Project Guide 02

Artificial Intelligence (AI), 2023-24

Degree on Computer Systems Engineering

Introduction

- This document presents the goal and the requirements for the development of project 02

- The goal of the project is to **implement distinct Machine Learning (ML) approaches and methods** to address a specific problem using a public dataset

- The work groups should be composed of **two or three students**

- The project should use a **single dataset** for classification, clustering and association rules

- If the work group finds that it cannot get any interesting results for the association rules, a distinct dataset can be used just for that task

GitHub Platform

- The project must be available in a **GitHub repository**

- Each student must create an account on the GitHub platform
  - One of the members of the group must create a private repository
  - Use the nomenclature "**IA23_G##**", where G## corresponds to the group
  - Add the other members of the group as collaborators
- The lecturer/teacher should also be added to the repository and the project
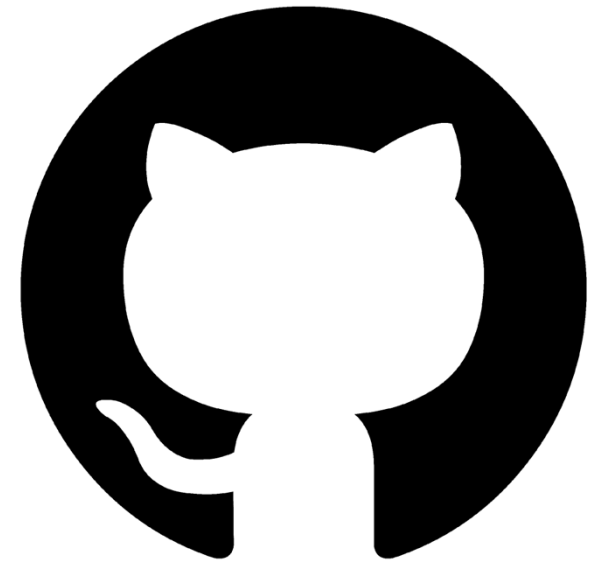
Project submission

- The **Github repository** must contain all files required to run the solution

- The code and the documentation must be merged into a **Jupyter Notebook**

- A **ZIP file** of the repository must be submitted on Moodle before the deadline

- The **link** for the repository must be included in the first markdown block of the Jupyter Notebook

- The project will be later presented and defended by the work group

Jupyter Notebook

Notebook structure

- **Introduction**
  - Identify the teammates: student name and number
  - Establish here the context and the purpose of project 02
  - The structure of the notebook can be adapted according to each project characteristics

- **Dataset**(s)
  - Provide the source of the used dataset(s)
  - Present a brief description
  - Include the dataset metadata

  - You can find many public datasets in https://www.kaggle.com/datasets

Notebook structure (2)

- **Automatic classification**
  - Define the business goals to be achieved
  - Select two algorithms and the parameters to be used
  - Present the data selection criteria
  - Explain how the data was prepared
  - Apply the ML algorithms and evaluate the generate models
  - Optimize the selected algorithm
  - Document the intermediate and final results

Notebook structure (3)

- **Clustering**
  - Define the business goal to be achieved
  - Present the data selection criteria
  - Explain how the data was prepared
  - Apply and evaluate the K-Means algorithm
  - Optimize the algorithm parameters
  - Document the intermediate and final results

Notebook structure (4)

- **Association rules**
  - Define the business goal to be achieved
  - Present the data selection criteria
  - Describe the data preparation steps
  - Apply and evaluate the Apriori algorithm, adjusting the algorithm parameters
  - Document the intermediate and final results

- **Conclusion**
  - Present an analysis of the results obtained in the previous sections based on performance metrics
  - Highlight the lessons learned from the execution of the project

Thank you!