



# Data Science & Machine Learning for Engineering Applications

## *Group Projects*



Data Base and Data Mining Group of Politecnico di Torino



2022-2023 - *Politecnico di Torino*



# Final Projects

- Activities to perform
  - Choose a use case from the four suggested cases
  - Characterize the dataset in terms of data distributions (i.e., analyze the dataset) and synthesize with plots
  - Perform multiple analysis sessions of the data set using Python
  - Analyze the results and summarize them in graphs
  - Discuss how the findings can be used in a specific technical/engineering application
  - Write a scientific report on the activities performed
    - Use the Latex or Word templates available on the course website
    - Adhere to the maximum number of pages (4 pages maximum)
    - Contribution section is mandatory



# Structure of the scientific report

- Abstract
  - The abstract summarizes what the report presents briefly.
- I. INTRODUCTION
  - Introduce the problem and briefly present the data.
- II. DATA CHARACTERIZATION
  - Characterize data distributions through plots and textual description.
- III. METHODOLOGY
  - Describe the data science pipeline by specifying briefly why the algorithms were selected and configured, Discuss the choices made for each data analytics activity
- IV. EXPERIMENTAL RESULTS
  - Present the results obtained, highlighting the strengths and weaknesses of the various proposals. If deemed useful, also report practical examples to facilitate understanding. State the parameters used for the results presented.
- V. CONCLUSION
  - Provide conclusions and state how the findings can be used in a specific technical/engineering application
- VI. CONTRIBUTIONS

State personal contributions

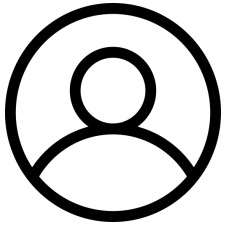


# Groups

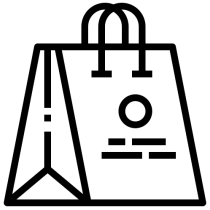
- Definition of the group
  - Fill out the Google form at <https://forms.gle/cYTLurXg1dsaKCMf8> by May 12, 2023
    - A team of 5 students
    - All fields must be filled
    - 1x per group
  - Choice of use cases
    - First choice
    - Second choice
  - Group member information
    - Student ID, first name, last name, master program
  - Assignment of use cases to groups will be announced by May 13, 2023



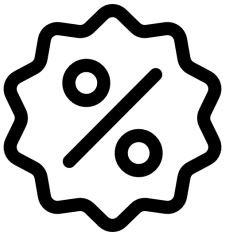
# Use Case 1: Clustering



Personal data



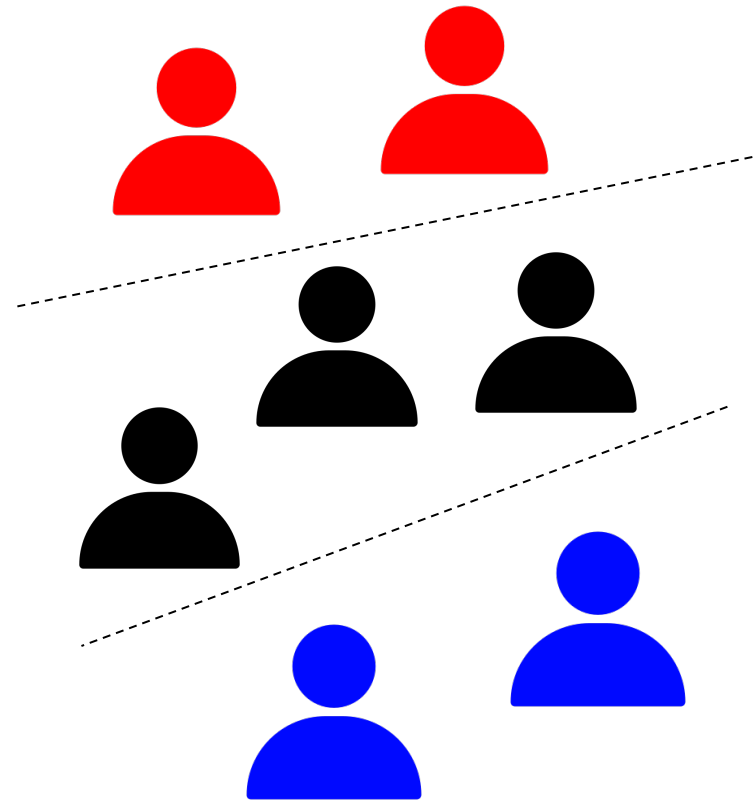
Products



Discounts

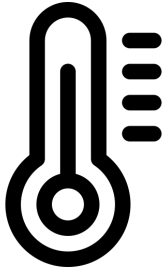


Purchasing channels





# Use Case 2: Binary Classification



Temperature



Torque

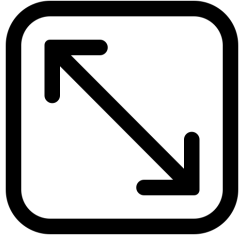


Rotational speed

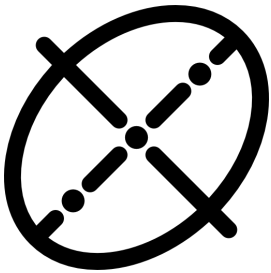




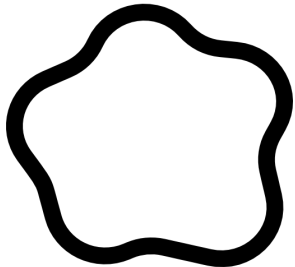
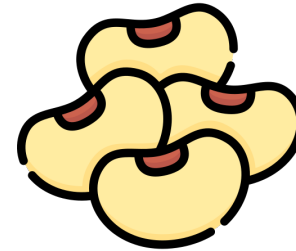
# Use Case 3: Multiclass Classification



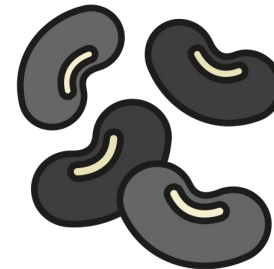
Area



Axle Length

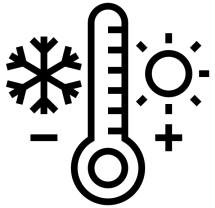


Shapes





# Use Case 4: Regression



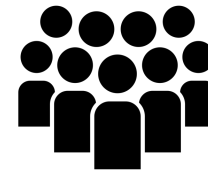
Temperature



Weather



Calendar







# Tutoring

## ■ Time slots

- Every Monday since May 15th, 2023
- Time slot 10:00-11:30 in Room 8C
- Time slot 14:30-16:00 in Room 5M

## ■ Tutors

- Salvatore Greco
- Jacopo Fior



# Submission details

## ■ Each group must submit

- The scientific report
- The Python code (e.g., the notebook)
  - The notebook should contain explanatory texts and comments
- A zip archive must be uploaded to a Dropbox folder (specific for group projects)

<https://www.dropbox.com/request/kIJp3zIKfp1w0fj5TEot>

## ■ Submission deadline

- Within 7 days before each date of the written test



# Project evaluation

- The evaluation of the group project is based on the performance and accuracy of the proposed solution, in terms of standard quality measures (e.g., prediction accuracy, f1-score), and completeness (i.e., in depth analysis of each phase of the designed process and motivation for selecting given techniques and algorithms).
- Each student in the group will be scored
  - Score from zero to twenty.
- The individual evaluation of the group project is valid until the exam session of January 2024 (included)