

# Computer Sciences & data sciences

The incredible amount of data that our modern societies are accumulating has led to the development of specific techniques for sorting, ordering, analysing or displaying these data. Data sciences is the discipline based on mathematical, statistical and computer tools, including machine learning, that encompasses these techniques. This course will introduce to the students the basic notions and techniques of data sciences, including machine learning, with practical applications in Biomedical or Environmental engineering. The course includes a significant practical component with programming in the Python language.

## Summary:

Lecture 1: Data : The data in data science

Lab 1 : Python basics adapted from the “Machine learning preparatory week @PSL” (Python & NumPy)

Lecture 2: Python and pandas : Tabular data in Python

Lab 2: Notebook on European past floods

Lecture 3: Machine Learning: history; applications; recent successes from the “Machine learning preparatory week @PSL”

Lab 3: Loading big datasets; Analyzing big datasets

Lecture 4: Introduction to machine learning from the “Machine learning preparatory week @PSL”

Lab 4: Dimensionality; Dimensionality reduction : Principal Component Analysis; Classification

Lecture 5: Supervised machine learning models from the “Machine learning preparatory week @PSL”

Lecture 6: Scikit-learn: estimation and pipelines from the “Machine learning preparatory week @PSL”

Lab 5: finish the previous notebooks

Project: Presentation, description and modeling

## References

### Python for data science

[Programming in Python for Data Science](#)

[Scientific Computing in Python: Introduction to NumPy and Matplotlib](#)

[Python Data Science Handbook](#)

### Python for geo data science

[Geo-Python course](#)

[Automating GIS-processes](#)

From:

<https://wiki.centrale-med.fr/msct-cse/> -

Permanent link:

<https://wiki.centrale-med.fr/msct-cse/csds>

Last update: **2023/07/12 12:39**

