

# Course unit: Data science project

## Course metadata

- Title in French: Projet data
- Course code: tba
- ECTS credits: 3
- Teaching hours: 60h
- Type: advanced course
- Language of instruction: French
- Coordinator: tba
- Instructor(s): Alexandre Chirié (S4M), Maximilien Defourné (S4M)
- *Last update 24/03/2021 by C. Pouet*

## Brief description

The course consists of a theoretical part and a practical part, simulating a business project.

## Learning outcomes

- Understand the workflow of a data science project in a business context
- Be able to account for business (collection of needs, project lifecycle, communication) and technical (data, machine learning, scaling) constraints

## Course content

1. Data science in business
  - The main issues
  - Examples of data project
2. Starting a data science project
  - The constraints of data science projects
  - Finding data
  - Acquiring information
  - Playing with data
3. Lifecycle of a project
  - The Bias-Variance tradeoff
  - Feature Selection
  - Feature Engineering
  - Defining a metric
4. The basic models
  - Regressions (linear, polynomial, penalized et logistic)
  - Decision trees (random forest and gradient boosting)
5. Focus Natural Language Processing (NLP)
  - Word Embedding
  - Example: Sentiment analysis

## Bibliography

Check the availability of the books below at [Centrale Marseille library](#).

- Zeng, A and Casari, A. Feature Engineering for Machine Learning. O'Reilly Media.
- Müller, A. and Guido, S. Introduction to Machine Learning with Python. O'Reilly Media.

From:

<https://wiki.centrale-marseille.fr/mmefi/> -

Permanent link:

<https://wiki.centrale-marseille.fr/mmefi/en:ddefiprod?rev=1629969457>

Last update: **2021/08/26 09:17**

