LABORATORY OF DATA SCIENCE

Anna Monreale & Roberto Pellungrini

Computer Science Department, University of Pisa

Data Science & Business Informatics Degree

Room, timetable, teachers

- **Timetable**
  - Monday, 11-12:45
  - Thursday , 09-10:45

- **Teachers**

  - Anna Monreale
    - Computer Science Department, Room 374/DO
    - Email: anna.monreale@unipi.it
    - Office hours: **Wed 11:00-13:00**

  - Roberto Pellungrini
    - Computer Science Department, Room 384/DO
    - Email: roberto.pellungrini@di.unipi.it
    - Office hours: **Thursday 14:00 – 16:00**

Lab of Data Science

Context

- The DS Lab
  - is at the 1st semester, 2 year
  - is **compulsory** for Business Informatics
  - is a **lab course** of topics presented at:
    - 667AA: Programming for data science
    - 600AA: Decision Support Databases
    - 420AA: Data Mining I & II
  - <u>requires</u> knowledge of
    - computer programming
    - datawarehousing
    - data mining

Module objectives

- The DS lab presents technologies and systems for accessing heterogeneous **data sources**, building and populating **datawarehouses**, querying data **cubes**, extracting knowledge by **data mining**

- The accent is on **problem solving** and on the **technology** in support of the process. _Theory is assumed from previous courses._

- Contents of the DS lab will demonstrate useful for the thesis or in industrial projects you will do in BI.

Teaching materials

- Web page
  - http://didawiki.di.unipi.it/doku.php/mds/lbi/
  - news and links
  - slides & video-records of lessons
  - textbook chapters
- Software installed on PCs or can be downloaded
  - Anaconda 3.5 for programming (Python 3.7)
  - SQL Server 2019 for data warehousing
    - only client & dev tools needed
    - optional, server downloadable with MSDN-AA licence

Exam

- **Project**
  - Given a set of data students will address the tasks required
  - Groups of 2 people
  - The dataset will be released during the first week of October
  - **Two dealines:** the first one around **mid-november** and the second one **before the end of the year.**

- **Oral**
  - Project presentation (with slides)
  - Open Questions on the program and on the details of project

Accounts & Mailing list

- We will publish in Teams a shared document where each student need to provide the following data:
  - name
  - surname
  - student-id
  - email

- You will get:
  - an account on a database server of the course
  - included a mailing list of urgent news