The practical task is organized as a competition between groups of 4 students. Each team represents a bank, that develops their own model for assessment of client's probability of default (**PD-model**) and suggests their **decision-making strategy** to optimize profit and attract clients with lower interest rates. After these two steps banks will score new clients (with hidden default indicator) and offer individual interest rates for them or reject someone at all.

Offered interest rates will be compared. If any client receives approval from several banks, they choose the bank with the best conditions (minimal offered interest rate). If a loan is issued to a good client (non-default), a bank has profit. Similarly, if a loan is issued to a bad (default) client, a bank possibly bares losses. If it turns out that several banks offer the same interest rate and it will be minimal, profit or loss will be divided by the number of such banks equally. Banks will be ranked by total profit on the data with new clients.

Register your team with the list of members on email **alex.dorofeyeva@gmail.com** by **Saturday, 9 Dec 2023, 18:00 Moscow time**.

## PD-model

Each bank has a challenge to develop scoring model (PD-model) to achieve its competitive advantage on a train dataset and make sure for its quality on a test dataset. In such tasks a baseline solution is to use logistic regression with WOE-transformed factors. It is recommended to choose this approach, but you can apply any type of models.

| Data | Submission format |
|---|---|
| • dataset: "data.csv"<br>• description of data: "columns_description.xls"<br><br>All files in this case contain historical data from the huge online lending platform Lending Club. | the Jupyter notebook with a solution, that contains:<br>• train/test split<br>• feature engineering (including WOE and other transformations)<br>• feature selection<br>• model construction<br>• model specification (final factors list, model coefficients, etc.)<br>• evaluation of model quality on train and test datasets (ROC curve, Gini)<br><br>**Warning!** Your Jupyter notebooks must be well organized and easy to understand. Please, provide it with comprehensive comments!<br><br>**Warning!** If you use additional data, include them to your submission files, add a description, referring to the source! |

For WOE-transformation you can use "scorecardpy" Python library or any other. Of course, you can program it yourself.

## Decision-making strategy

Each bank develops its own decision-making strategy and applies it to make a decision about new unknown clients, offering them personal credit terms or rejecting someone at all.

| Data | Submission format |
|---|---|
| • dataset: "data.csv"<br>• description of data: "columns_description.xls"<br>• data for decision-making strategy application: "new_clients_for_scoring.csv"<br><br>All files in this case contain historical data from the huge online lending platform Lending Club. | the file "teamNumber_answer.csv", that contains columns:<br>• "client_id" from file "new_clients_for_scoring.csv"<br>• probability of default "PD"<br>• annual "rate_offered" (if a team wants to reject a client, the "rate_offered" should be empty)<br><br>**Warning!** In "teamNumber_answer.csv" separator must be "," and decimal sign must be ".”! |

Assumptions:

- the bank borrows funds at annual interest rate **f** compounded monthly;

- payments, made by clients, are deposited by the bank to an interest-bearing account with annual risk-free interest rate **r0**;
- term of the deposit for raising funds and term of the loan coincide;
- **r0 = f = 8%**;
- loans can be issued to clients under annual interest rate **r**;
- **EAD * LGD = 0.6 * FV**;
- total calculated PNL on "data.csv" sample must be not less than zero;
- credit terms, requested by clients, are provided in the data.

## Presentation

In the last class each team gives a short (10-15 min) presentation, that must contain at least the following:

- datasets, that you use, and how you create them (train/test split);
- the way, how you create, choose and transform features (feature engineering and feature selection);
- model construction, model quality;
- model specification (final factors list, model coefficients, etc.);
- evaluation of model quality on train and test datasets (ROC curve, Gini);
- the way, how you develop decision-making strategy.

Every team can have one (or more) representative, who will make the report. The presentation file should be named "teamNumber_presentation.pdf".

## Submission and Deadline

Send your solution on email **alex.dorofeyeva@gmail.com**: Jupyter notebook with the model and the decision-making strategy, additional data, if you use them, the file with new scored clients and the presentation file (see the details above) by **Saturday, 16 Dec 2023, 18:00 Moscow time**.

## Assessment

- feature engineering and feature selection: + max 2 points;
- model construction and model specification: + max 2 points;
- evaluation of model quality on train and test datasets: + max 1 point;
- decision-making strategy definition: + max 3 points;
- getting into the top three: + (4 – place).

Final Score = min (Total Points, 10).