

# Applied Machine learning with Python

**Lecturer: Andrea Giussani**

## Course language

English

## Course description and objectives

The purpose of this course is to give you a solid introduction to modern applied Machine Learning (ML) methods and pipelines that are generally available for practitioners in the field of Machine and Statistical learning. In particular, you will be guided through the construction of an end-to-end pipeline using both linear and nonlinear methods. In particular, each method will be briefly described and accompanied by hands-on, practical snippets written in Python.

At the end of this course, each participant will be able to independently apply modern machine learning techniques to elementary problems related to Economics and Social Sciences.

## Audience

The course is aimed for people involved in applying ML methods in the industry, especially for those interested in working in data analytics, or for those who want to extend their knowledge in modeling and statistical inference.

Upon successful completion of the course (attendance of at least 75% of the scheduled hours and passing the final exam), students will get 2 credits and an Open Badge, sharable across the web (LinkedIn) or personal CV.

## Prerequisites

The student must have a solid knowledge of Probability and Statistical Inference, and great knowledge of the Python language. All students are strongly encouraged to bring their own laptop.

It is strongly recommended that students have been exposed to the following libraries: Numpy, Pandas, Matplotlib. We are going to use them a lot, so please be sure you have a good understanding of them.

## Duration

24 hours

## Teaching mode

This course will take place exclusively in **synchronous, in-person sessions**. Remote participation will not be offered.

The **final test** of the course will take place on the **last day of class**.

## Calendar

Lesson	Date	Time	Room
1	Thu 05/02/2026	18.15 - 19.45	N27 (L. Del Vecchio)
2	Fri 06/02/2026	16.30 - 18.00	N27 (L. Del Vecchio)
3	Fri 13/02/2026	14.45 - 16.15	N27 (L. Del Vecchio)
4	Fri 13/02/2026	16.30 - 18.00	N27 (L. Del Vecchio)
5	Thu 19/02/2026	18.15 - 19.45	N27 (L. Del Vecchio)
6	Fri 20/02/2026	16.30 - 18.00	N27 (L. Del Vecchio)
7	Fri 27/02/2026	14.45 - 16.15	N27 (L. Del Vecchio)
8	Fri 27/02/2026	16.30 - 18.00	N27 (L. Del Vecchio)
9	Fri 06/03/2026	14.45 - 16.15	N27 (L. Del Vecchio)
10	Fri 06/03/2026	16.30 - 18.00	N27 (L. Del Vecchio)
11	Thu 12/03/2026	18.15 - 19.45	N27 (L. Del Vecchio)
12	Fri 13/03/2026	16.30 - 18.00	InfoAS04/05

## Syllabus of the course

Lecture Topics	Book reference
<b>1 Introduction to Machine Learning</b> <ul style="list-style-type: none"> <li>- Objectives of the course</li> <li>- Introduction to Machine Learning (ML)</li> <li>- The Python ML toolkit: a gentle introduction</li> </ul>	

<b>2</b>	<b>Modern Machine Learning Pipeline: Preprocessing</b>	<b>Ch. 1</b>
	<ul style="list-style-type: none"> <li>- Feature Distribution and Scaling</li> <li>- Normalization</li> </ul>	
<b>3</b>	<b>Modern Machine Learning Pipeline: Preprocessing (2)</b>	<b>Ch. 1</b>
	<ul style="list-style-type: none"> <li>- Imputation</li> <li>- Dealing with Categorical Variables</li> <li>- Model Selection</li> </ul>	
<b>4</b>	<b>Modern Machine Learning Pipeline: Modeling</b>	<b>Ch. 2</b>
	<ul style="list-style-type: none"> <li>- A simple ML Pipeline</li> <li>- Overfitting and Underfitting</li> <li>- Model Selection</li> </ul>	
<b>5</b>	<b>Modern Machine Learning Pipeline: Modeling (2)</b>	<b>Ch. 2</b>
	<ul style="list-style-type: none"> <li>- Linear vs Logistic Regression</li> <li>- Shrinkage Methods</li> <li>- Lab: Classification</li> <li>- Lab: Regression</li> </ul>	
<b>6</b>	<b>Modern Machine Learning Pipeline: Modeling (3)</b>	<b>Ch. 3</b>
	<ul style="list-style-type: none"> <li>- Nonlinear Models: why do we need them</li> <li>- Gradient Boosting</li> <li>- Random Forest</li> <li>- XGBoost</li> <li>- Explain Machine Learning Pipeline using Shap</li> <li>- Fraud Detection</li> </ul>	
<b>7</b>	<b>Introduction to Natural language processing</b>	
	<ul style="list-style-type: none"> <li>- Preprocessing with unstructured data</li> <li>- Working with Embeddings</li> <li>- Standard NLP Pipeline with scikit-learn</li> <li>- Lab: Text Classification</li> </ul>	

## Software used

Python 3.9.x (or greater)

### Suggested bibliography

- Giussani A., *Modern Applied Machine Learning with Python*, EGEA, 2020
- Hastie T., Tibshirani R., Friedman J., *The elements of statistical learning*, Springer, 2009

### Available seats

This activity is limited to **110** participants and reserved for **students enrolled in the Master of Science programs**. Registrations cannot be carried out once this number has been reached or after the registration period ends.