

Individual Project Assignment

General overview: The goal of the project is to demonstrate your understanding of the course content and prove your skills in implementing basic machine learning models. The project will require to analyze a given dataset using machine learning techniques, evaluate the trained models and critically discuss the results.

Technical notes: The project should be implemented in Python using the scikit-learn framework. You can explore additional libraries and software tools; however, please make sure you understand the code you are writing because during the oral exam you might be asked specific questions related to your implementation. The Python notebook must be uploaded in a Moodle section dedicated to the project assignment at least 10 days before the chosen exam date. Make sure you properly comment the code and the results using “text blocks” in the notebook.

Datasets: At least two different real datasets will be made available through the course Moodle (CSV format). Each dataset will include a classification task and a regression task. You can also consider using a different dataset of your choice: in that case, please get in touch with the professor to discuss it.

Grade: The maximum grade for the individual project will be **20 points**. The oral exam will then contribute to the final grade with other additional 10 points. The *laude* is reserved for students that implement outstanding projects and will be able to answer more challenging questions during the oral exam.

Points will be assigned based on the correct implementation of the following tasks:

- **5 points:** Load the dataset, exploit data visualization techniques (e.g., histograms) to understand the distribution of input and output variables, normalize/standardize the input variables (if required), use a weighted loss function (if required), split the data into train/validation/test sets.
- **5 points:** Implement proper regression/classification models and justify your model choices.
- **5 points:** Tune the models' hyperparameters using grid/random search and cross-validation.
- **5 points:** Evaluate and compare the models using proper evaluation metrics (e.g., accuracy, confusion matrices, ROC curves, F1 score) and produce informative plots to visualize your results.

FAQs:

- 1) Do I need to submit a report describing the project results?

No: the advantage of using Python notebooks is that the code can be integrated with figures and textual descriptions. This dispenses from the need to submit a PDF report, however please make sure you properly describe your implementation choices and comment the results.

- 2) Is it possible to work on the project in pairs / groups?

No: in order to give you a representative grade, I need to evaluate your individual contribution and that would be difficult in a collaborative setting. You are encouraged to consult with your colleagues during project development, but please do not copy their Python code. Plagiarism will be negatively evaluated.

- 3) Is it possible to resubmit the project several times?

Yes, but you will receive only one evaluation per exam call. You can resubmit the project to improve your grade, but in that case you will also need to retake the oral exam.

- 4) What should I do if I encounter problems during project implementation?

You could ask support to your colleagues or seek advice through the Moodle Forum. You are also strongly encouraged to look for solutions online (search engines are a great source of information!). As a last resort, you could also seek advice from the Teacher Assistant, Sina Shafieezadeh (sina.shafieezadeh@studenti.unipd.it).