

# Statistical Learning

Bachelor in Data Science and Engineering

Second Homework: Supervised Learning

Deadline: December 21, 2025 at 23:59

# General Objectives

- Apply supervised-learning tools to an open data set
- Select one important categorical variable as the target
- Develop all the supervised tools you can (including hyperparameter tuning) to predict the target
- Explain and interpret the relationships between the target and the predictors
- The bigger and more diverse the dataset, the better
- Get insightful conclusions from the performance of the tools, considering ideas from risk learning
- Compare the results using different sets of predictors based on your feature selection analyses

# Evaluation

- ① Data preprocessing and visualization tools: 2 points
- ② Classification (emphasis on prediction): 3 points
- ③ Classification (emphasis on interpretation): 3 points
- ④ Feature Selection and comparison of predictor sets: 2 points

**Important:** you can take inspiration from other notebooks but you need to always cite the source.

## Files to be evaluated

The evaluation process will be given by the following four files:

- **Document** (.pdf). A document with a maximum length of 5 pages that is a report of the work done. The report should contain a **summary** of the whole process, with some graphs, results, R outputs, and references. The conclusions shown in the paper must be reasoned from the perspective of the course.
- **Data**. The data file must be present.
- **Notebook** (.Rmd). The file must be fully executable. Errors will not be corrected. If not executed, the assignment will not be evaluated.
- **Notebook** (.html). The notebook will also be delivered in html format with comments of what the student is doing to verify with the document.

Upload the 4 files in a single zip folder to Aula Global. The **absence** of any file will result in the assignment **not being evaluated**.

# Instructions

- The assignment has to be done in pairs.
- The assignment will be uploaded by both members of the pair.
- The evaluation will be the same for both members.
- **No assignment will be accepted after the deadline.**