Project Directives

Machine Learning 2025 Project Guidelines

The project grade accounts for 70% of the final grade: 30% for the presentation and 40% for the report. The remaining 30% are for the exam. Grades are from 1 to 6 (Absent=0), rounded to the closest 0.1. The final grade (40% report + 30% presentation + 30% exam) is rounded to the closest 0.5.

Organization of Groups

The size of groups is 3 students. You may now register via our moodle page. Because we do not know the final number of participants yet, we may create groups of 4 by the end of the grouping process. By default, only groups of 3 can register for the moment.

Content

Your work must match the structure of analysis and of reporting similar to a scientific paper using machine learning. To guide you, 10 scientific papers have been selected as examples of found here . Note that these papers **are not about developing new machine learning techniques** but rather about predictions in some domain using machine learning techniques.

For this work, you must find an interesting case (scientific domain of application), make an extensive literature review about the use of machine learning in this context, find appropriate data set(s), define appropriate methodologies of analysis, apply them, and report the results in a scientific paper

The paper must contain

- An abstract: summary of the paper including the main results.
- An introduction: context and objective of the study.
- A literature review: state of the art, previous works, and appropriate citations.
- A methodology exposure: methods and models that are used in the paper.
- Results: all the interesting results under the shape of tables, figures, interpretations.

- Conclusion: limitations, discussion, and outlook subsections.
- Appendix: extra elements from the paper body.
- References: in an appropriate style (e.g., APA).

The project must cover methods seen in class such as:

- Cleaning of the data, creation of new features if needed.
- Exploratory Data Analysis.
- Supervised learning analysis: several models, appropriate metrics, a tuning of hyperparameters, data splitting strategies, interpretation of the model.
- Unsupervised analysis: clustering and/or dimension reduction.

Deliverables

- Report: a pdf, between 10 to 20 pages (not including appendix or supplementary material)
- Presentation: the slides of your presentation.

Deadlines

You must meet the following deadlines:

- Project report: Wednesday the 21st of May 2025 at 23h59
- Presentation slide: Sunday the 25th of May 2025 at 23h59

Not meeting the deadline (presentation and/or report) penalizes the grade by 0.1 per started hour of delay. No maximum penalty for the project report and presentation slide.

Presentations

Presentations will be organized on site on Monday the 26th of May 2025 in a 15+10-minutes format (presentation + questions). If the number of groups is large, then either the presentation time will be reduced, or, if not possible, another oral presentation session will be organized later (possibly during the exam session).