00 - Introduction

Data Science and Management

Corso di Laurea Magistrale in Ingegneria Gestionale

Marco Mamei, Natalia Hadjidimitriou, Fabio D'Andreagiovanni, Matteo Martinelli

{marco.mamei, selini, fabio.dandreagiovanni, matteo.martinelli}@unimore.it

- Course Program
- Course Organization
- Exam
- Software
- Books
- Calendar

Course Programme (1/2)

Data Science

- Some statistics background
- Data exploration and visualization
- Machine learning and data mining
- Python
- ML and Optimization

Course Programme (2/2)

Data Management
- Data warehouses
- Big data
- Postgress
- ERP

Course Material

YouTube Channels for last years

- https://www.youtube.com/playlist?list=PLm63_RCCYVFt2PymuODsGG555FkFM2dLw

Course Organization

Class lectures and "virtual lab" sessions
- **All the teaching material available on Moodle**

Schedule
- Tuesday, 14:00-18:00
- Thursday, 14:00-17:00 / 18:00

Course Organization

Additional notes
- Total amount of hours: **81**
- The course should finish **by mid December**

Course Organization

Office hours

- Online, using Teams or Google Meet
- Please send me an e-mail in advance
- You can also ask me (quick) questions via e-mail

Course Organization

Please **subscribe** to the team **and** Moodle website!
- I will send communications through the platform
- Changes in lecturing timeline, new slides, etc…

2024 – Marco Mamei, Full Professor (marco.mamei@unimore.it)

Exam

The exam will consist of **two parts**

- Project (weight ~50%)
- Oral exam (weight ~50%)

Exam

About the project (more info throughout the course):
● You can work in teams of **up to 3 people**
● We will assign a project to each student / team, or you can propose a topic
● The project requires:
  ○ **Code**: implementation or experimentation
  ○ **Report:** 4-pages document in english presenting the problem, the solutions and the results. It is important to **motivate** the choices you made
  ○ **Presentation:** slides for 10 mins project presentation

Oral exam will start with project discussion followed by questions about course's topics

Dates
● Dates to be confirmed…
● 3 January/February + 2 June + 1 September

Software

For the data science part, we will use Python 3+

Available for Windows, Linux, MacOSX

You can use any IDE you like: during the course - I will use VS Code.

https://code.visualstudio.com/download

https://www.jetbrains.com/pycharm/ (community)

https://jupyter.org/

For the data management part of the course, we will use Postgress

https://www.postgresql.org/

Books

- J. VanderPlas. Python Data Science Handbook, O'Reily

- A. Géron. Hands-On Machine Learning with Scikit-Learn, Keras, and Tensorflow: Concepts, Tools, and Techniques to Build Intelligent Systems, O'Reily

- R. Kimball, M. Ross. The Data Warehouse Toolkit, Third Edition, Wiley

- M. Golfarelli, S. Rizzi. Data Warehouse: Teoria e Pratica della Progettazione, Seconda Edizione. McGraw-Hill

Tentative Caledar

Data	Argomento	Ore	Docente
17-set-24	Introduction + DS Introduction	3	Marco Mamei
19-set-24	recap pythyon, numpy, panads	4	Matteo Martinelli
24-set-24	Statistics	4	Natalia Hadjidimitriou
26-set-24	recap pythyon, numpy, panads	4	Matteo Martinelli
01-ott-24	Statistics	4	Natalia Hadjidimitriou
03-ott-24	Data Visualization + matplotlib	4	Matteo Martinelli
08-ott-24	ML Intro	4	Natalia Hadjidimitriou
10-ott-24	KNN – Nearest Neighbor	4	Natalia Hadjidimitriou
17-ott-24	Decision Tree Classification	4	Natalia Hadjidimitriou
22-ott-24	Datawarehouse + MySQL	4	Natalia Hadjidimitriou
24-ott-24	Datawarehouse + MySQL	3	Natalia Hadjidimitriou
29-ott-24	Linear Models	4	Marco Mamei
12-nov-24	Neural Networks	4	Marco Mamei
14-nov-24	Neural Networks	4	Marco Mamei
19-nov-24	ML & Optimization	3	Fabio D'Andreagiovanni
21-nov-24	ML & Optimization	3	Fabio D'Andreagiovanni
26-nov-24	ML & Optimization	3	Fabio D'Andreagiovanni
28-nov-24	Regression	4	Marco Mamei
03-dic-24	Unsupervised Learning	4	Marco Mamei
05-dic-24	Unsupervised Learning / RL	4	Marco Mamei
10-dic-24	Bigdata	3	Matteo Martinelli
12-dic-24	ERP	3	Matteo Martinelli