# Project Guide 02

Artificial Intelligence (AI), 2025-26

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 fo the maximum of **five students**

- The project should use only **one or two datasets** 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

- The presentation and defense of the project will take place on a date to be scheduled later.
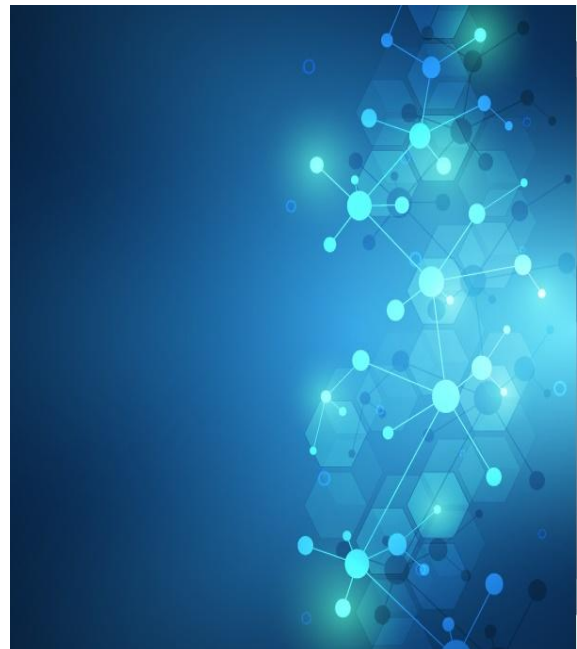
# Collaborative Platforms

The project should be hosted in a private **GitHub** repository
- Use the nomenclature "**IA25_P02_G##**", where G## corresponds to the group

- The lecturer must be added to the repository

The project must be submitted as a ZIP file labelled **IA_P02_G##.zip**, where G## corresponds to the group

The file must contain
- Jupyter notebook;
- readme.txt file with the link to the repository (Github);
- other files needed to run the notebook.
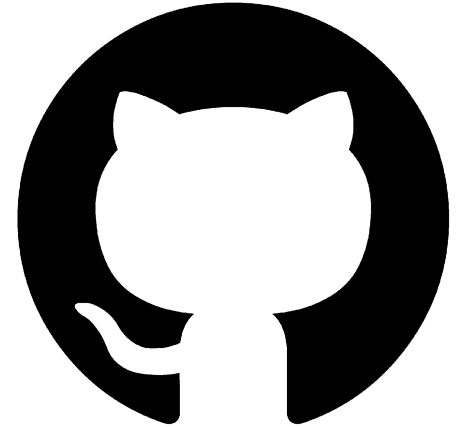
# Jupyter Notebook

# 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 an exploratory data analysis (EDA) for each dataset
- Include the dataset metadata

# Notebook structure (2)

**Dataset**(s)

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

- You should select data sets related with "**sport**"

Source: https://br.freepik.com/vetores-gratis

# Notebook structure (3)

**Automatic classification – notebook 1**

- Define the business goals to be achieved

- Select two or three algorithms

- Present the data selection criteria

- Explain how the data was prepared

- Apply the ML algorithms and evaluate the generate models

- Optimize the selected algorithm adjusting the hyperparameters

- Document the intermediate and final results

Source: https://www.elastic.co/guide/en/machine-learning/current/ml-dfa-classification.html
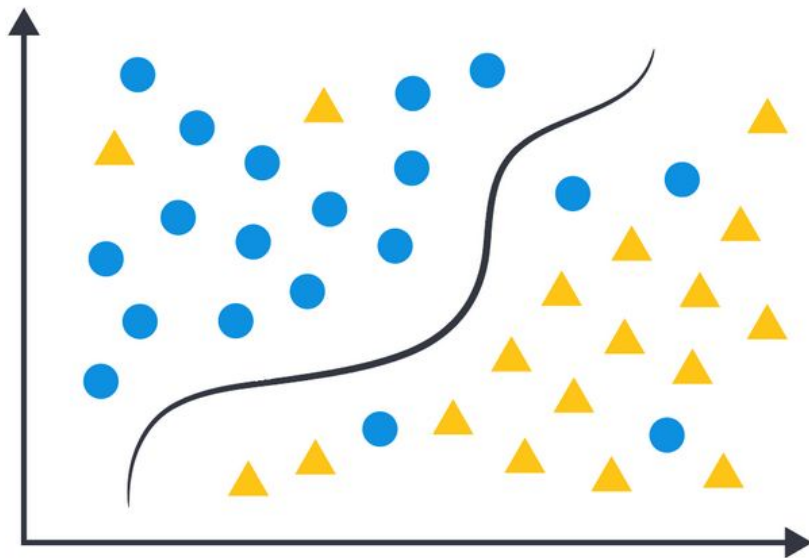
# Notebook structure (4)

**Clustering – notebook 2**

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

- It is important to analyze the characteristics of each cluster
  - Number of elements in each cluster
  - Average, standard deviation, max and min of each numerical attribute
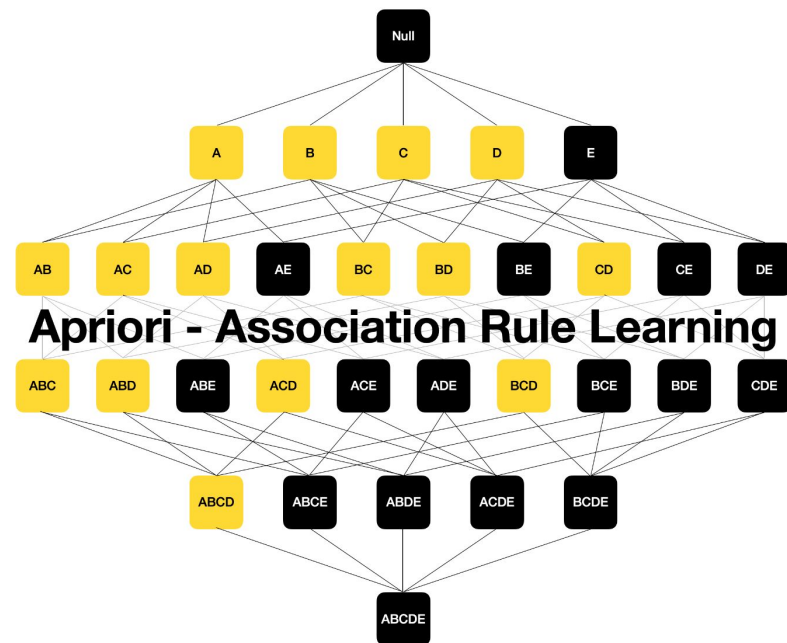  - Distribution of categorical attributes
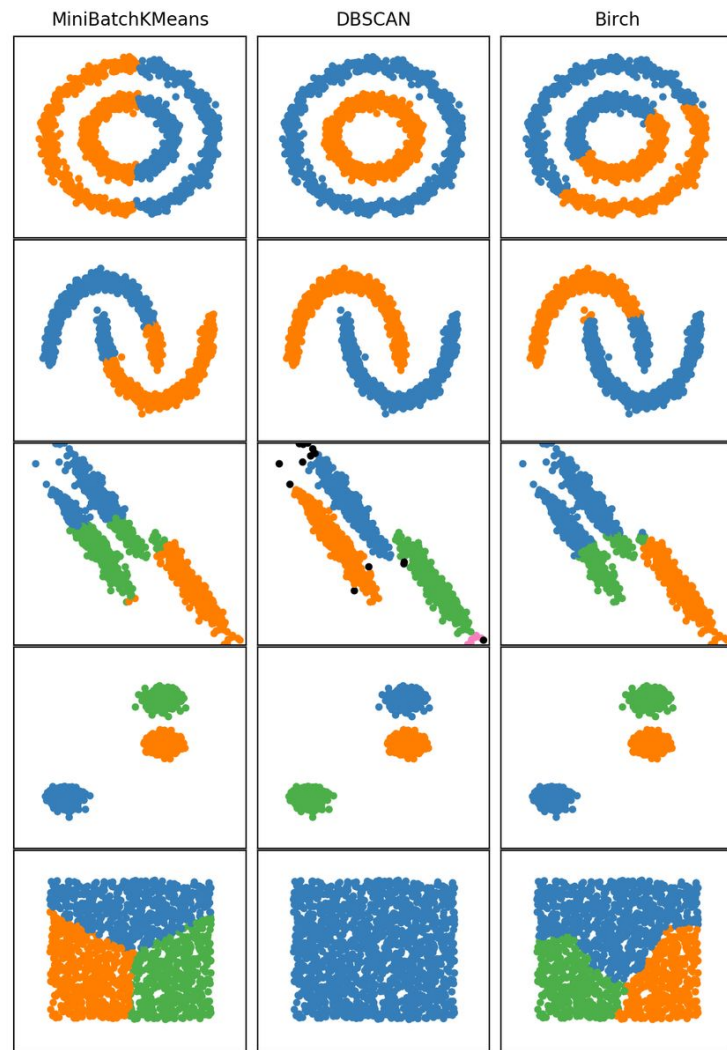
# Notebook structure (5)

- **Association rules – notebook 3**
  - Define the business goal to be achieved
  - Describe the data preparation steps
  - Apply and evaluate the Apriori algorithm, adjusting the algorithm parameters
  - Document the intermediate and final results

- **Conclusion**
  - In each notebook, pPresent an analysis of the results obtained based on performance metrics
  - Highlight the lessons learned from the execution of the project



Source:_https://towardsdatascience.com/apriori-algorithm-for-association-rule-learning-how-to-find-clear-links-between-transactions-bf7ebc22cf0a

Thank you!