

Data Mining

Course Overview

Anna Monreale & Mattia Setzu
Computer Science Department
a.y. 2024 - 2025

Data Mining

Data Mining (9 CFU) Code: 309AA

- for students in **Computer Science**
- Students from Data Science & BI and Informatica Umanistica (Data Mining: 420AA)

Information

- **Anna Monreale**
 - KDDLab, Università di Pisa
 - ✉ [anna \[dot\] monreale \[at\] unipi \[dot\] it](mailto:anna.monreale@unipi.it)
- **Mattia Setzu**
 - KDDLab, Università di Pisa
 - ✉ [mattia \[dot\] setzu \[at\] unipi \[dot\] it](mailto:mattia.setzu@unipi.it)

Teaching Assistant:

- * **Lorenzo Mannocci**
 - University of Pisa
 - ✉ [lorenzo \[dot\] mannocci \[at\] phd \[dot\] unipi \[dot\] it](mailto:lorenzo.mannocci@phd.unipi.it)

Day of Week	Hour	Room
Tuesday	11:00 - 13:00	Room C1
Thursday	09:00 - 11:00	Room C
Friday	09:00 - 11:00	Room C1

Data Mining

- **Web Page**
 - <http://didawiki.cli.di.unipi.it/doku.php/magistraleinformatica/dmi/start>
- **Learning Material:**

Title	Authors	Edition
Introduction to Data Mining	Pang-Ning Tan, Michael Steinbach, Vipin Kumar	2nd
Introduction to Data Science: A Python Approach to Concepts, Techniques and Applications	Laura Igual, Santi Seguí	2nd
Python Data Science Handbook: Essential Tools for Working with Data	Jake VanderPlas	1st
Deep Learning	Ian Goodfellow, Yoshua Bengio, Aaron Courville	
Introduction to Linear Algebra	Gilbert Strang	5th

- **Slides of the lectures, Python Notebooks**
- **Data mining software**
 - **Scikit-learn**: python library with tools for data mining and data analysis

Exam

- **Project**

- Given a set of data students will address the tasks required
- Groups of 3 people (**to be defined within the end of Sept**)
- The dataset will be released during the first week of October
- **Two deadlines:** the first one around the **mid-November** and the second one **before the end of the year.**
- **To be delivered:** source code, report on the analysis

- **Paper presentations:**

- Papers made available by the teacher
- Presentation in December during the classes
- Students need to answer questions from teacher and other students
- It is not mandatory

- **Oral**

- Project presentation (with slides) – 15 minutes
- Open Questions on the details of project and on the parts of the program that are not covered by the project
- Students who decide to skip the paper presentation need to answer questions on the whole program

Mailing list

- We will publish in Teams a shared document where each student need to provide the following data:
 - name
 - surname
 - student-id (if you have one)
 - Email