University of Padova
Computer Science
a.y. 2021/2022

Data Mining
Teacher: Annamaria Guolo

Practical examination: August 23, 2022

INSTRUCTIONS

Analyze the data described below according to the two questions listed. Prepare a brief report to summarize the main findings of your analysis. It is not necessary to include `R` commands in the report. Please indicate the seed you use for the analysis, if the seed is necessary.

Include your name, surname and enrollment number in the report. Reports without this information will not be corrected. Please, add the date. Print a copy of your report and return it to the teacher.

Dataset included in `phonecall.RData` refers to a study about the probability of purchase by customers contacted through phone calls.

- `age`: Age of the customer (in years)

- `job`: Job of the customer

- `marital`: Marital status (divorced, married, single)

- `housing`: Does the customer intend to ask for a housing loan? (yes/no)

- `loan`: Does the customer intend to ask for a customer loan? (yes/no)

- `phone`: Type of phone of the customer (cellular/landline)

- `month`: The month of the call

- `weekday`: The weekday of the call

- `CPI`: Consumer price index at the time of the call

- `CCI`: Consumer confidence index at the time of the call

- `purchase`: Indicator of purchase (1 for purchase, 0 for no purchase)

- `edu`: Years of education

FIRST QUESTION.

Consider the dataset composed by `purchase`, `job`, `age`, `housing`, `edu`. Construct the most appropriate model for the purpose of the analysis. Which variables are associated to the probability of purchase?

SECOND QUESTION.

Consider all the variables in the dataset. Construct the most appropriate model for the purpose of the analysis. Which variables are associated to the probability of purchase?