Responsible Data Science, Spring 2021 Course Project

A Nutritional Label for an ADS Predicting Loan Default Risk

Erin Choi, Vicky Lin

ADS Background

- ADS: Predicts the loan repayment abilities of customers of Home Credit
- Goal: Broaden financial inclusion for underbanked and low-income populations by focusing on lending to people with little or no credit history
- ADS Implementation: “Start Here: A Gentle Introduction", by Will Koehrsen

HOME CREDIT
0%*
EASY EMI OPTIONS

Input and Output

Input

The 122 features in the training set include clients' income, clients' highest level of education, credit amount of the loan, and three external sources, among many others.

The sensitive attributes in the data include clients' gender and age in days at the time of the application.

Output

The target variable in the training set represents the clients' repayment abilities;

- 1 – indicates that a client had payment difficulties
- 0 – all other cases (successful repayment).

The outcome is a value between 0 and 1, representing the probability that the loan will not be repaid (probability of the client being classified with the value 1).

ADS Implementation

Data Cleaning and Preprocessing

Encoding Methods

One-hot encoding and label encoding

Missing Value and Anomalies

The anomalous values are replaced with nan values. Missing and nan values are filled with the median of the respective features by the imputer.

Feature Construction Methods

Domain knowledge features

Scaling Method

MinMax Scaler

Description of Implementation

Random Forest Classifier with Domain Features

A random forest classifier was created with 100 trees and a seed.
Domain knowledge features were included in the training and testing sets.

ADS Validation

AUC score: the higher the value, the more accurate the model.

Using domain features increased the AUC from 0.678 to 0.679.

ADS Outcomes: Accuracy

Overall and Group Accuracy Based on Sensitive Attributes

	Age (>= 43 | < 43)	Gender (M | F)	Income (>= $147,150 | < “”)
Overall Accuracy	0.919305	0.919228	0.919321
Privileged group accuracy	0.93736	0.897438	0.922877
Unprivileged group accuracy	0.901104	0.930683	0.915758

ADS Outcomes: Fairness

Disparate Impact and FP/FN Rate Difference Based on Sensitive Attributes

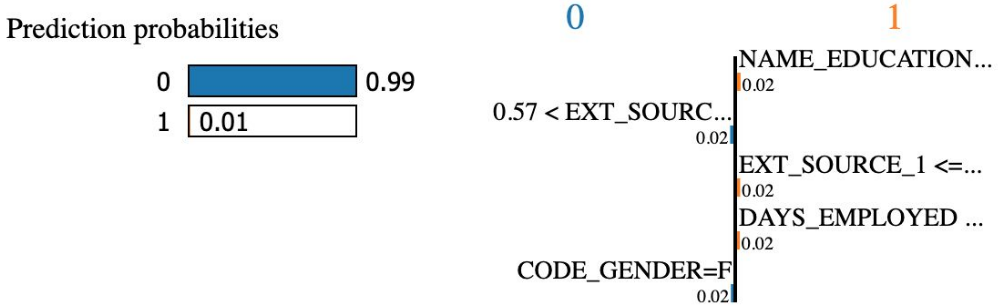
	Age (>= 43 | < 43)	Gender (M | F)	Income (>= $147,150 | < “”)
Disparate Impact	8.064516	0.347657	2.337966
False Positive Rate Difference	0.000036	-0.000026	0.000036
False Negative Rate Difference	-0.00179	0.001239	-0.001052

ADS Outcomes: LIME

Prediction Explanations for 2 Important Clients in the Data

Misclassification

```
Actual class:  1
```

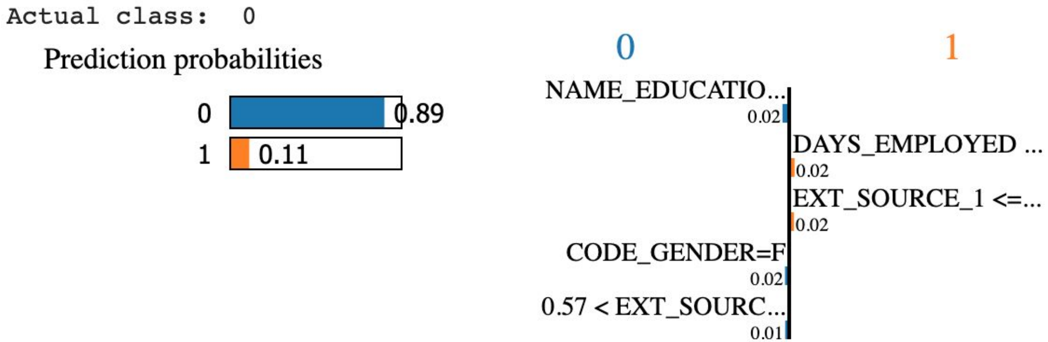


Feature	Value
NAME_EDUCATION_TYPE=Secondary / secondary special	True
EXT_SOURCE_2	0.62
EXT_SOURCE_1	0.51
DAYS_EMPLOYED	365243.00
CODE_GENDER=F	True

ADS Outcomes: LIME

Prediction Explanations for 2 Important Clients in the Data

Correct Classification



Feature	Value
NAME_EDUCATION_TYPE=Higher education	True
DAYS_EMPLOYED	-84.00
EXT_SOURCE_1	0.22
CODE_GENDER=F	True
EXT_SOURCE_2	0.65

Summary of Findings

Was the data appropriate?

Yes.

The data was provided by Home Credit for competitors to use to create a system that would potentially be deployed by the company themselves.

Robustness, Accuracy, and Fairness

- Some robustness from pre-processing anomalous values, but potentially decreased robustness from uninterpretable values of External Source variables
- Fairly accurate overall – accuracy scores around/over 0.9, AUC scores around/over 0.7
- Not fair enough – large disparate impact values, could lead to certain group membership causing unfavorable classification

Summary of Findings

Should the ADS be deployed?

We **do not think** this ADS should be deployed, at least in its current state.

- Contradiction between Home Credit's mission and system's fairness
- Should provide additional guidance for those who may be less likely to repay loans on time

Potential Improvements

- Increased transparency for the data collection
- Description of the External Source features
- Processing: balancing proportion of clients in the privileged and unprivileged groups based on sensitive features
- Hyperparameter tuning based on disparate impact and some accuracy measure