Credit card usage and default risk

by Team Oak
Alexander Timofeev
Martin Molina

The Erdos Institute, Fall 2022 Bootcamp

Overview

Problem: To predict the default risk of a client on a financial product using only bank card transactions

Stakeholders: A large bank interested in assessing whether a client applying for a credit will default on it

KPI: Increasing the proportion of applicants correctly predicted to result in eventual default, minimizing losses caused by defaulting clients, reducing application processing time

Example: A bank stores detailed card transaction and default history of its clients. One day one of its client applies for a loan or a new credit card. Should the bank approve this client's application based on her card transactions and data from other clients?

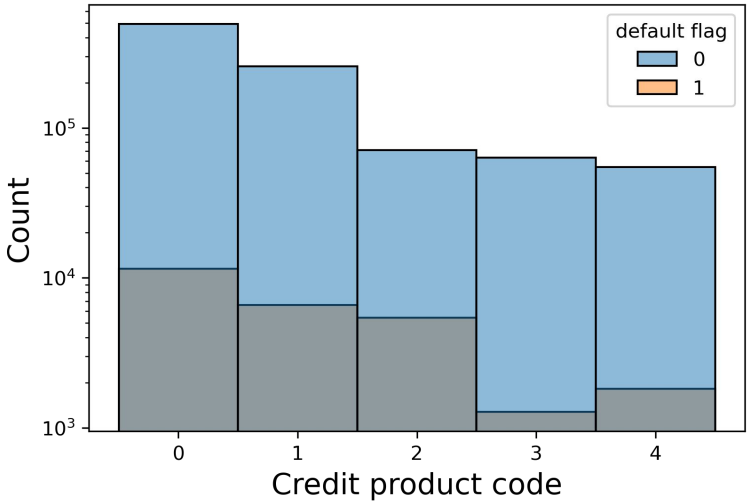
Dataset description

1.5 million anonymous **records of** approved credit **applications** and **450 million card transactions total** (4.2 GB of train data in parquet files).

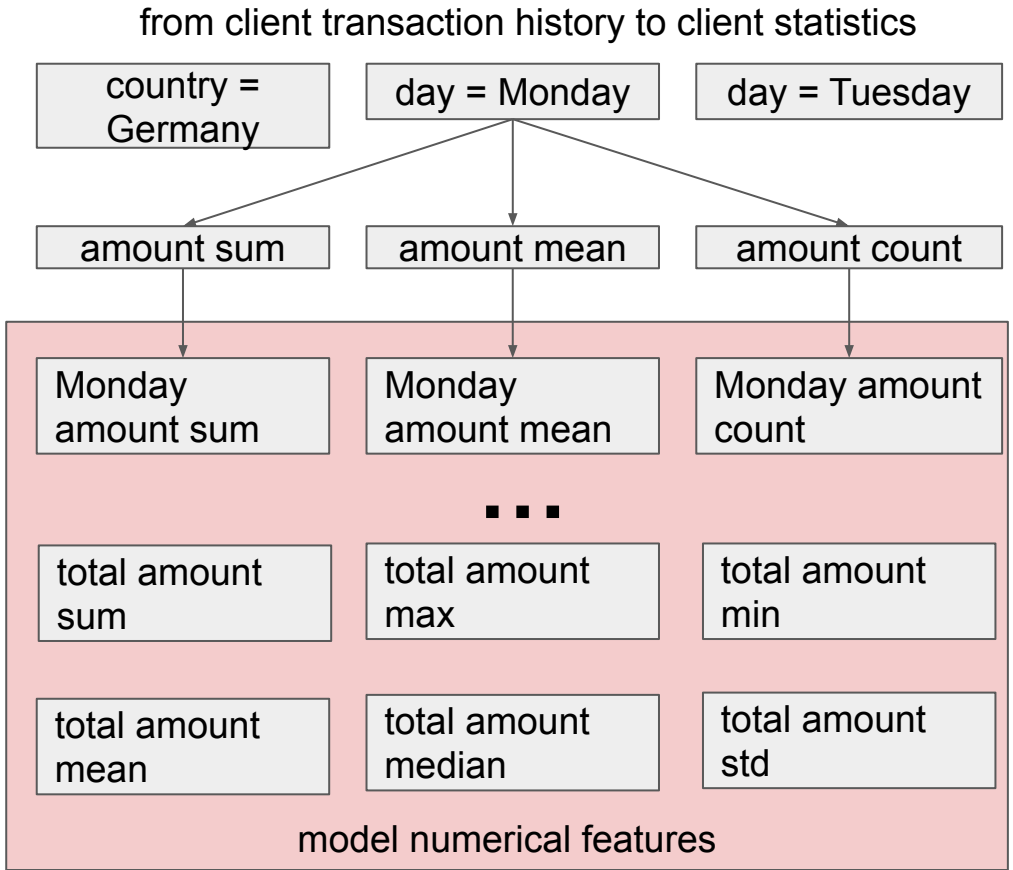
For each credit product application, the dataset includes detailed history of the client's bank card transactions over the previous year to approval.

Each card transaction has **18 features** (amount, time, card type, currency, location, e-commerce flag, payment system, MCC, etc).

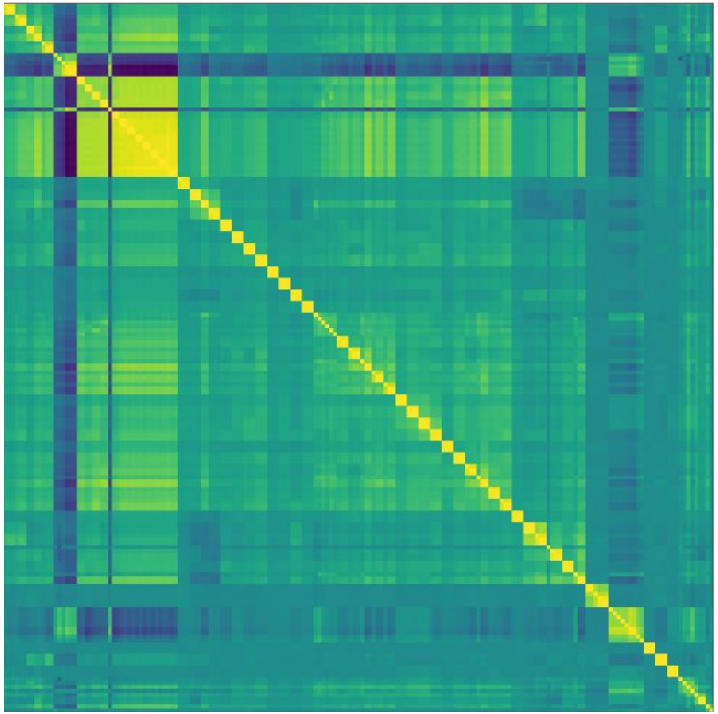
EDA & Feature extraction



Imbalanced classification
=> use ROC AUC metric



EDA & Feature extraction



Numerical features

- small clusters of highly correlated features
- PCA: 90 % explained variance
- before reduction: 183 features
- after reduction: 79 features

Categorical features

- only credit product code

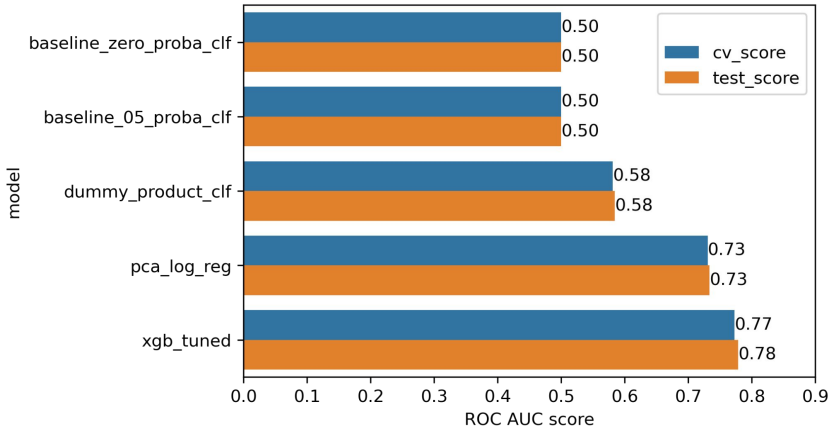
Models

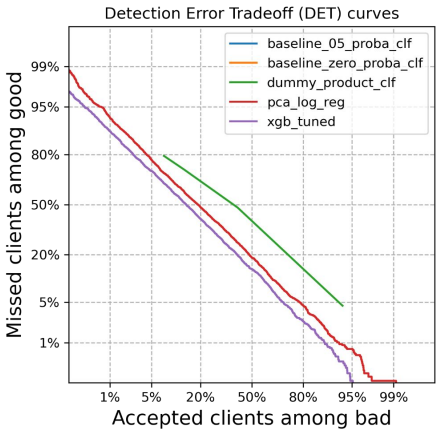
Baselines

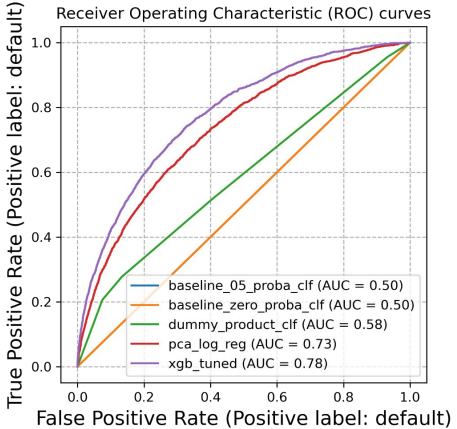
- default probability = 0
- default probability = 0.5
- default probability is a prior default probability conditioned on a credit product type

Logistic regression with PCA reduction for numerical features and one-hot encoding for product type

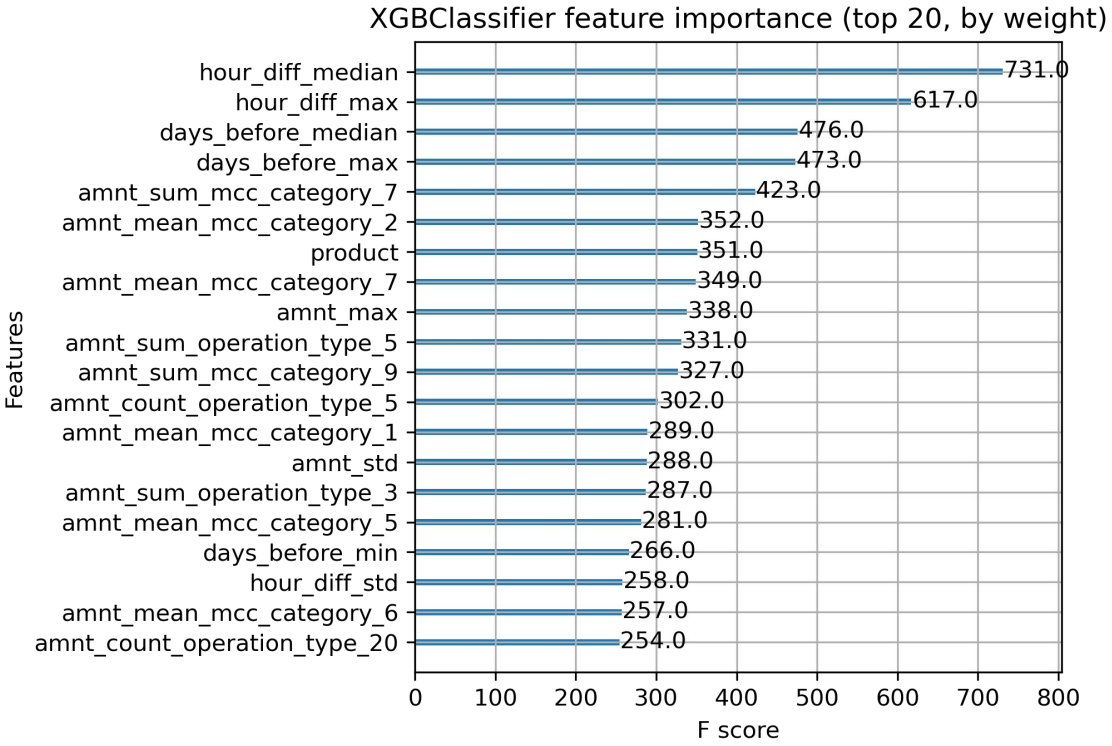
XGBoost classifier (GPU for hyperparameter grid search with cross-validation and CPU for demo deployment)

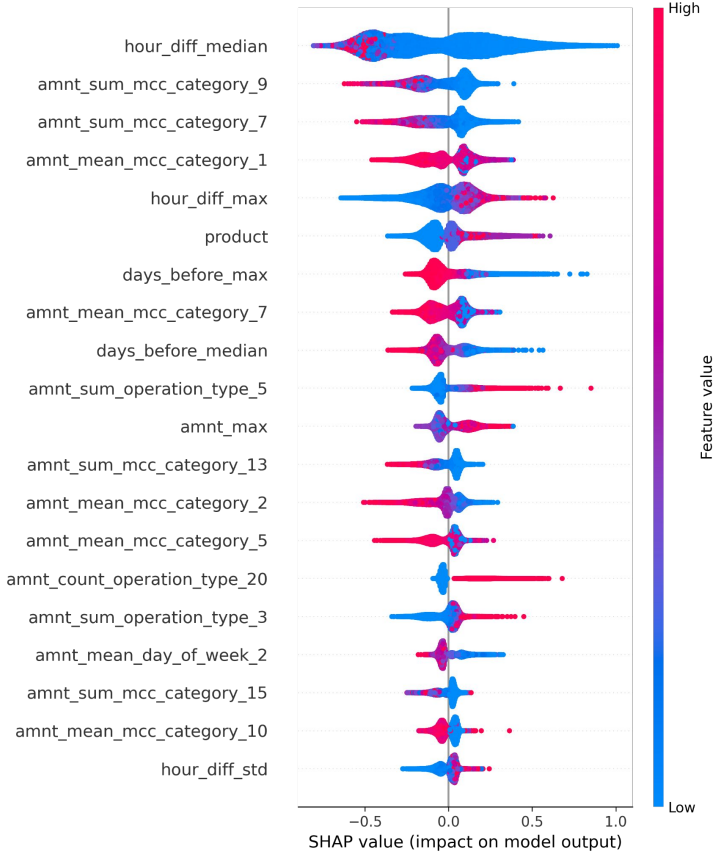




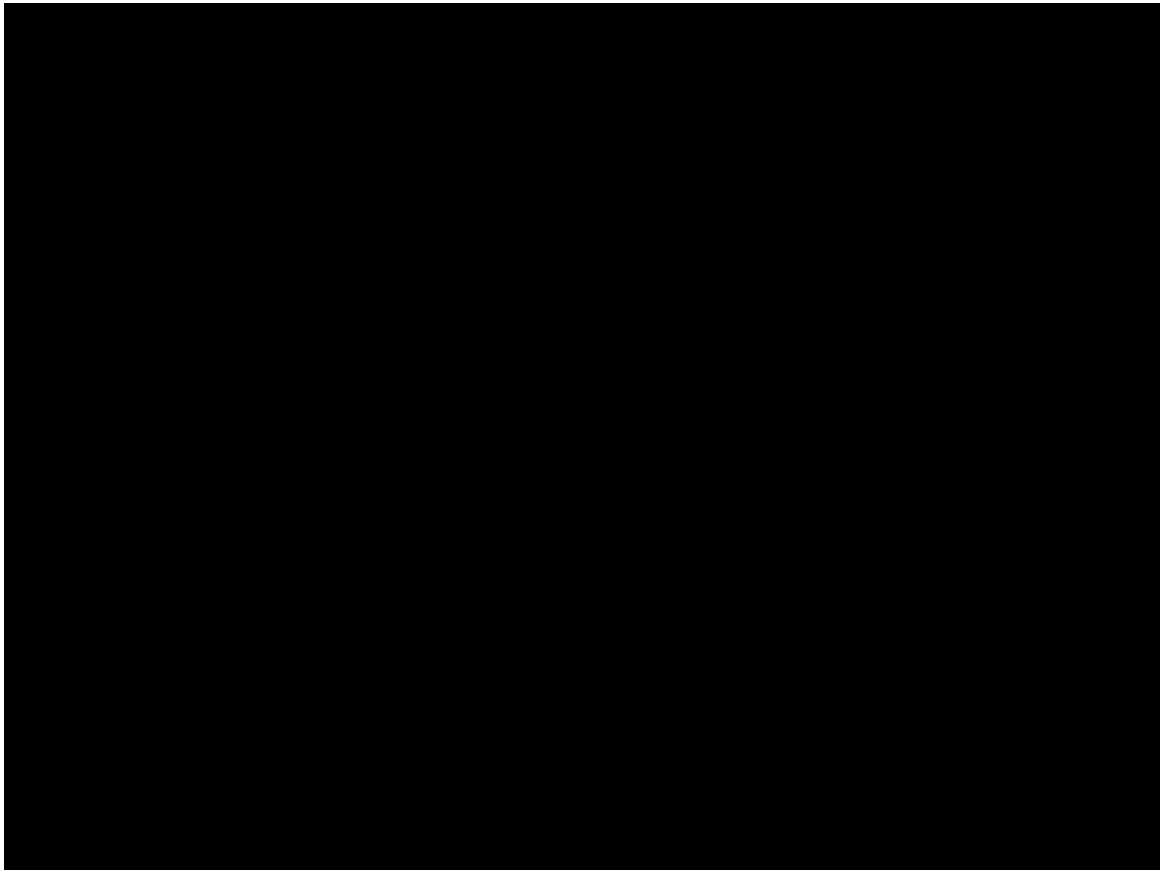


XGBClassifier feature importance





Demo application



Links

Github: https://github.com/fall22-oak/oak_main

Demo: https://huggingface.co/spaces/alex42t/CreditScore

Alexander: https://www.linkedin.com/in/rs42

Martin: https://www.linkedin.com/in/martinmf

Special thanks to Kash Bari for his help during this project.

Results and future directions

We were able to considerably increase the proportion of applicants correctly predicted to default on a credit application, as measured by the ROC AUC metric.

Our model increased ROC AUC score by up to 56 percent compared to our baseline model.

Because our model does not require third party information about clients it could be used as a fast early screening tool to reduce application processing times and provide an easier experience for applicants.

Although our original objective was to rely solely on anonymous data, our model could be improved if third party information and personal data were used such as credit scores and demographic information.