

# Data Mining & Machine Learning in Python

## Duration

30h (5 days x 6h)

## Description

Learning Data Mining or Machine Learning is not an easy task. It's transversal to computer science, programming, information theory, statistics, and artificial intelligence.

Becoming a professional in the field through self-teaching is quite challenging, and you'll often feel that you didn't make it in the end.

This course gives extensive coverage of the subject and teaches all the latest techniques in extraction, transformation, and data analysis to solve real problems, more focused on practice instead of theory.

## Audience

- Everyone that needs to put into practice projects that involve data mining/machine learning
- Data analysts/researchers
- Business intelligence practitioners

## Prerequisites

- Experience in Python programming or completion of our *Python Fundamentals* course
- Experience in data analysis or completion of our *Python for Data Analysis* course

## Objectives

After completing the course, the students should:

- Understand the actual panorama of a world governed by data, and the importance of data engineering
- Understand what data mining is and what is the role of machine learning
- Learn and use the tools available in Python for these fields

## Course Outline

1. Welcome
  - a) Course Overview
  - b) Installation and Setup
2. New Data Ecosystem
  - a) The Big Data paradigm
  - b) What is Data Mining
  - c) Use Cases for Data Science/Mining

3. Data Engineering
  - a) ETL process (Extract, Transform, Load)
  - b) Extracting from multiple sources and file formats
    - i. Excel
    - ii. CSV
    - iii. REST API's (JSON)
    - iv. Web Scraping
    - v. FTP
    - vi. Google Drive
    - vii. Databases
  - c) SQL has the universal language for data engineering
  - d) Data lake
  - e) Unique data models
  - f) Data warehousing
4. Machine Learning
  - a) What is Machine Learning
  - b) Introducing Scikit-Learn
  - c) Training/Test datasets
  - d) Hyperparameters and Model Validation
  - e) Feature Engineering
  - f) ML techniques overview:
    - i. Naive Bayes
    - ii. Random Forests
    - iii. Linear Regression
    - iv. Decision Trees
    - v. SVM
    - vi. K-Means
    - vii. KNN's
5. Advanced examples
  - a) EDA
  - b) A/B Testing
  - c) Fraud detection
6. Capstone project
7. Goodbye
  - a) Course Recap
  - b) Course evaluation
  - c) Q&A