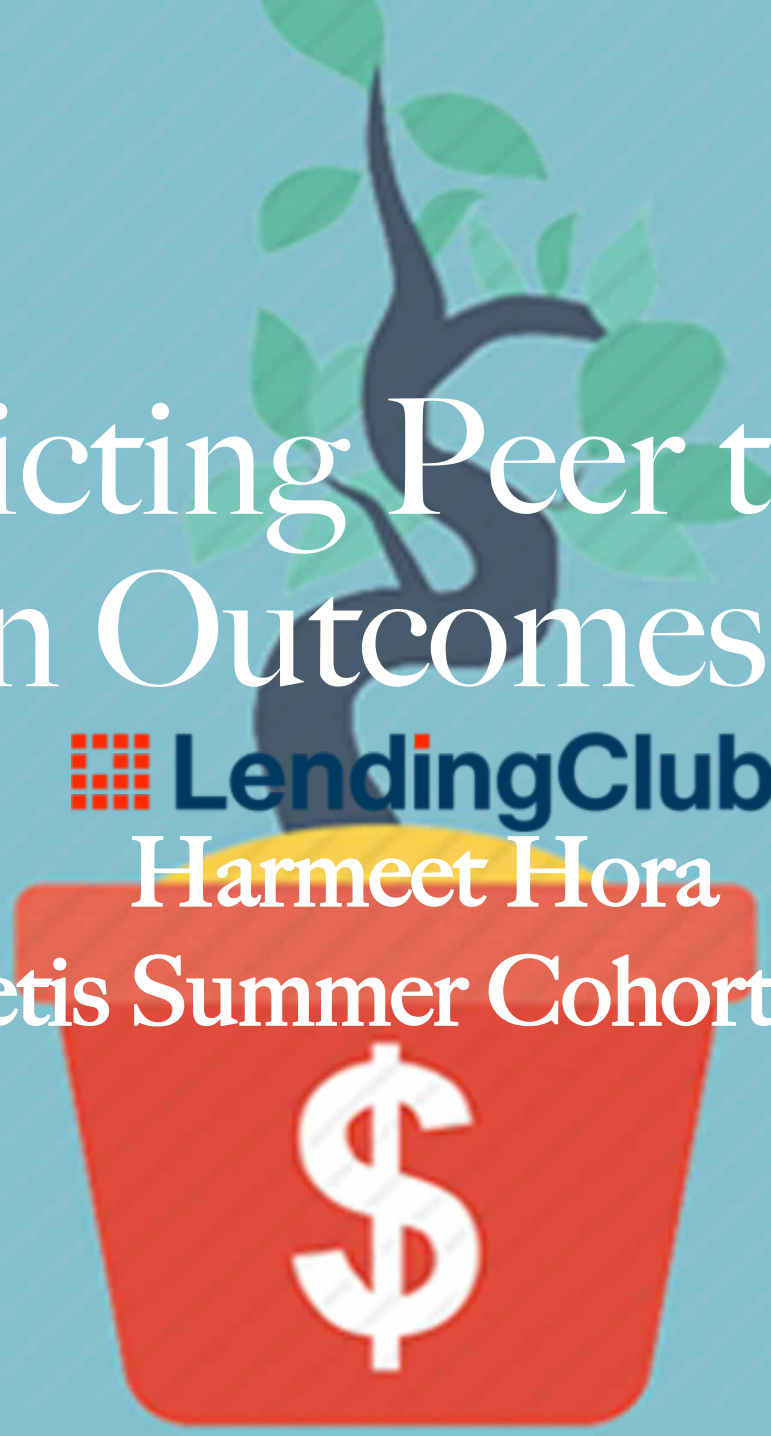
Predicting Peer to Peer Loan Outcomes With LendingClub

Harmeet Hora
Metis Summer Cohort 2018



The Data

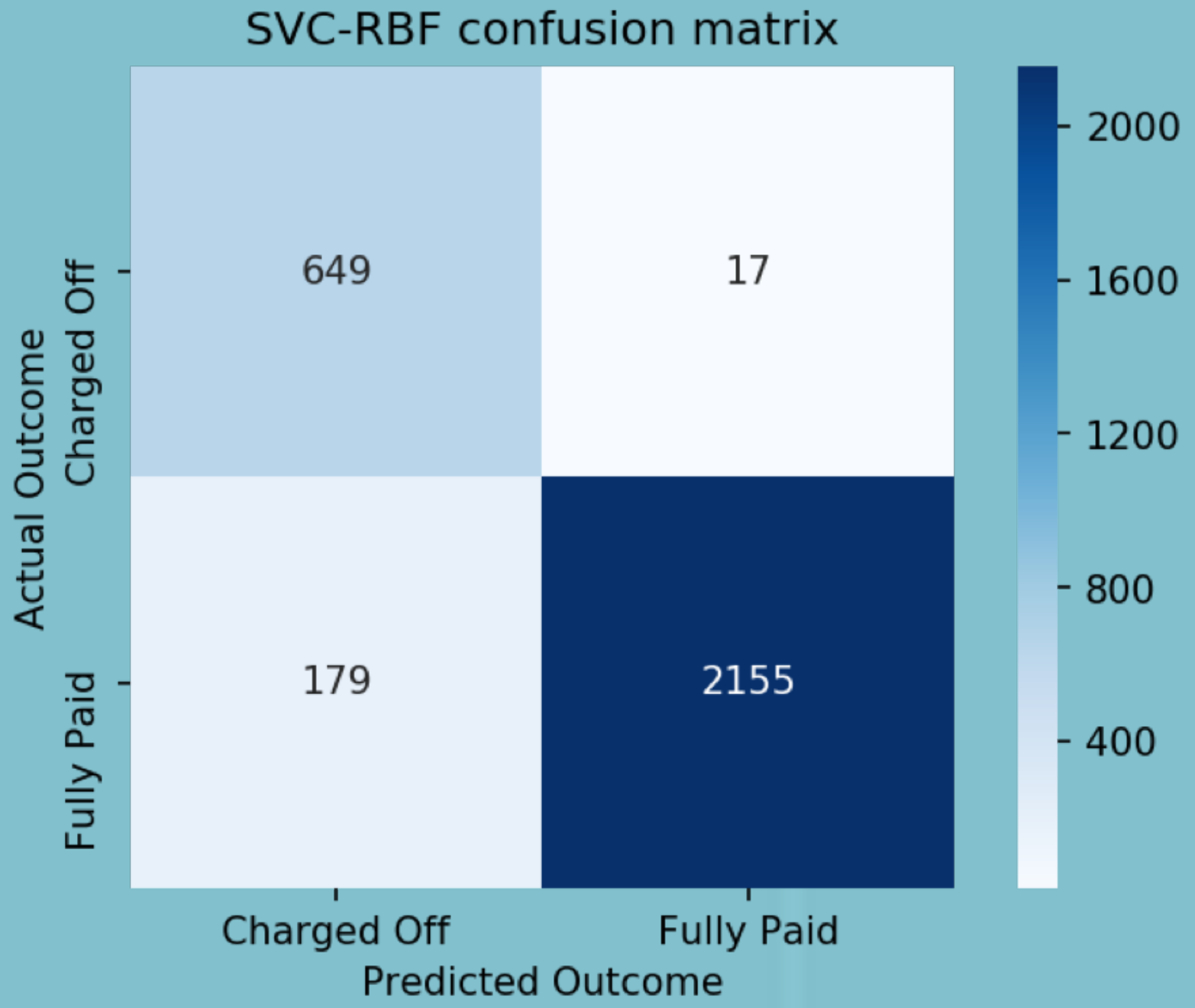
- LendingClub.com loan data for all loans accepted in 2015
- Assumed independent
- Binary outcome (“Full Paid” or “Charged Off”)
- 300k observations => 10k

Use Case

- Minimize Risk
- Focus on precision of “Paid Off”
- Profit!

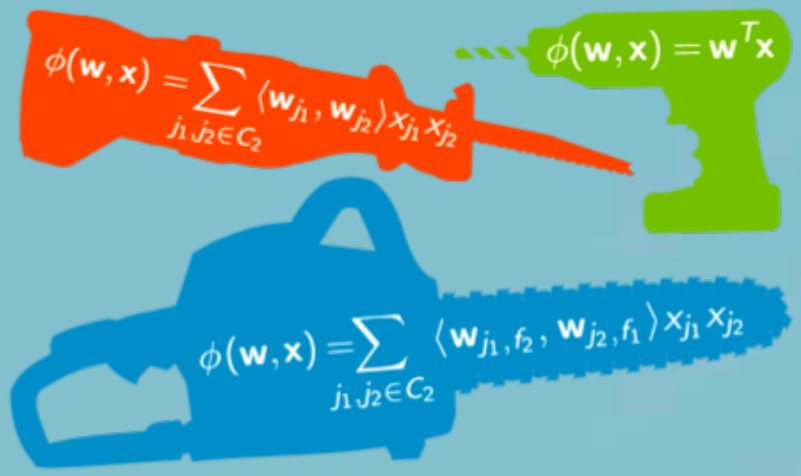
Results

Algorithm	Precision on Test Set
SVC	0.92
XGB Classifier	0.89
k-Nearest Neighbors	0.87
Decision Tree	0.87
Linear SVC	0.86
Random Forests	0.79
Dummy Classifier	0.77
Logistic Regression	0.69
Gaussian Naïve Bayes	0.65

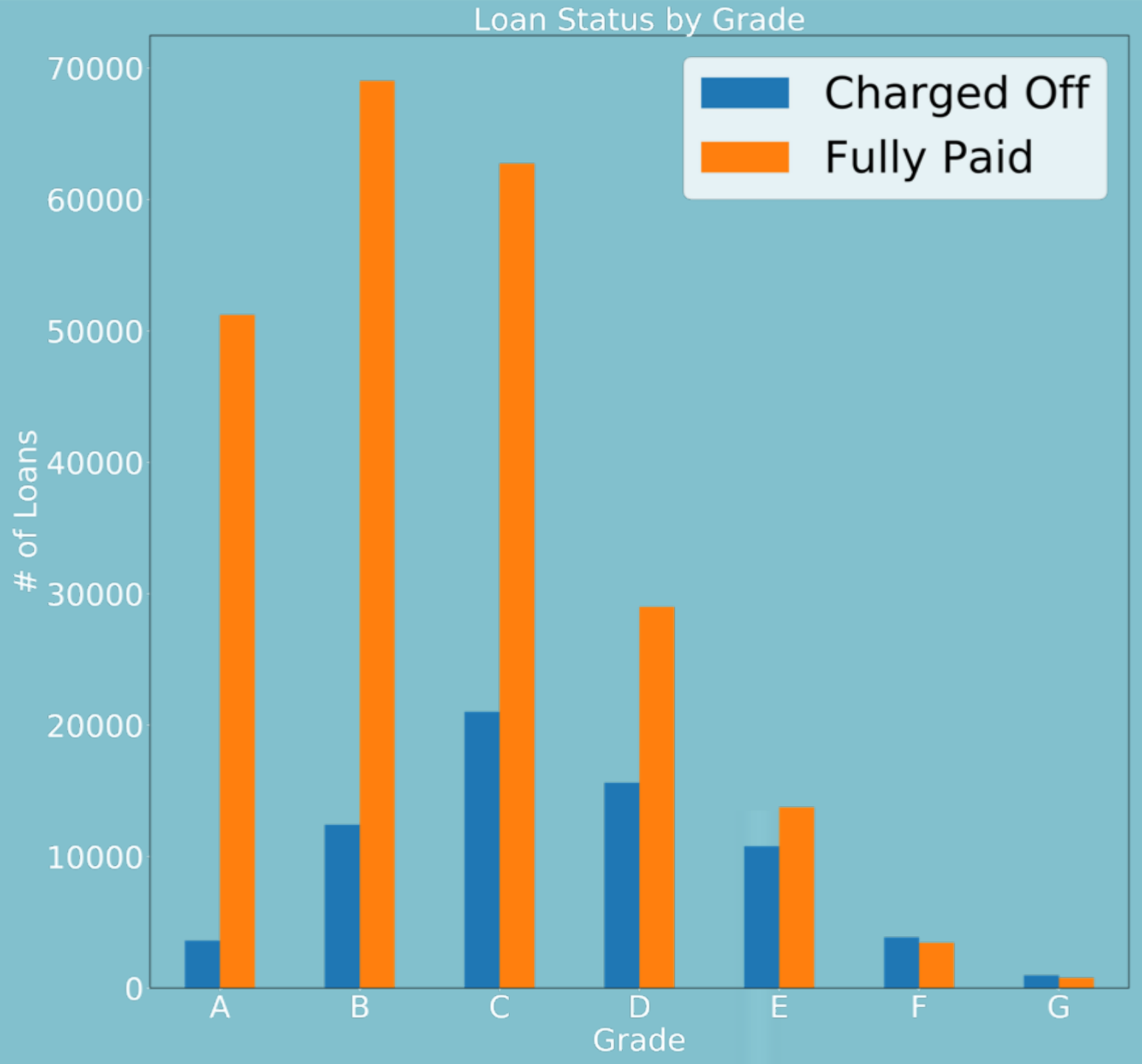

SVC-RBF confusion matrix
649
17
179
2155
Actual Outcome
Charged Off
Fully Paid
Charged Off
Fully Paid
Predicted Outcome
2000
1600
1200
800
400

Challenges

- Feature Engineering
  - Non-predictive features
  - >100 to 63
- Optimizing run times



Top 10 Features
Importance
0.04
0.03
0.02
0.01
0.00
dti
Mos Oldest Bank Instl
Avg Balance
Income
Mos Old Revolving
Loan Amt
Revolving Util Rate
Total Cur Balance
Total Open on Cards
Features


Loan Status by Grade
Charged Off
Fully Paid
70000
60000
50000
40000
30000
20000
10000
0
# of Loans
A
B
C
D
E
F
G
Grade

Next Steps

- Non-binary classifier (by grade, etc)
- Further reduced dimensionality
- Deploy in production

Thank you.