Instituto Superior
de Engenharia
Politécnico de Coimbra

Practical Evaluation – M1 [15% of final grade]

Machine Learning /1st year MEI

2025/26

1. Introduction

The first milestone (M1) of practical assessment aims to produce a systematic and organized approach to the preparation of a dataset for its subsequent submission to Machine Learning algorithms.

Thus, the selection of the dataset should take into account that M1 will support the development of the next step: M2 (supervised learning; unsupervised learning).

2. Dataset Selection

The selection of the dataset must be carried out considering that it must:

- contain real data[1];
- set up a **binary classification** problem;
- have between 8 and 30 features;
- have a dimension that allows the constitution of a training set and a test set with a sufficient number of instances to ensure a consistent validation. Ideally, the total number of instances must be greater than $N \geq 1000$;
- have at least 1 attribute of each type (numerical, categorical, ordinal).

Taking into account the established requirements, students are free to search and select the dataset.

[1] In this context, the date of data acquisition is not relevant.

3. Milestone M1

This assessment must be implemented in accordance with the following requirements:

- A complete description of the data must be made, as well as the corresponding visualization;
- The data must be prepared to be submitted to the Machine Learning algorithms (applied in M2 step);
    - All data cleaning/transformation/coding operations must be properly explained;
    - Data transformation pipelines must be implemented.

4. Evaluation

This evaluation should observe the following rules:

- groups are formed by 2 students;
- a short presentation (ppt; maximum 10 minutes) must be prepared to be presented on 11th and 12th of November 2025. The two elements of the group must have an equivalent participation during the presentation;
- a small document must be prepared in order to describe the followed approach. The document must be as objective and condensed as possible (max. 10 pages);
- Jupyter Notebook based implementation should also be uploaded;
- this evaluation goal represents 3 values (15%) of the final grade;
- A short form will be made available with the several slots for carrying out the presentation;
- all assessment elements (*.ipynb; *.csv; *.doc; *.ppt) must be delivered on the nonio platform (inforestudante), in a compressed file (.zip), by 11:59 PM on 9th of November 2025, with the following nomenclature:

firstLastNameStudent1_studentNumber1_firstLastNameStudent2_studentNumber2.zip

Example: SamAltman_1231234_DemisHassabis_1234123