Fintech Lending Game

Jean-Edouard Colliard (HEC Paris)

Johan Hombert (HEC Paris)

November 2023

This business simulation was created by Jean-Edouard Colliard and Johan Hombert for the course "Introduction to Finance for Data Scientist" in the Master for Data Science for Business X-HEC.

Context

You run a fintech lending platform making consumer loans to individuals. You receive loan applications from a large number of individuals and must decide whether and at which interest rate you offer to lend money to each loan applicant. You are in competition with two other lenders (two other teams of students) who also make loan offers to these individuals. Borrowers may default on their loan, so it is important to choose the interest rate at the right level to compensate for the risk of default.

You have information about loan applicants to estimate their risk of default. Some of the information about loan applicants is available to all lenders. Each lender also has its own source of private information about loan applicants extracted from digital footprints. Because different lenders use different algorithms, they obtain different private signals capturing different dimensions of loan applicants' creditworthiness. All lenders also have access to data on past loans to train a credit scoring model.

Data

Past loans. All lenders have access to the same dataset `PastLoans.csv`, which contains information on loans made in the past for which it is known if the borrower eventually defaulted or not. The data include the following variables:

- `sex`: 0= female; 1= male
- `employment`: employment status (four categories)
- `married`: 1= married; 0= other

- `income`: annual income in euro (top coded at 1M euros)
- `digital1`: digital footprints extracted by lender 1 (coded from 0 to 1)
- `digital2`: digital footprints extracted by lender 2 (coded from 0 to 1)
- `digital3`: digital footprints extracted by lender 3 (coded from 0 to 1)
- `default`: 1= the borrower defaulted on the loan; 0= the loan was repaid

Note that past loans data contain the digital footprints variable of all three lenders. The reason might be that the banking regulator mandates lenders to disclose their information after a certain amount of time (this is not the case in practice). Another reason might be that lenders have been hacked and their data leaked in the public domain. No matter what the reason is, the fact is that all lenders have access to the exact same data on past loans to train their credit scoring model.

NB: All these data are simulated data, not real data. However, they are meant to "feel" real in terms of how the different variables are distributed in the population and how they correlate with default.

New loan applications. All lenders receive the same 100,000 new loan applications from new potential clients. Each loan applicant asks for a loan of 10,000 euros. The three lenders do not have the same information set about new applicants. While all lenders have the variables sex, employment, married, and income in their database, each lender only observes its own digital footprints variable. The loan applications are in the file `LoanApplications.csv`.

The new loan applicants are drawn from the exact same population as the borrowers in the first dataset. In particular, the determinants of default are exactly the same in the first dataset and in the new pool of loan applications.

Organization of the Loan Market

Interest rate. Your job is to decide whether and at which rate you make a loan offer to each of these 100,000 loan applications. An offer is an interest rate at which you would be willing to make a loan.

Loan applicants select from which lender they take a loan as follows. Denote the three lenders by $k = 1, 2, 3$ and the loan applicants by $i = 1, \ldots, 100000$. Denote the interest rate offered by lender k to loan applicant i by $r(k, i)$. For example, $r(k, i) = 0.05$ if the interest rate is 5%. If the lender makes no offer to this applicant, we denote $r(k, i) = \infty$. There are three types of loan applicants:

- Type 1 (one-third of the loan applicants): They have a preference for lender 1 and are ready to pay 2% extra to get a loan from lender 1. Formally, they choose the lowest among $r(1, i) - 0.02$; $r(2, i)$; and $r(3, i)$.

- Type 2 (one-third of the loan applicants): They have a preference for lender 2 and are ready to pay 2% extra to get a loan from lender 2. Formally, they choose the lowest among $r(1,i)$; $r(2,i) - 0.02$; and $r(3,i)$.
- Type 3 (one-third of the loan applicants): They have a preference for lender 3 and are ready to pay 2% extra to get a loan from lender 3. Formally, they choose the lowest among $r(1,i)$; $r(2,i)$; and $r(3,i) - 0.02$.

For example, if lender 1 makes a loan offer at 5%, lender 2 at 6%, and lender 3 at 8%:

- A type 1 applicant takes lender 1's offer at 5%.
- A type 2 applicant takes lender 2's offer at 6%.
- A type 3 applicant takes lender 1's offer at 5%.

Therefore, a loan offer you make is not necessarily accepted. It can be rejected because one of your competitors makes a cheaper offer to the same borrower or because the borrower has a preference for another lender. Conversely, some borrowers may accept your offer even if it is not the cheapest.

Payoffs. When a borrower chooses your loan offer, your profit on that loan depends on the interest rate you offered and on whether the borrower defaults or not. Therefore:

- If the borrower does not default, you earn the interest rate you offered times the size of the loan. Your profit on that loan is $r(k,i)$ times 10,000 euros.
- If the borrower defaults, you lose the amount you lent (the recovery rate is zero). Your profit on the loan is negative 10,000 euros.

Your total profit is the sum of the profits and losses you make on all the borrowers who take your offer.

Instructions

Your job is to predict the probability of default of the loan applicants and to decide whether and at which rate to make them offers. Please submit a csv file with the list of the 100,000 loan applications and two columns containing the variables:

- `id`: loan applicant identifier provided in the data set `LoanApplications.csv` (running from 1 to 100,000).
- `rate`: interest rate you offer to the applicant. Please input 0.12 for an interest rate of 12%. You are not allowed to offer interest rates above 100%. If you don't want to make an offer to a loan applicant, leave the interest rate variable missing for this applicant.