# Intelligent Data Analysis

Exam: Creditworthiness (Project 4)

This project is part of the exam *Intelligent Data Analysis*. Each project assignment is to be resovled by a single student on his/her own. The student is supposed to present the solution as part of the oral exam. The student is required to present a printed version of the Python code together with diagrams, tables, etc. that summarize the results. The specific way of how the project is presented is up to the student's choice.

## Problem setting

A bank wants to predict the creditworthiness of its customers. Based on the customer records, the credit history, etc., a customer should be classified as creditworthy or unworthy of credit. It is five times more 'expensive' for the bank to rate a customer who is unworthy of credit as creditworthy than vice versa. In addition, not all information is available for all customers. For 1,000 representatively selected customers, the creditworthiness is known. For these customers the following data has been collected. (Features for which not all values are known are marked with the addition "incomplete".)

- Status of existing checking account
  - A11: ... $< 0$ EUR
  - A12: $0 \leq ... < 200$ EUR
  - A13: ... $\geq 200$ EUR / salary assignments for at least 1 year
  - A14: no checking account
- Duration in month
- Credit history
  - A30: no credits taken/ all credits paid back duly
  - A31: all credits at this bank paid back duly
  - A32: existing credits paid back duly till now
  - A33: delay in paying off in the past
  - A34: critical account/ other credits existing (not at this bank)
- Purpose (incomplete)
  - A40: car (new)
  - A41: car (used)
  - A42: furniture/equipment
  - A43: radio/television
  - A44: domestic appliances
  - A45: repairs
  - A46: education
  - A47: (vacation - does not exist?)
  - A48: retraining
  - A49: business

  - A410: others

- Credit amount

- Savings account/bonds

  - A61: ... $<$ 100 EUR
  - A62: 100 $\leq$ ... $<$ 500 EUR
  - A63: 500 $\leq$ ... $<$ 1000 EUR
  - A64: .. $\geq$ 1000 EUR
  - A65: unknown/ no savings account

- Present employment since (incomplete)

  - A71: unemployed
  - A72: ... $<$ 1 year
  - A73: 1 $\leq$ ... $<$ 4 years
  - A74: 4 $\leq$ ... $<$ 7 years
  - A75: .. $\geq$ 7 years

- Installment rate in percentage of disposable income

- Personal status and sex

  - A91: male: divorced/separated
  - A92: female: divorced/separated/married
  - A93: male: single
  - A94: male: married/widowed
  - A95: female: single

- Other debtors / guarantors

  - A101: none
  - A102: co-applicant
  - A103: guarantor

- Present residence since

- Property

  - A121: real estate
  - A122: if not A121: building society savings agreement/ life insurance
  - A123: if not A121/A122: car or other, not in attribute 6
  - A124: unknown / no property

- Age in years

- Other installment plans

  - A141: bank
  - A142: stores
  - A143: none

- Housing
    - A151: rent
    - A152: own
    - A153: for free
- Number of existing credits at this bank
- Job (incomplete)
    - A171: unemployed/ unskilled - non-resident
    - A172: unskilled - resident
    - A173: skilled employee / official
    - A174: management/ self-employed/
    - A175: highly qualified employee/ officer
- Number of people being liable to provide maintenance for
- Telephone
    - A191: none
    - A192: yes, registered under the customers name
- Foreign worker (incomplete)
    - A201: yes
    - A202: no
- Creditworthy
    - 1: yes
    - 2: no

You were asked to develop a predictive model that assesses the creditworthiness of future clients. It can be assumed that for these clients as well, there are missing values in the features “Purpose”, “Present employment since”, “Job” and “Foreign worker”.

### Aufgabe

Read the data into Python and proprocess them. Replace missing values ?using? linear regression or classification. Identify a suitable method for solving the prediction problem, implement it in Python, and train the model. Evaluate the model in terms of the bank’s cost model. Briefly motivate and document all the steps you have taken.