

Data Mining and Machine Learning

Project Guidelines

Academic Year 2024-2025

José Luis Corcuera Bárcena – Francesco Marcelloni

Outline

- Assessment method
- Project development steps
- Timeline
- Project proposal template



Assessment method

- Students must carry out a DMML project developed in Python
- The project can be carried out **in group (up to 2 students)** or **individually**
- In case of «group project», both students must participate in the project discussion
- The exam will consist of:
 - 15-minute presentation of the project
 - Question(s) about the project: bring your own device since we can ask you to show some aspects of your project
 - Question(s) about theoretical aspects
- It is possible to discuss the project and take the oral part of the exam in different exam sessions (anyway, you can take part in the oral part only after a positive evaluation of your project)



Project development steps

- Identify a subject of your interest and motivate the relevance of DMML techniques
- The project must deal with the topics covered during the lectures; in addition (not as replacement), you can explore topics related to DMML not covered during the lectures
- Demonstrate mastery of DMML, e.g.:
 - Exploratory analysis
 - Data cleaning and data preprocessing
 - Feature extraction, feature selection
 - Classification / regression / clustering / outlier detection / pattern analysis
 - Model selection and performance evaluation
- Adequately report each step and discuss the results obtained
- A comparison with the *state of the art* would be desirable
 - e.g., critically review at least one scientific work that uses the same dataset, if any



Timeline

- **Step 1: Project proposal (see next slides)**
 - Submit your proposal by filling the form <https://forms.gle/rbm1s5Pr8AXAfmTA9>, at least **one month before the exam** session which you intend to take
- **Step 2: Material submission**
 - The whole material (codebase, pptx, report/user guide) must be submitted via e-mail (or as a Google Drive folder, or github repository) to jose.corcuera@ing.unipi.it
 - The material must be submitted few days (typically around 5) before the date of the exam. Details will be provided timely as notes in the exam portal



Project proposal template

Slide 1/3: Problem description

- *Brief description of the problem*
- *Relevance of DMML techniques*
- *Brief description of how you intend to face the problem (classification, clustering, pattern analysis, outlier detection, etc.)*
- ...



Slide 2/3: Dataset description

- *Procedure of data collection if you made, or plan to make*
- *Link to the source, if any*
- *Properties of the dataset (raw data features, dataset size, ...)*
- *Description of input and output formats*
- ...



Slide 3/3: References (if any)

- *Bibliography / references / related works*
- ...

