Puzzle Overview

As a credit company, it is important to know beforehand who is able to pay their loans and who is not. The goal of this puzzle is to build a statistical/machine learning model to figure out which clients are able to honor their debt.

Structure

Besides these instructions, the ZIP file contains 2 datasets in CSV format:

- puzzle_train_dataset.csv : training dataset
- *puzzle_test_dataset.csv* : test dataset. It contains the same columns as the training dataset, except for the `default` variable

Your goal is to predict the **probability of default** , which is identified by the *default* variable in the training dataset.

Deliverable

You must send us an email with:

- an explanation of what you did (maximum length of 2000 characters with spaces) ;
- your own code, which we should be able to run (you can assume that we have access to the same resources that you used);
- We would prefer a single file structure for your project (as we have a lot of code to review), you also don't have to use container to make your code exactly reproducible
- instructions about how to run the code (e.g., software, programming language, required dependencies, commands to run);
- a *.csv* file called *predictions.csv* with 2 columns: the IDs from *puzzle_test_dataset.zip* and the predicted probabilities under the column *predictions* ;
- the expected performance score of your predictions;

Don't identify yourself in any files! We do blind reviews so people are not supposed to know who you are while evaluating your work.

Please try to follow these instructions as closely as you can.

What we will evaluate

Your ability to:

- Handle/clean data
- Write good quality code (e.g., reproducible, readable)
- Apply machine learning models to real problems
- Split complex and real problems into solvable pieces
- Your explanation on the choices you made

We are giving 3 days to complete this puzzle but you shouldn't spend more than 1 to 5 hours on it