Data Mining
Course Overview

Anna Monreale

Computer Science Department

a.y. 2022 - 2023

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

- **Instructors**
  - Anna Monreale
    - Computer Science Department, Room 374/DO
    - Email: anna.monreale@unipi.it
    - Office hours: **Tue 11:00-13:00 -** Online using Teams or in my Office (Appointment by email).
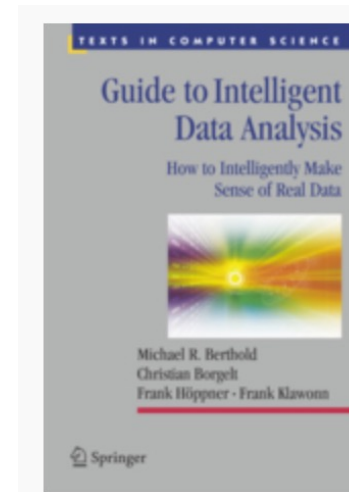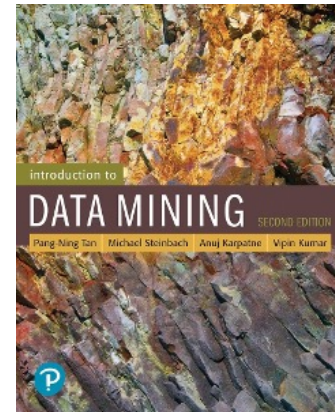- **Timetable**
  - Wednesday   09:00 - 11:00   Room E
  - Thursday   11:00 - 13:00   Room C
  - Friday   09:00 - 11:00   Room C

Data Mining

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

- **Learning Material:**
  - **Textbook:** Pang-Ning Tan, Michael Steinbach, Vipin Kumar
    Introduction to DATA MINING Addison Wesley, 2018
  - **Textbook**: Berthold, M.R., Borgelt, C., Höppner, F., Klawonn, F**. GUIDE TO INTELLIGENT DATA ANALYSIS.** Springer Verlag, 1st Edition., 2010.

- **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 **end 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 program and on the details of project
  - Students who decide to skip the paper presentation need to answer questions on the whole program

UNIVERSITÀ DI PISA

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

Lab of Data Science

UNIVERSITÀ DI PISA