Predicting repayment of a small loan

Anna Zubova

Project repository:
https://github.com/AnnaLara/predicting_loan_repayment

Loans up to 500$

Opportunity:

1 in 10 American consumers has no credit history (2015)*.

Source: *https://files.consumerfinance.gov/f/201505_cfpb_data-point-credit-invisibles.pdf*

Seattle financial startup

- Grants loans up to 500$ to people with no credit score
- Reports positive outcome to credit authorities
- Uses bank transactions data

Loans up to 500$

Problem: How to balance costs?

$ loss : 1 bad loan

=

$ profit: 6 good loans

Data

- Real data from a financial start-up
- 4145 approved applications with labels
- 2927 rejected applications
- 36 bank transaction features

days_until_next_payroll
overdraft_transactions
pay_advances
title_loan
total_loan_payments
government_aid_weekly_income
nsf_transactions
income_length
reverse_transactions

Data Preparation

Modeling

Models:

- Logistic Regression
- Random Forests
- XGBoost
- Gradient Boosting

AUC : **0.62 - 0.65**

Balanced 0 (default) and 1 (paid back) classes

Final choice: Logistic Regression

Evaluation: class 0 - default, 1 - paid back

Cross-validated Log loss :

-0.56 ($p \approx 0.58$)

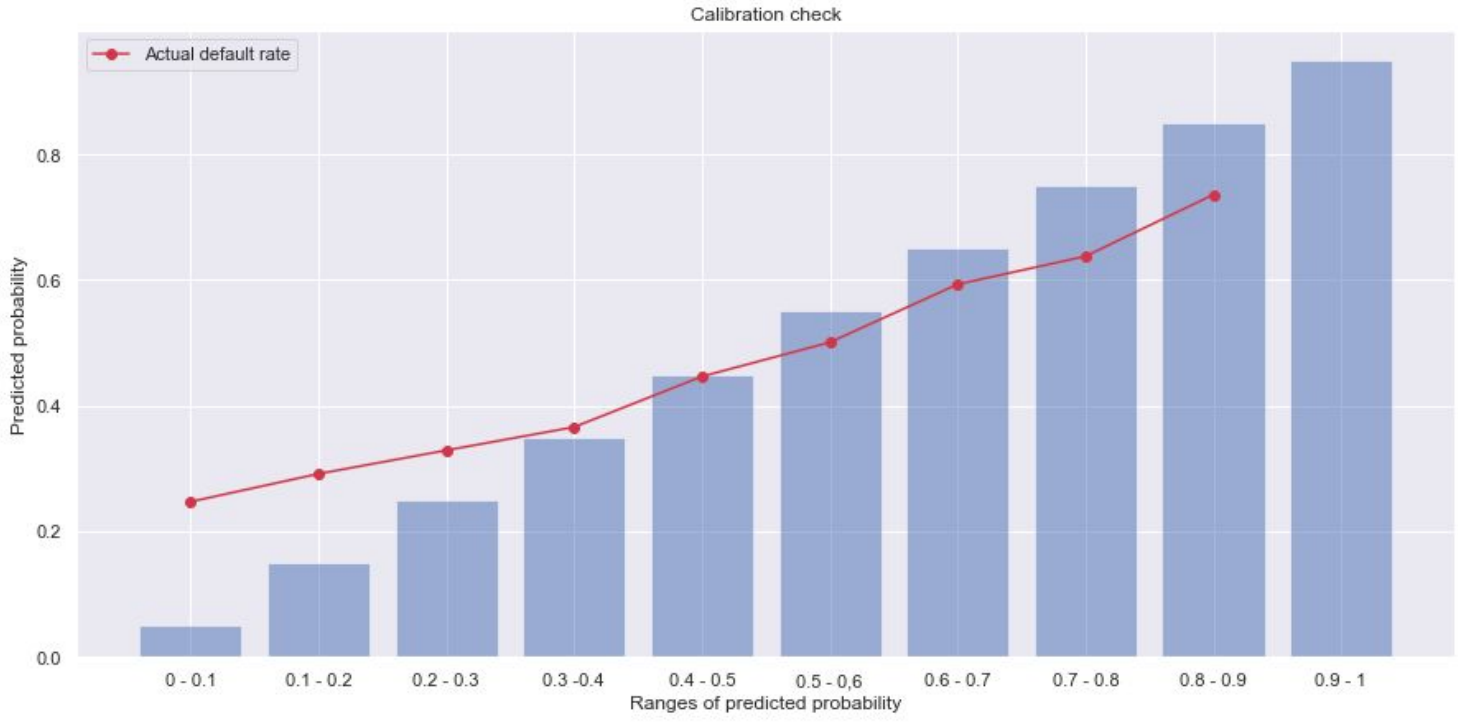
Cross-validated AUC:

0.63

	precision	recall	f1-score
0	0.23	0.97	0.38
1	0.92	0.09	0.16

Threshold: 20% for 0 class

Predictive power (calibration check)



Deployment: Flask App on AWS EC2

plr.annalara.net

Next steps

- Prediction model by state
- Improve model’s prediction performance

Thank you!

Get in touch:

Email: anna@leolog.com

LinkedIn: https://www.linkedin.com/in/annazubova/

GitHub: https://github.com/AnnaLara/