

Data Mining

Course Overview

Anna Monreale & Lorenzo Mannocci

Computer Science Department

a.y. 2025 - 2026

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)

- **Lorenzo Mannocci**

- KDDLab, Università di Pisa
- ✉ [lorenzo \[dot\] mannocci \[at\] di \[dot\] unipi \[dot\] it](mailto:lorenzo.mannocci@di.unipi.it)

Day of Week	Hour	Room
Monday	11:00 - 13:00	Room C
Tuesday	14:00 - 16:00	Room C1
Thursday	09:00 - 11:00	Room C

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 Track**
 - 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 by the first week of October
 - **One Presentation per project** for showing the ongoing work **by the end of Nov**
 - **Deadlines:** before the end of the year
 - **To be delivered:** source code, report on the analysis
- **Written test track:**
 - Written exam to be delivered during the standard exam sessions and can include both theoretical questions and exercises.
- **Oral**
 - Project presentation (with slides) – 15 minutes and open questions on the details of project to finalize the project score
 - Open questions on the program
 - Students who decide to skip the project 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