

## Data Mining

Teacher: Annamaria Guolo

Practical examination: July 11, 2023

### Dataset loan

Dataset included in `loan.RData` contain information about characteristics of 892 bank customers, applying for a bank loan. The aim of the analysis is the evaluation of the variables associated to the probability that the customer will return the loan.

- `duration`: Duration of the loan (in months)
- `history`: Had the customer repaid previous loans (regular/nonregular)?
- `purpose`: Purpose of the loan application (business/car/education/house/other)
- `amount`: Amount of the loan
- `bankaccount`: Has the customer a bank account? (no/yes)
- `gender`: Gender of the customer (F/M)
- `job`: Has the customer a job? (no/yes)
- `otherloans`: Other loans of the customer
- `property`: Properties of the customer (no/other/realstate)
- `age`: Age of the customer (in years)
- `house`: Ownership or not? (ownership/rent)
- `unemployed`: Is the customer unemployed? (no/yes)
- `maintenance`: The number of people being liable to provide maintenance for (1/ more than 1)
- `foreign`: Is the customer foreign? (no/yes)
- `customer`: Did the customer return the loan or not? (1=yes/0=no)

#### FIRST QUESTION.

Consider the dataset composed by `customer`, `duration`, `amount`, `history`, `purpose`. Build the most suitable model for the purposes of the study. Which variables are associated to the probability that the customer will return the loan?

#### SECOND QUESTION.

Consider all the variables in the dataset. Build the most suitable model for the purposes of the study. Which variables are associated to the probability that the customer will return the loan?