



Loan Default Prediction

Katie Fan

Introduction

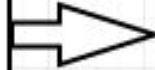
How Lending Club works



Borrowers apply for the loan
Investors open an account



Borrowers get funded
Investor build a portfolio



Borrowers repay automatically
Investors earn & reinvest

Motivation & Goal

- Motivation:
help investors make more objective and data-driven assessment of loan applications to minimize risk
- Goal:
use data insights to predict loan will be defaulted or not and find the important factors

Data

- Data Source:

LendingClub

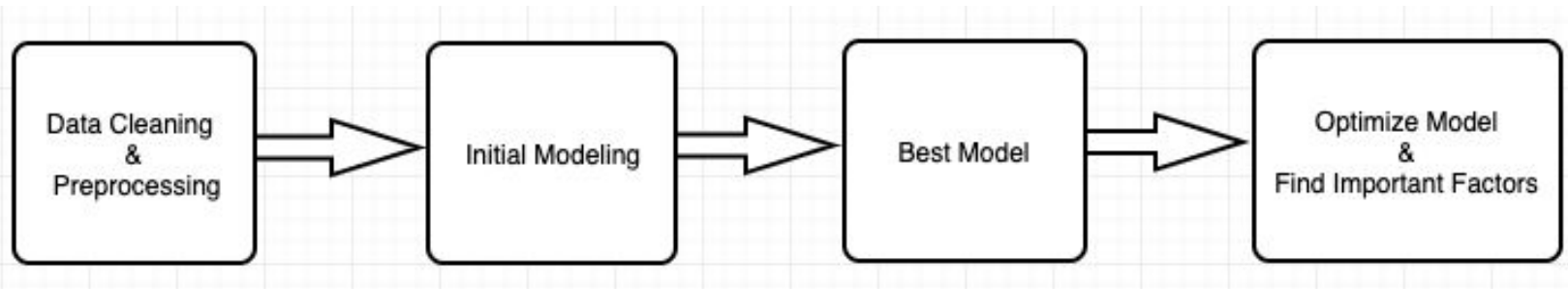
- Data Size:

2,260,701 rows and 145 columns

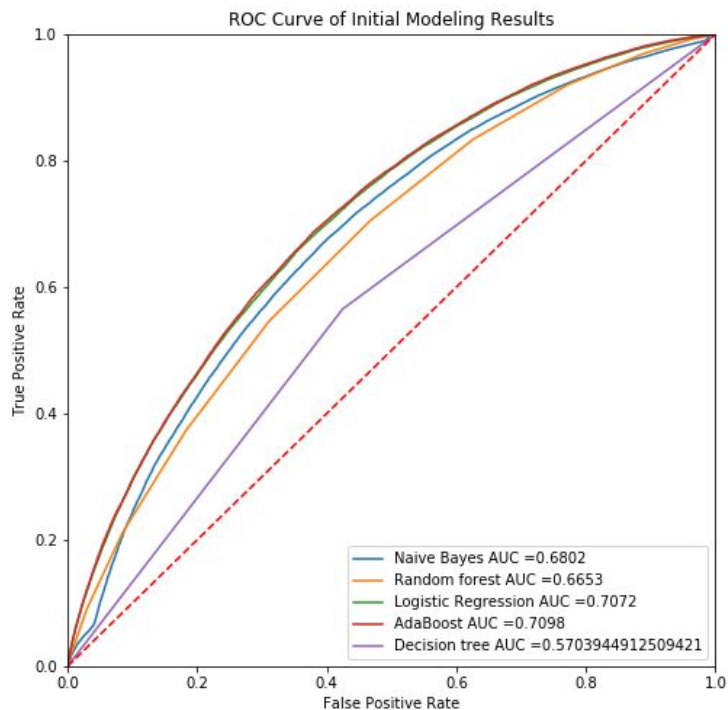
- Target Variable:

Loan Status: default, fully paid

Methodology



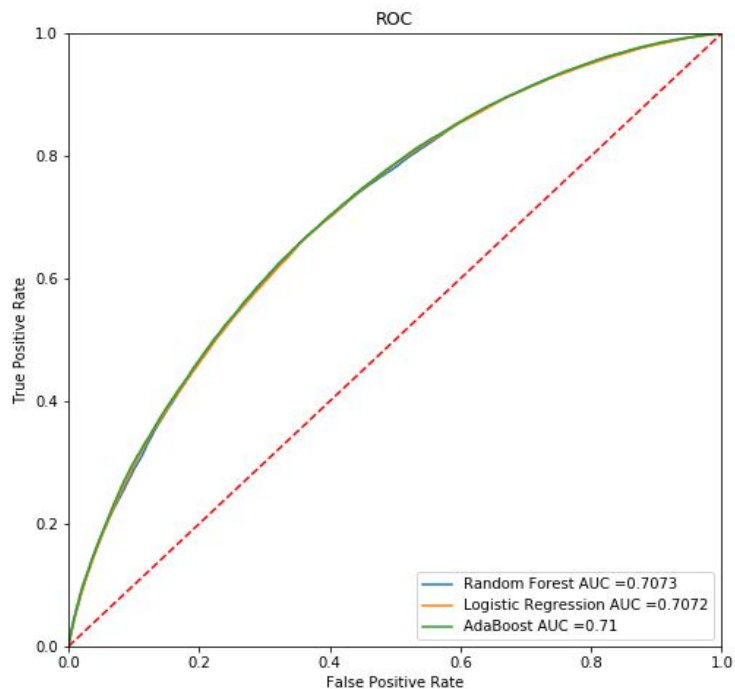
Initial Modeling



Based on performance (accuracy, time, etc.), choose below models as candidate models

- Random Forest
- Logistic Regression
- AdaBoost

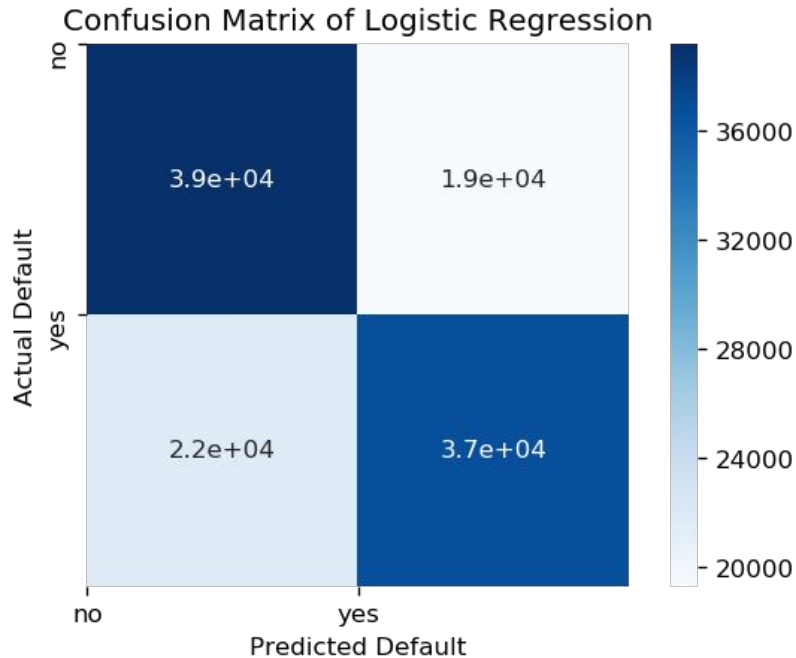
Model Evaluation



- Best model: Logistic Regression

Best Model

Results of Logistic Regression

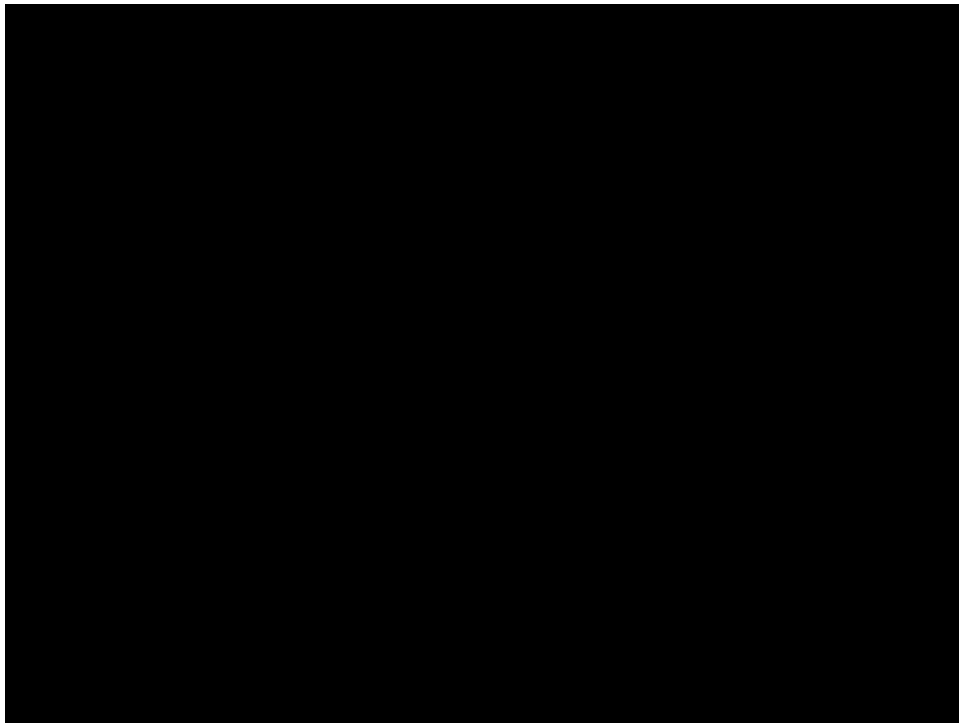


Top 5 Important Factors

Important Factors

- interest rate
- grade
- debt to income ratio
- number of mortgage accounts
- number of open credit lines in the borrower's credit file

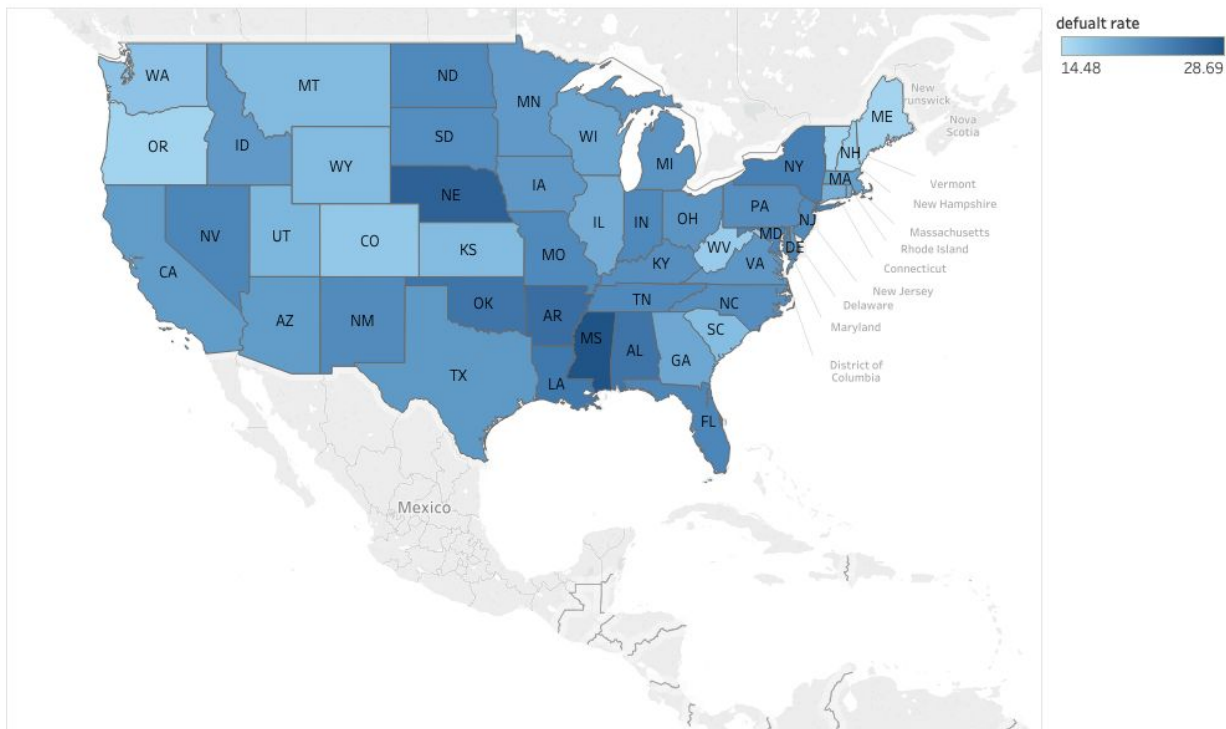
Prediction



Accuracy: 12/20

Appendix

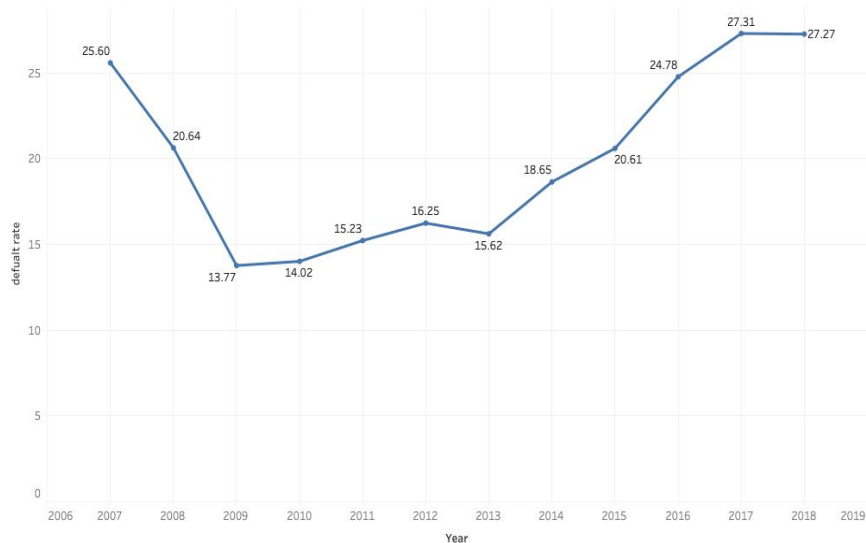
Default Rate by State



Map based on Longitude (generated) and Latitude (generated). Color shows default rate. The marks are labeled by Addr State. Details are shown for Addr State.

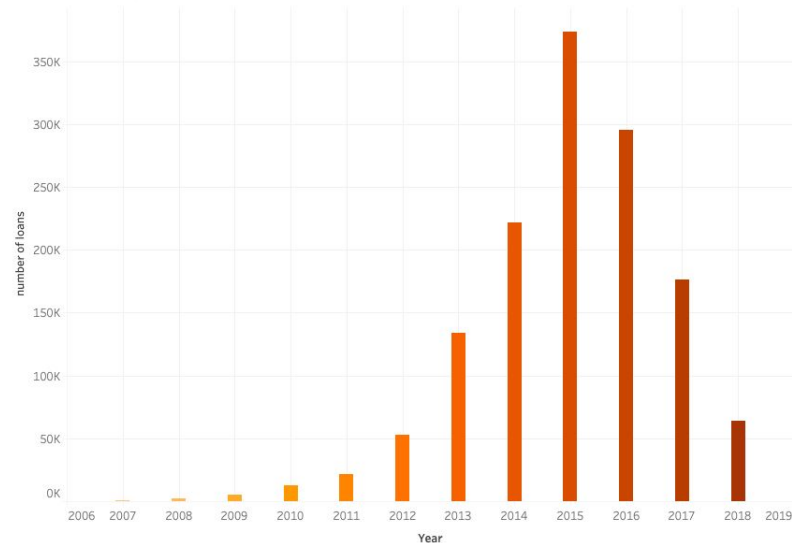
Appendix

Default Rate by Year



The trend of default rate for Year. The marks are labeled by default rate.

Loan Funded by Year



The plot of number of loans for Year. Color shows details about Year.

Thank you!