Loan Default Prediction - Lending Club

Jiamin Han

Introduction

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.

Peer-to-peer lending

Beijing police shut down P2P lending demonstration

Chinese police shut down a planned protest by groups of investors angry at government inaction over their losses from the peer to peer lending crisis

Save to myFT

Yesterday Produced by Tom Griggs

Data

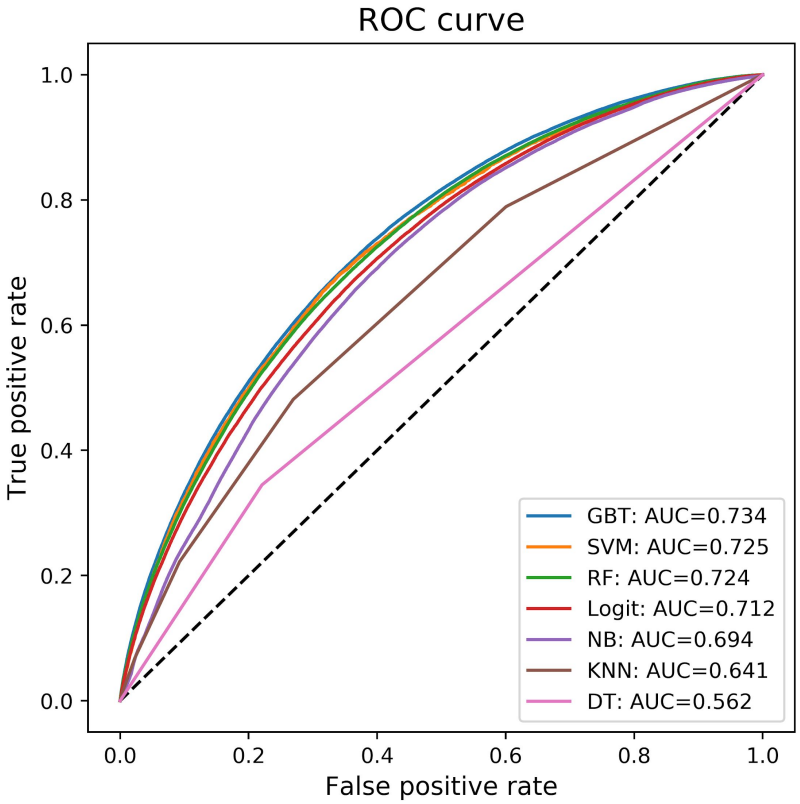
- Source: Lending Club Website
- Size: Completed loans from 2007 to 2018 (n= 1,059,979)
- Features: 28 features at the time of loan application, including loan information, application type and borrower's financial and demographic information
- Dependent variable: loan status - bad or good (paid off on time)

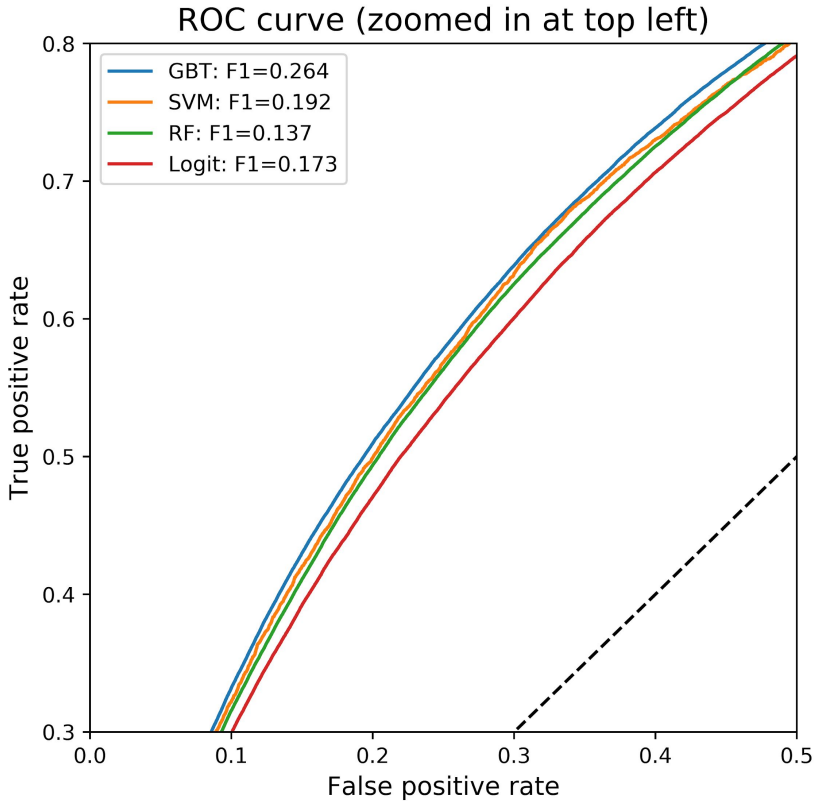
LendingClub

Methods

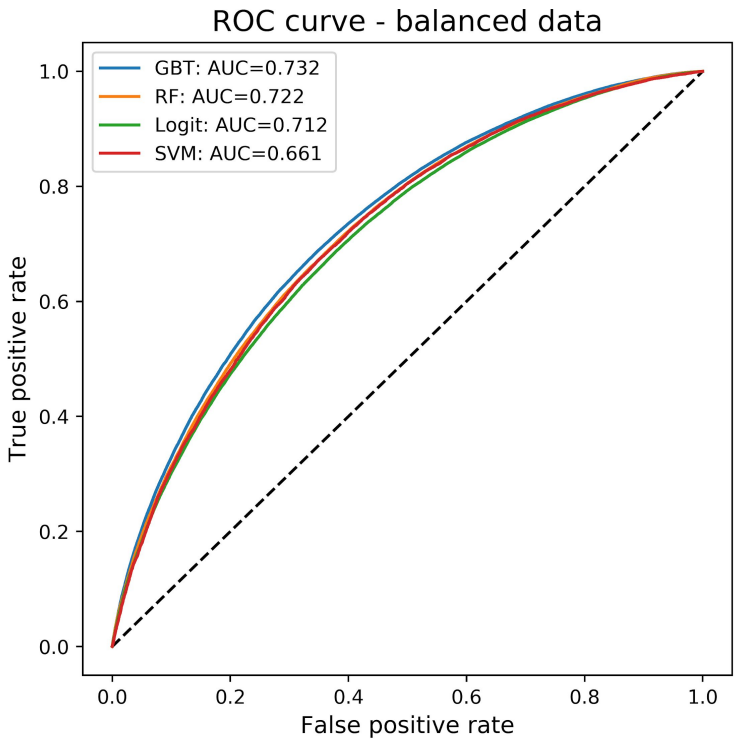

Data cleaning
Transformation
KNN
Logistic Regression
Naive Bayes
SVM
Decision tree
Random forest
Gradient boosting
Cross validation
Gridsearch
AUC
F1 score
Time
EDA
Data process
Feature engineering
Model training
Model tune
Model selection

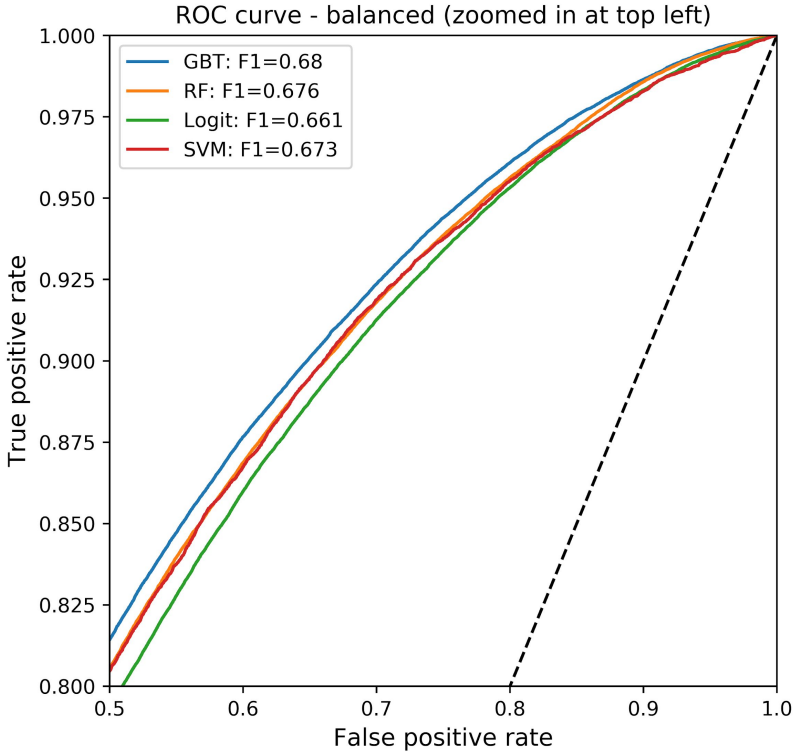
Results - Full data





Results - Balanced data





Key Findings

Feature importance - Gradient Boosting

Average FICO score	0.092
Charge off times within 12 mths	0.074
Total credit revolving balance	0.064
Debt to income ratio	0.061
total number of credit lines	0.058
Interest rate	0.057
Loan description length	0.056
Times of delinquency in 2yrs	0.056
Number of derogatory records	0.055
Loan application year	0.044

Application

- Flag risky loans
- Develop risk-based interests rate

Future Work

Feature engineering using NLP

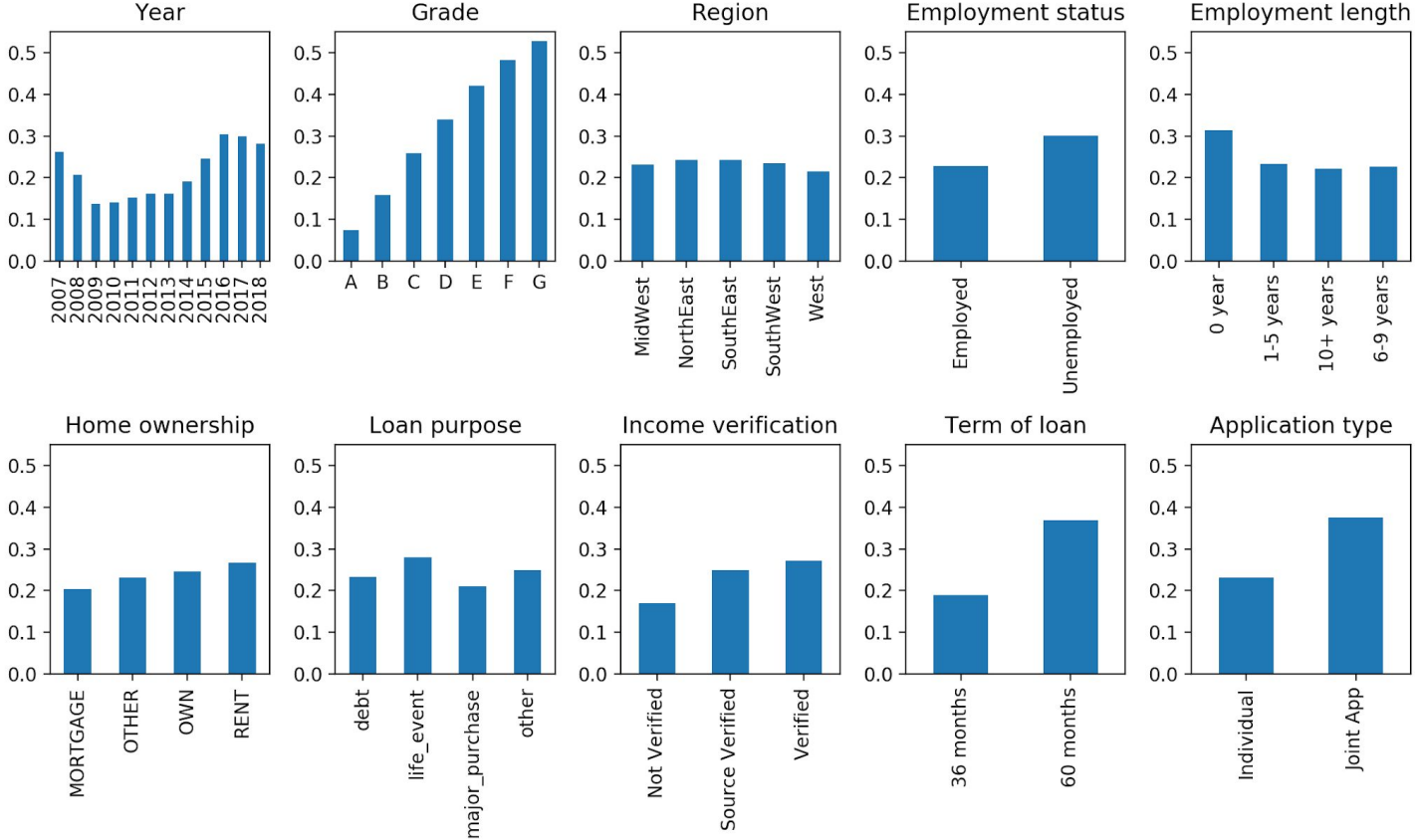
- Loan description/purpose
- Job title

Dimension reduction

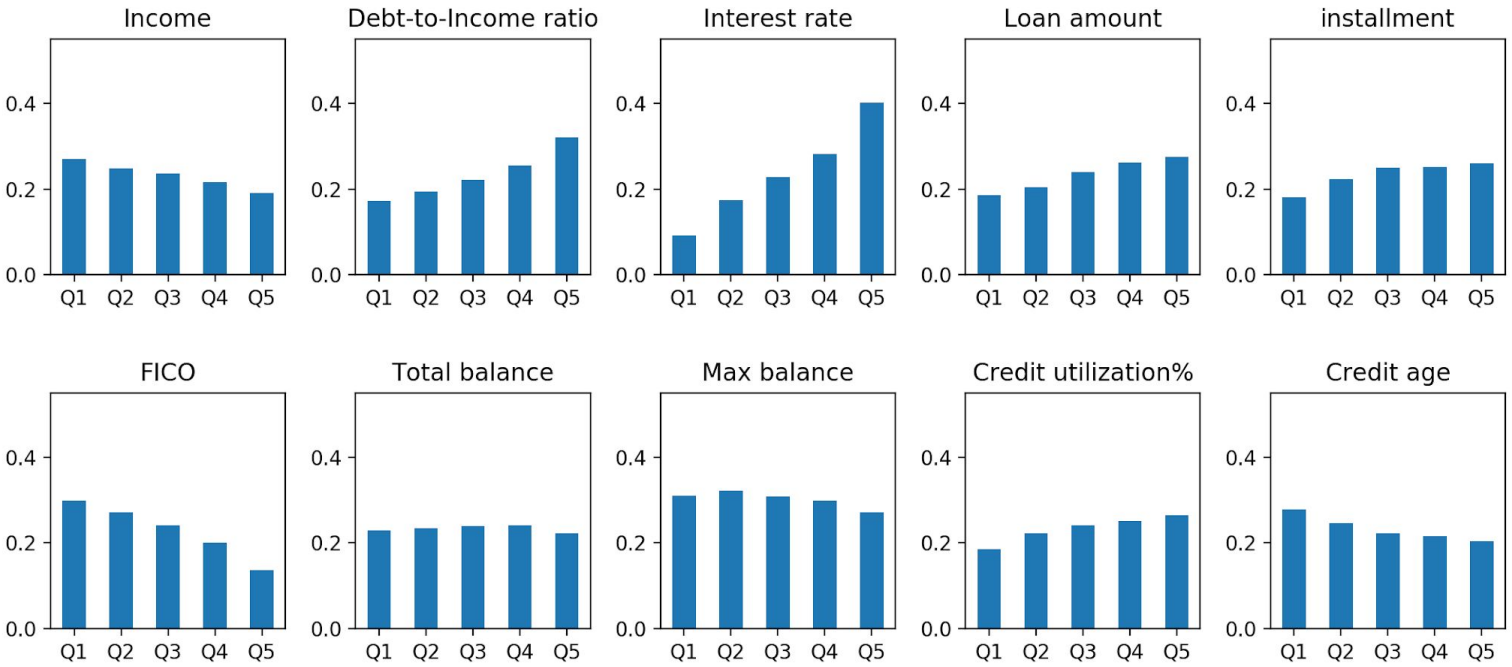
Thank you!

Appendices

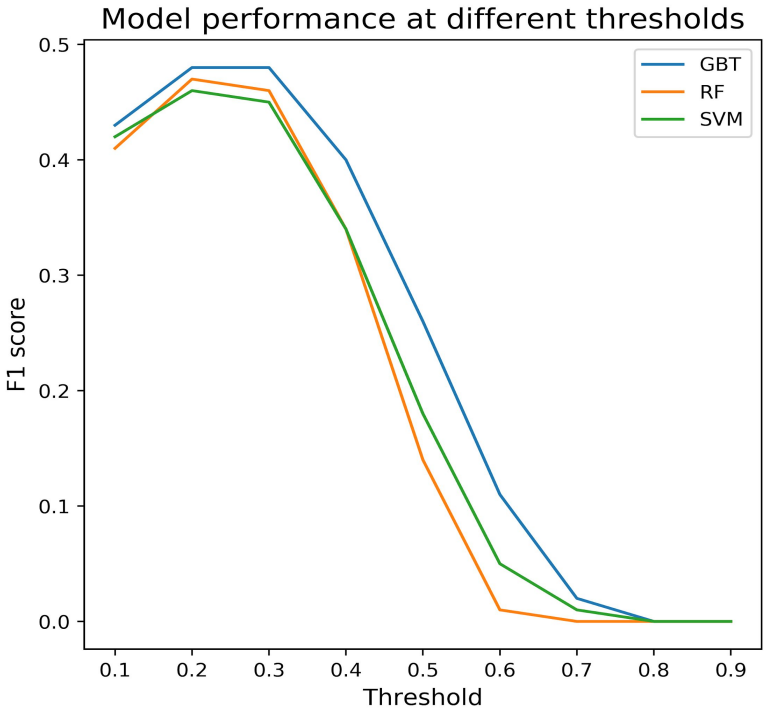
Default rates by feature categories



Default rates by quintiles of features

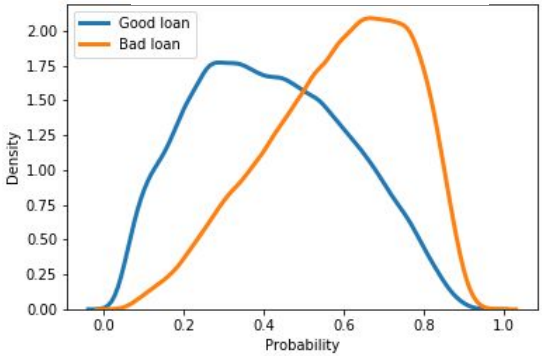


F1 score at different thresholds

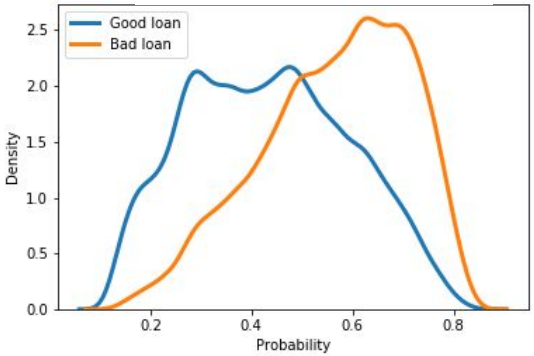


Probability density distribution

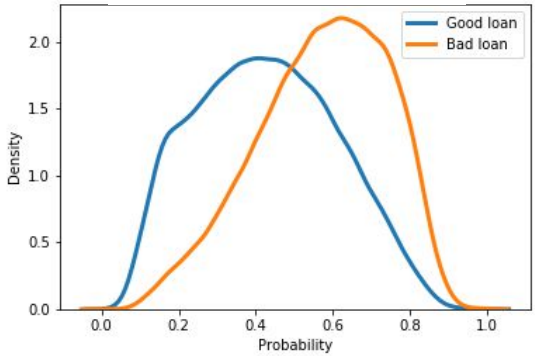
Gradient boosting



Random forest



Logistic regression



Confusion Matrix