

**DATA MINING AND MACHINE LEARNING**  
**Academic Year: 2020/2021**  
**MSc in Artificial Intelligence and Data Engineering**  
**Project Guidelines**

1. Each student must design and develop an application which embeds one or more machine learning algorithms discussed during the course.
2. The project can be carried out even in collaboration with another student.
3. The design of the application, in terms of requirements, basic use case diagrams and simple UML class analysis should be carried out.
4. An appropriate reference dataset must be chosen. This dataset should be used for the design of the model(s) that will be embedded in the application.
5. The design of the model(s) must include all the classical stages of KDD (data pre-processing, attribute selection, model learning and validation).
6. Usually, different approaches for data pre-processing (attribute selection, class-rebalancing and so on) and for learning descriptive/predictive models should be used. For example, if the application will deal with a classification task, different classifiers and attribute selection algorithms should be experimented for selecting the most suitable model. An analysis using the cross-validation methodology should be carried out.
7. Some examples of typical usage of the application must be discussed.
8. The teachers are available for discussing the requirements of the application before starting with the implementation itself.
9. The students will discuss the project during the day specified by the teachers (before the official examination day).
10. A presentation (pptx or pdf) which summarizes all the stages of the project must be prepared by the student(s).
11. At the end of the project, the final documentation and the presentation must be uploaded on Google Classroom. All the other artifacts (code, database dump and executable files) must be uploaded on a git server (provide the server address in the documentation).
12. If the project has been carried out in collaboration with another student, just one student must submit the materials using the eLearning platform.

Notice that, if a different agreement has been stated between teachers and students, the design and the implementation of an application can be skipped. This is the case, for example, in which a specific dataset has been identified (and discussed with the teachers) and a deep KDD process must be carried out. In any case, the presentation must discuss all the stages carried out during the student work.