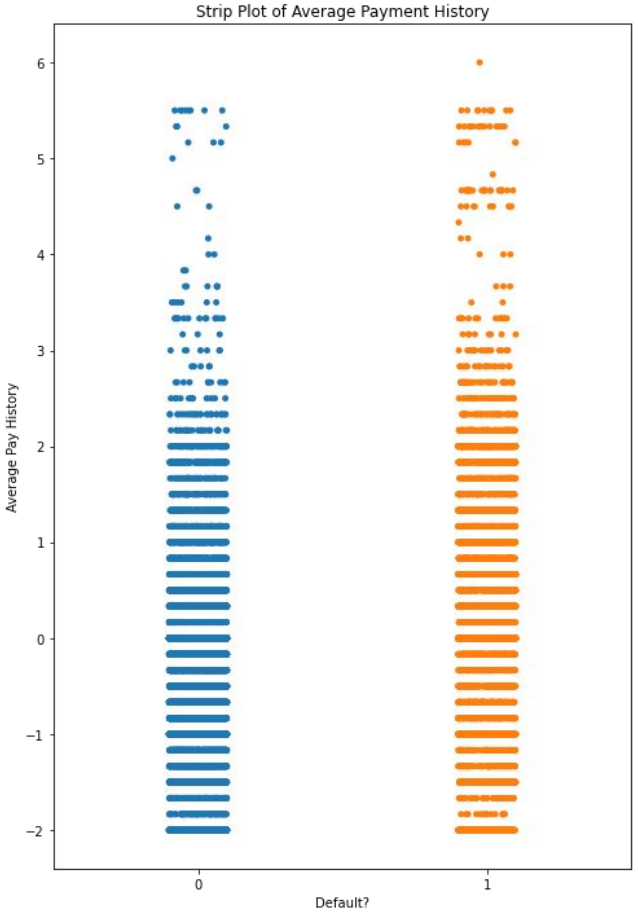
# The Data

- Categorical demographic data
  - Marital status
  - Education
  - Sex
  - Age
- Payment status history
- Payment amount history
- Bill amount history
- Credit limit

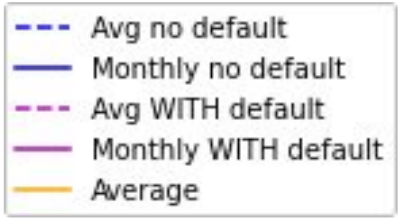
The data was unbalanced

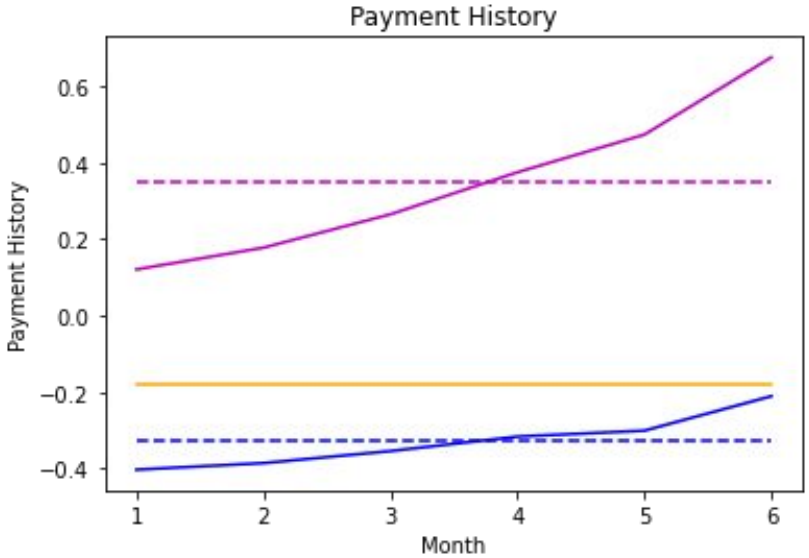
# The Data

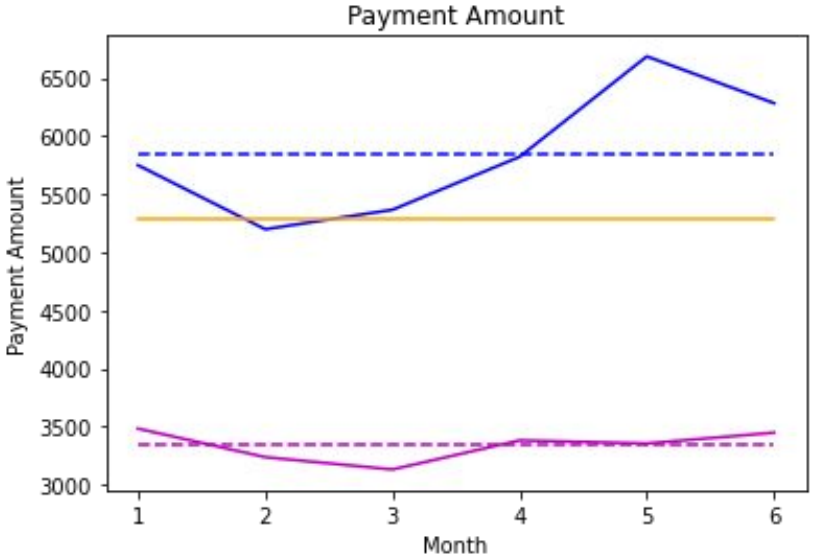

Average Payment Status
Default Status
Payment Status
1.0
0.8
0.6
0.4
0.2
0.0
−2
−1
0
1
2
3
4
5
6


Strip Plot of Average Payment History
Average Pay History
Default?
6
5
4
3
2
1
0
−1
−2
0
1

# The Data

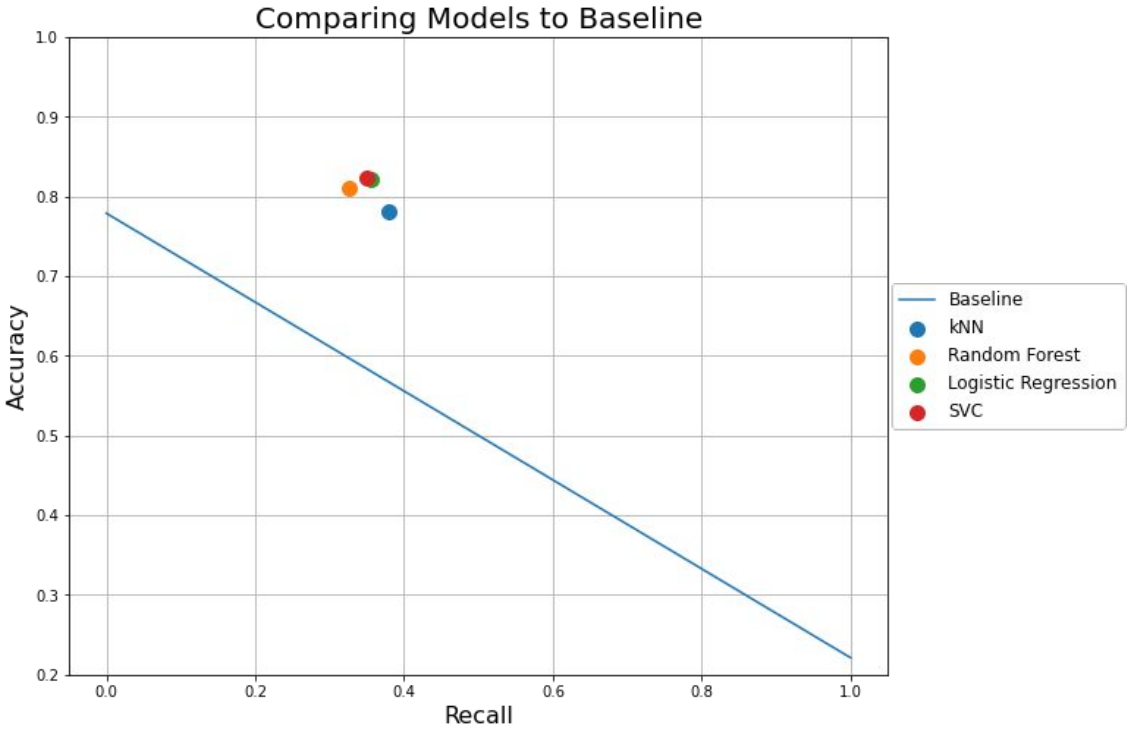

Avg no default
Monthly no default
Avg WITH default
Monthly WITH default
Average


Payment History
Payment History
0.6
0.4
0.2
0.0
−0.2
−0.4
1
2
3
4
5
6
Month


Payment Amount
Payment Amount
6500
6000
5500
5000
4500
4000
3500
3000
1
2
3
4
5
6
Month

# First Models

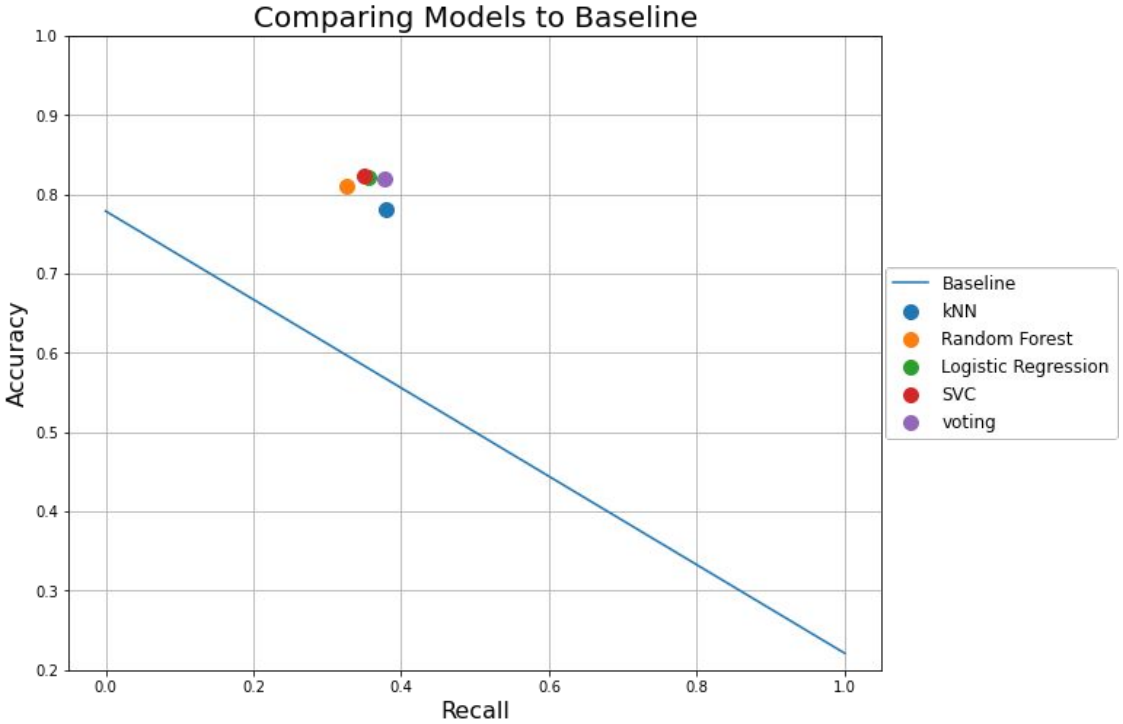
- Nearest Neighbors
- Random Forest
- Logistic Regression
  - Recent Payment Status
  - Average Payment Status
  - Total Unpaid Months
- Support Vector Classifier
  - Recent Payment Status
  - Average Payment Status
  - Total Unpaid Months



# Voting Model

- Different features for different models
- Break ties by assuming default

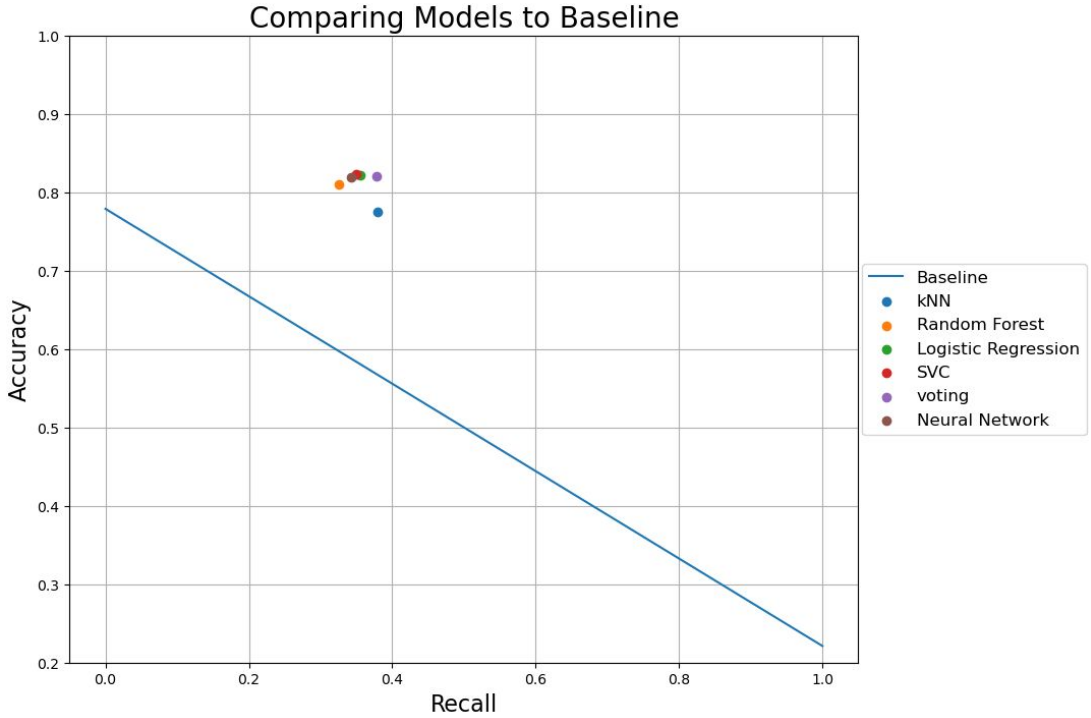
- Accuracy: 82.03%
- Recall: 37.80%



# Neural Network

- Multilayer Feed Forward
  - Recent Payment Status
  - Average Payment Status
  - Total Unpaid Months

- Accuracy: 81.90%
- Recall: 34.19%



# Final Model and Recommendations

- Final Model: Voting
  - Best accuracy and recall of all models
- Recommendations:
  - Could use model to make policy decisions
    - Send warnings to those more likely to default
    - Implement/change leniency policy
  - Could go with different models as well:
    - Logistic regression, SVC, Neural Network only used three features
    - Great for lowering data storage