

Instituto Superior de Agronomia, ULisboa

Master's in Green Data Science 2022-2023

Practical Machine Learning/Aprendizagem Automática Aplicada

Instructor: Manuel Campagnolo

## Final project guidelines

### Project proposal (due May 5, 2023)

Your project proposal should include the following information:

- What is the problem that you will be investigating? Why is it interesting?
- What are the challenges of this project?
- What dataset are you using? How do you plan to collect it? You can use your own data, if you have, or gather data from online data repositories.
- What method or algorithm are you proposing?
- How will you evaluate your results? What kind of analysis will you use to evaluate and/or compare your results (e.g. what performance metrics or statistical tests)?

Format of the project proposal: Your proposal should be a PDF document, giving the title of the project, the project category (tabular data, image classification, image segmentation, other—to specify), the full names and student IDs of team members (1-2 team members), and a 300-500 word description of what you plan to do.

Submission (due June 9, 2023): Create a GitHub repository which contains your report, and a separate notebook or script with the code. In alternative, you can create a notebook that combines both. Send the repository's URL to [mlc@isa.ulisboa.pt](mailto:mlc@isa.ulisboa.pt)

1. The report should contain a comprehensive account of your project. The report is expected to be thorough, yet concise, and organized by sections like "introduction", "data", "methods", "results", "discussion and conclusions", and "references". Broadly, it should cover the following:
  - Good motivation for the project and an explanation of the problem statement (you can re-use the project proposal);
  - A description of the data. It is possible that your data is not ready to be used, due to data errors, inconsistencies and formats. Perform and document the necessary data cleaning and transformation steps to prepare your data. Also identify the data types. Describe Data Cleaning, Feature Selection and Feature Engineering;
  - Describe training, validation and test sets and how those are stored in disk.
  - Which ML model or models are used to address the problem: describe hyperparameter and architecture choices that were explored
  - Presentation of results: either tabular or graphical

- Analysis of results
- Any insights and discussions relevant to the project
- References
- Contributions: include a section that describes what each team member worked on and contributed to the project.

Format: A ~4-6 pages document, which extra pages for appendices and references is needed (but the main document should be self-contained)

2. Code: Python notebook or script with the code.

Grading (up to 10): The final report will be judged based off of the clarity of the report, the relevance of the project to topics taught in class, the novelty of the problem, and the technical quality and significance of the work.

Useful links:

- The Kaggle Machine Learning Project Template  
<https://www.kaggle.com/general/187601>