Safe to Loan

Safer. Simpler. Smarter

YUEYING(SHARON) ZHANG

Motivation

Banks struggle to target customers with insufficient or non-existent credit histories and have a hard time to decide whether to grant loans to those customers.

This project aims to provide a positive and safe loan experience to a U.S bank. Using alternative data sources, the project can help the bank evaluate the repayment ability of those “underserved” customers, expand the loan servicing and generate more revenue.

Data

- **application** (~300K): Static data of the applicants such as age, income, education, housing and family.

- **bureau** (~1.7M): Previous credits provided by other financial institutions that were reported to Credit Bureau such as current debt on Credit Bureau credit

- **bureau_balance** (~27M): Monthly balances of previous credits in Credit Bureau.

- **POS_CASH_balance** (~10M): Monthly balance snapshots of previous POS (point of sales) and cash loans that the applicant had with Home Credit

- **credit_card_balance** (~3.8M): Monthly balance snapshots of previous credit cards that the applicant has with Home Credit

- **previous_application** (~1.7M): All previous applications for Home Credit loans of applicants

- **installments_payment** (~14M): Repayment history for the previously disbursed credits in Home Credit related to the loans in the sample

Model

- 14 features selected in the model including age, employment, payment rate, income-credit ratio, id and phone changing and mean days before due on installments

- Tuned and compared RandomForest, Xgboost and LightGBM

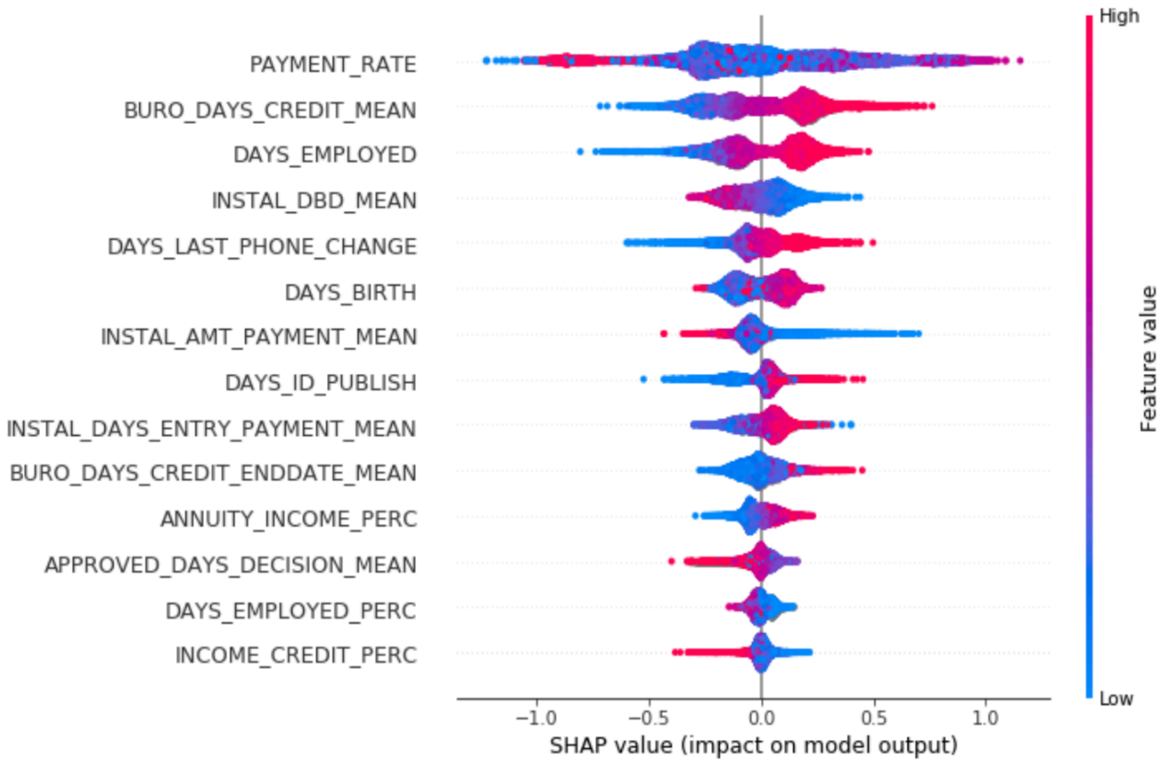
- Selected **Xgboost** with number of tree 300, learning rate 0.2 and max depth 3 as the final model

- AUC of Xgboost on test data is **0.706**

- With threshold 0.08, classification accuracy is **66%** and true positive rate/recall/sensitivity (the percentage of late-payment people who are correctly identified as having late payment) is **65%**

- Success criteria:
  - 1) Test AUC exceeds 0.70
  - 2) The additional revenue from lending to those underserved customers reaches the finance goal of the bank (e.g. less than 2% default risk, generating more than 1 million revenue from those customers)

Insights



- **Unstable**: The more recently a client changes id or phone number, the more likely he/she will make late payment.

- **Low Income**: A client with high annuity-income ratio and low income-credit ratio has larger chance to make late payment.

- **Borrow Recently**: A client who has applied for credit from credit bureau recently is more likely to make late payment.

*Days are recorded as negative in the dataset. For instance, if a person was born 1 year ago, DAYS_BIRTH is recorded as -365.

Contact

For more information on this project,

please refer to https://github.com/sharonZzz96/MSiA423-final-project-2019

or contact yueyingzhang2019@u.northwestern.edu

Thank you!