# Loan Default Prediction

Diane Quillao

# LendingClub

## How Lending Club Works
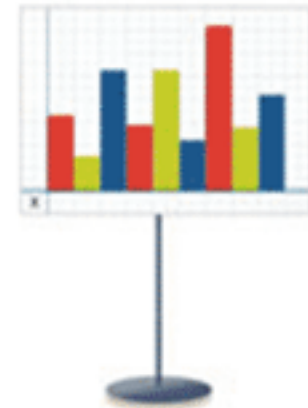
**Borrowers** apply for loans.
**Investors** open an account.

**Borrowers** get funded.
**Investors** build a portfolio.

**Borrowers** repay automatically.
**Investors** earn & reinvest.

# What is a bad loan?

- **Default**: payment is120 days past due

- **Charged Off**: payment 150 days past due

- From 2007-2017, Charged Off Rate: 7.08%

  - 2014: 10.15%

  - 2015: 10.09%
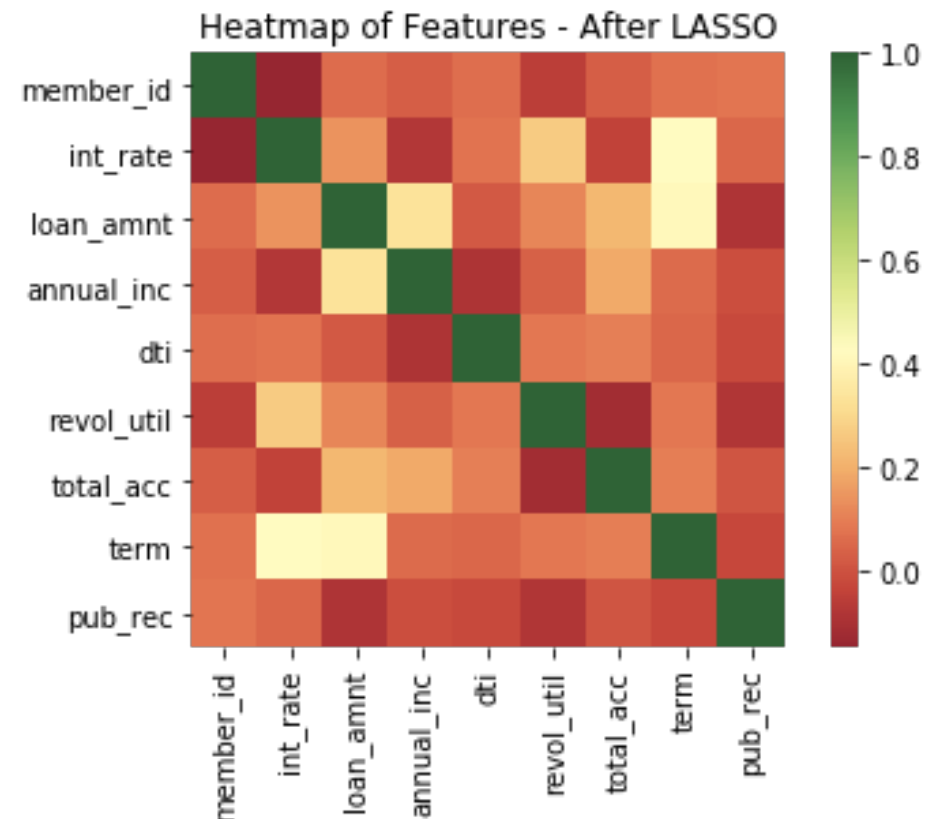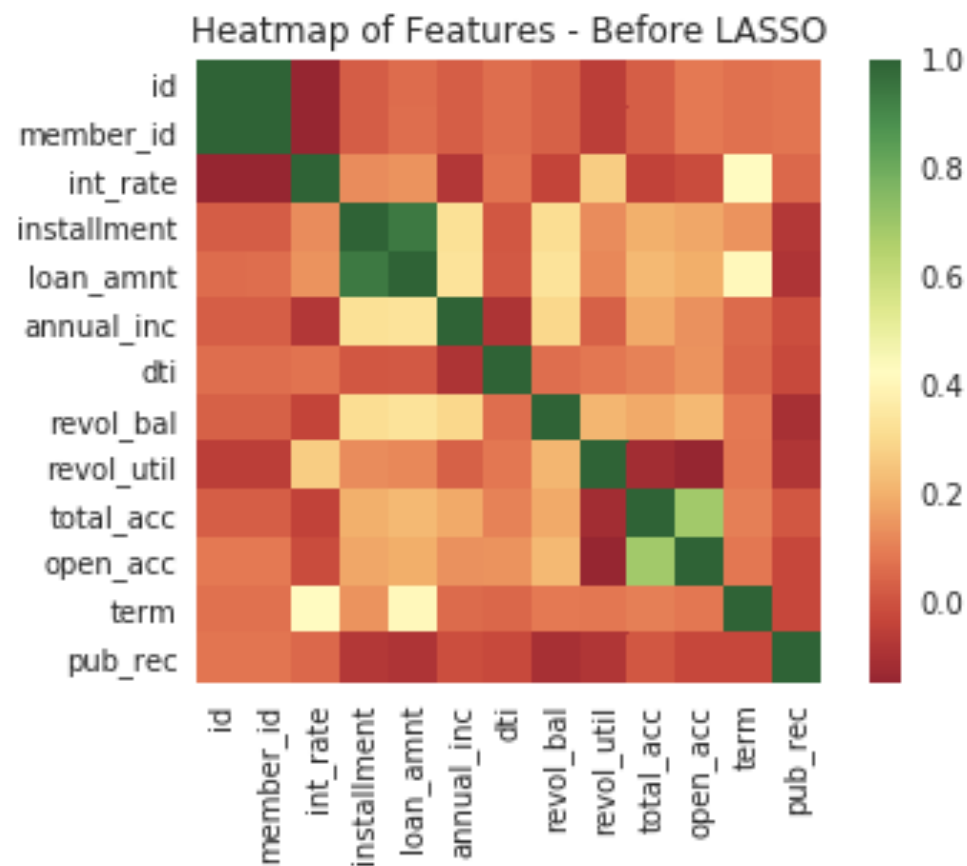
  - 2016: 7.37%

  - 2017: 1.07%

# 2007-2015 Lending Club Loan Data

- ~850,000 observations

- 74 columns
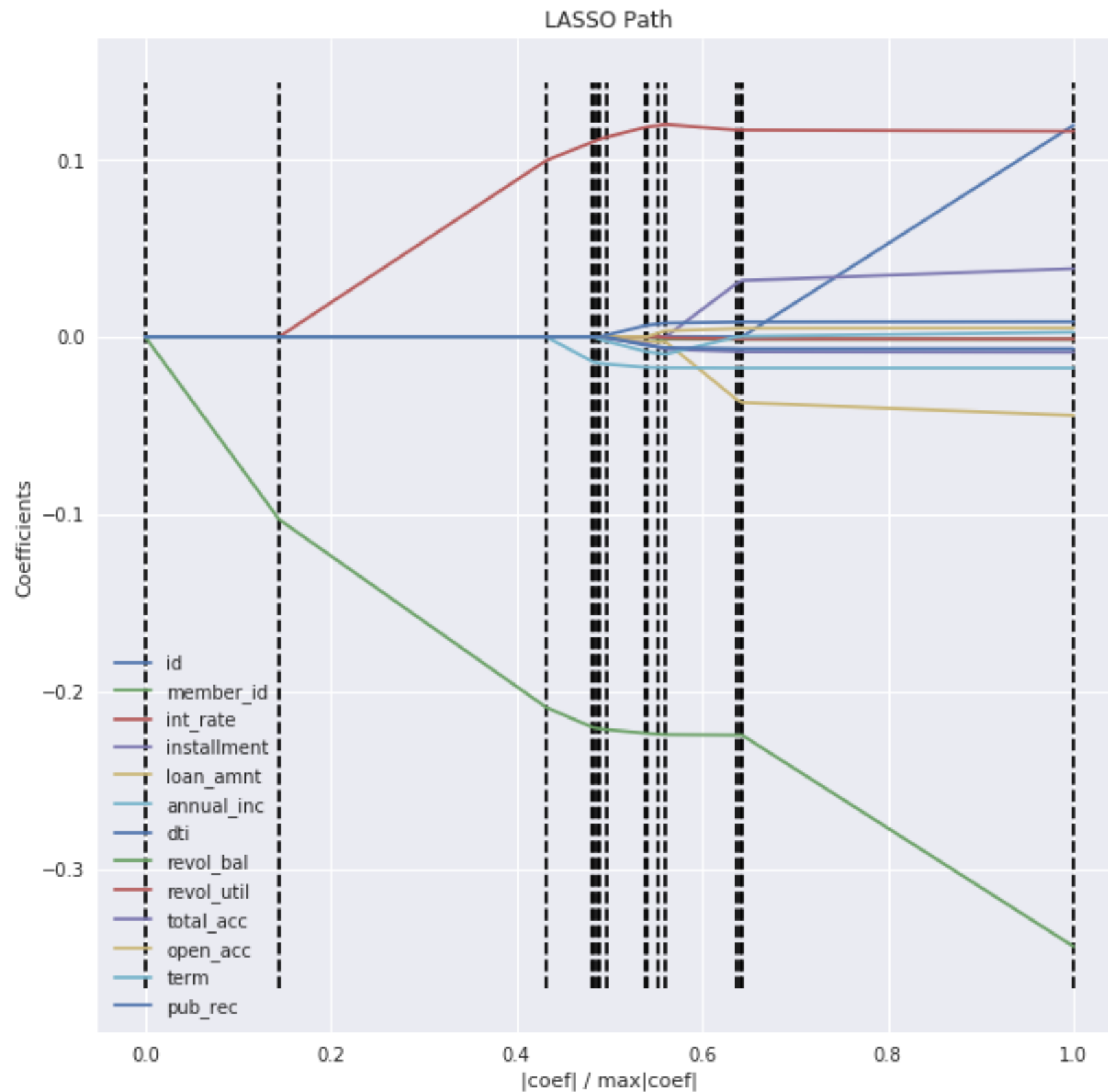
- Target Variable: Loan Status
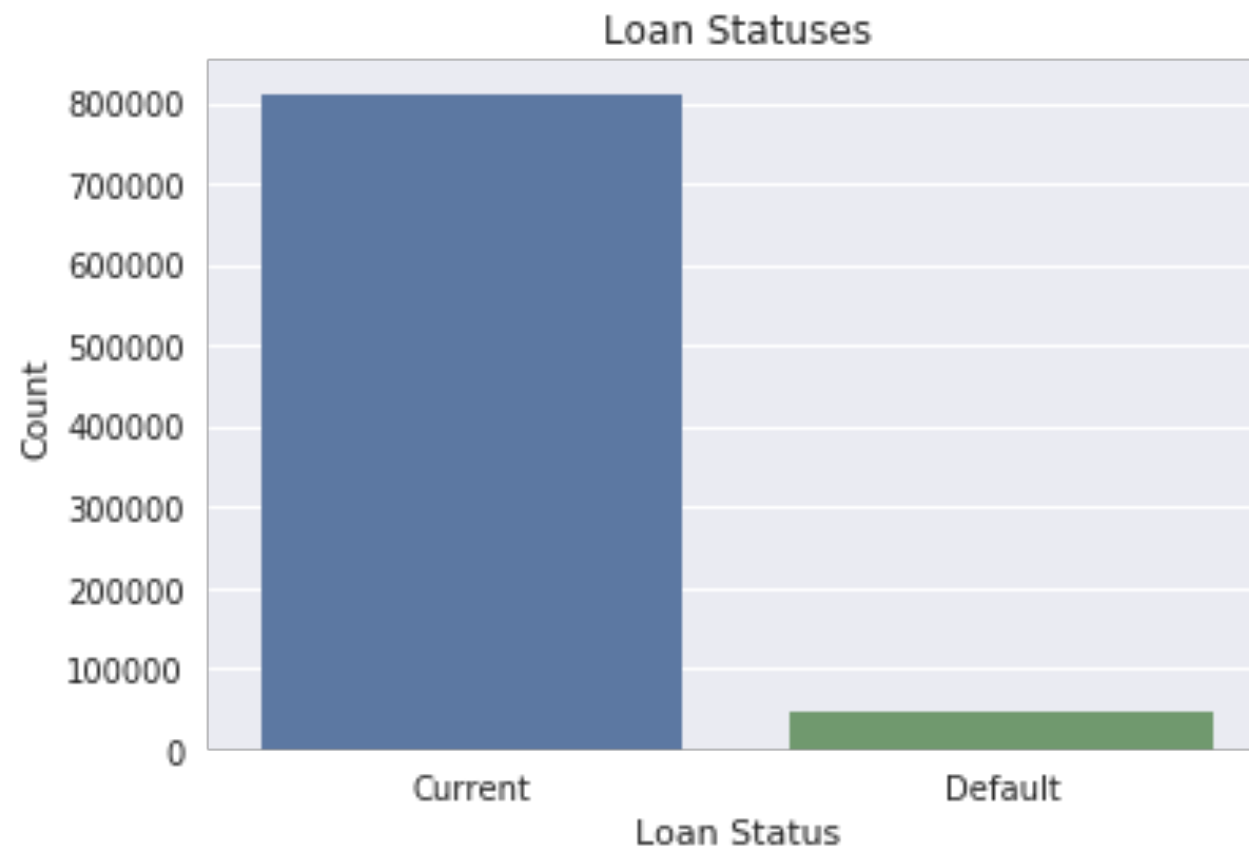
# Logistic Regression Model

- Scaled with StandardScaler

- Oversampled: 'Default' = 40% of data

- LogisticRegressionCV

- LASSO with alpha = 0.001



Heatmap of Features - Before LASSO



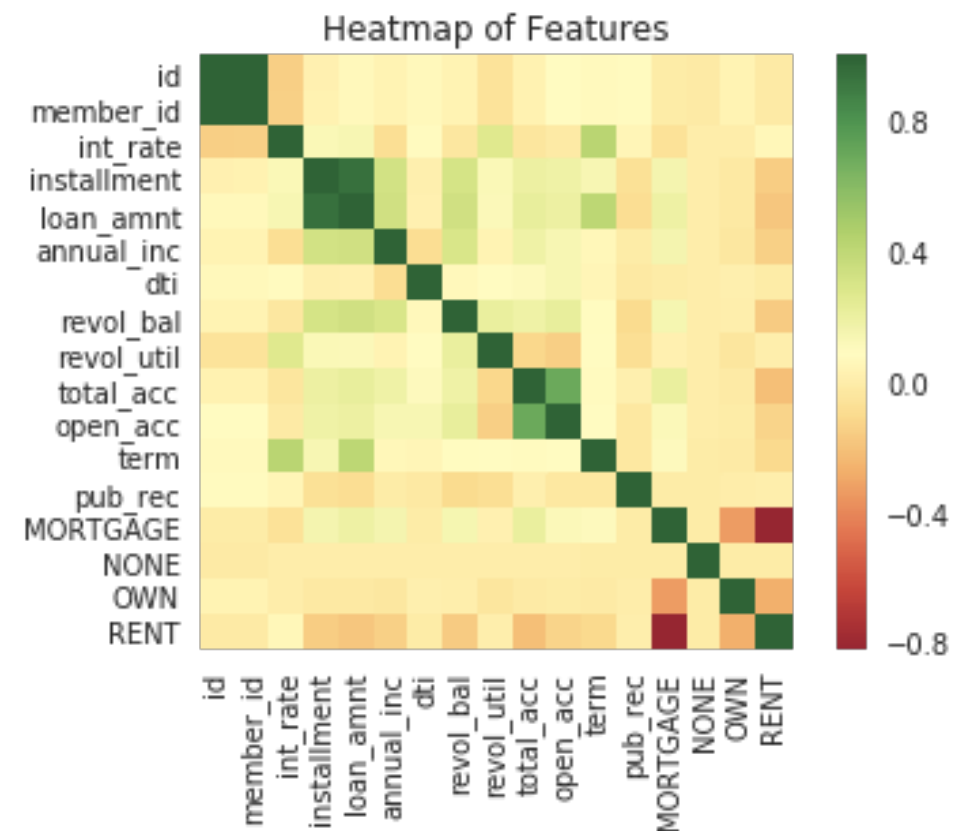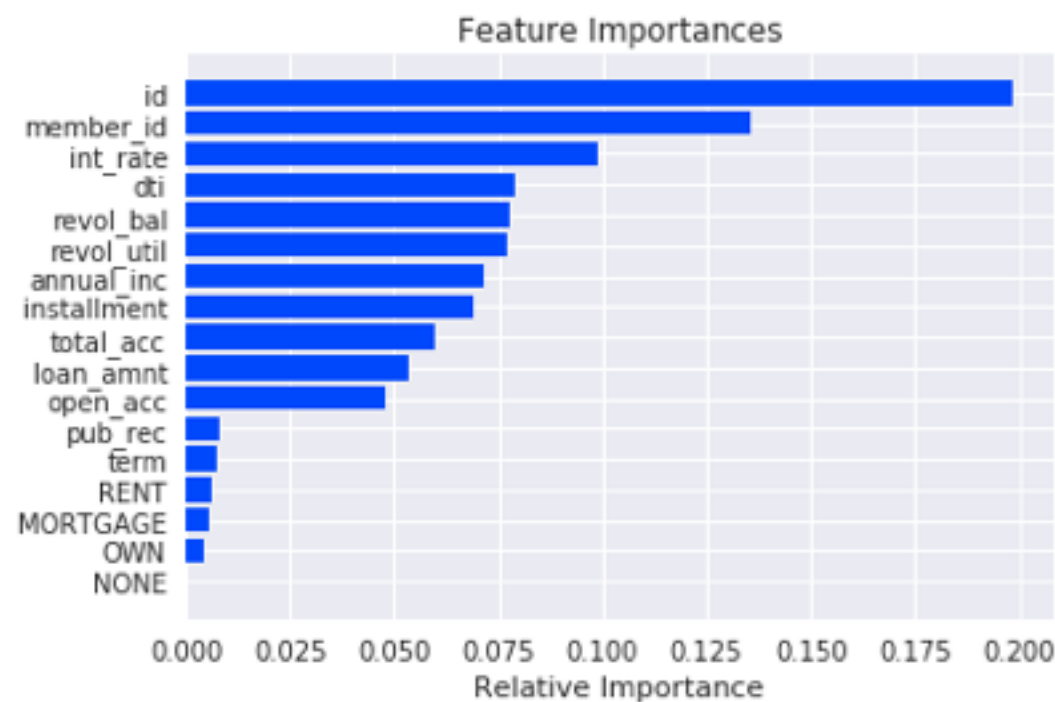Heatmap of Features - After LASSO

# LARS Path

# Random Forest Model

- Oversampled: 'Default' = 40% of data

- RandomForestClassifier

# Logistic

Accuracy Score: 0.8519

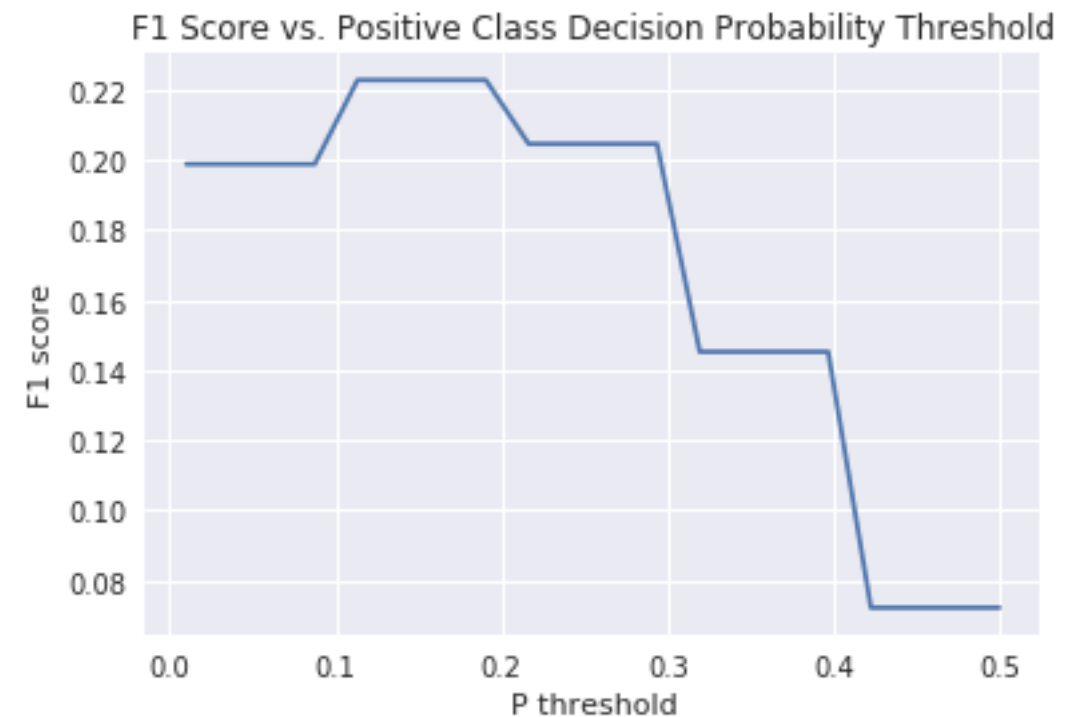|          | Precision | Recall | F-1  |
|----------|-----------|--------|------|
| Current  | 0.97      | 0.87   | 0.92 |
| Default  | 0.18      | 0.49   | 0.27 |



F1 Score vs. Positive Class Decision Probability Threshold
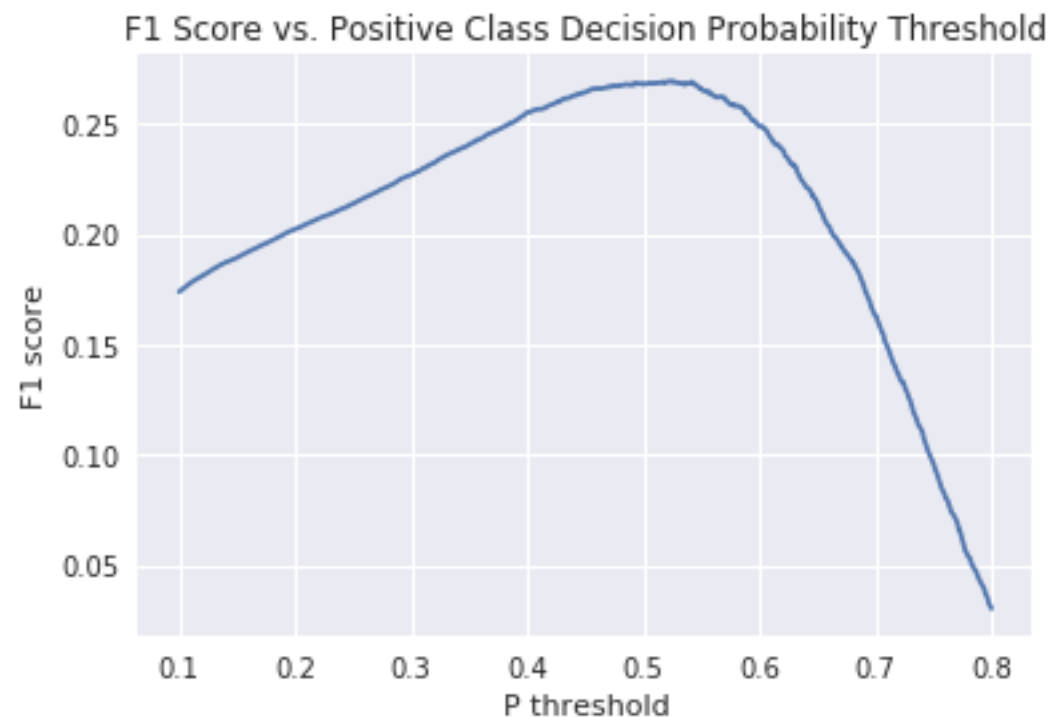
# Random Forest

Accuracy Score: 0.9435

|          | Precision | Recall | F-1  |
|----------|-----------|--------|------|
| Current  | 0.95      | 0.99   | 0.97 |
| Default  | 0.24      | 0.03   | 0.06 |



F1 Score vs. Positive Class Decision Probability Threshold

# Logistic Regression Coefficients

- Member ID: -1.3771

- Interest Rate: 0.6264

- Loan Amount: 0.0226

- Annual Income: -0.2443

- Monthly Debt Payment to Annual Income ratio: 0.2223

- Credit Utilization Rate: -0.03563

- Total Credit Lines: -0.02039

- Loan term: -0.02039

- Number of Derogatory Public Records: -0.007529



Feature Coefficients