

# 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.

## 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 mode** in the **classroom**. Online mode will not be provided.

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

## Calendar

Lecture	Date	Time	Room
1	Thu 08/02/2024	18.15 - 19.45	N38 (Velodromo)
2	Fri 09/02/2024	14.45 - 16.15	N38 (Velodromo)
3	Fri 09/02/2024	16.30 - 18.00	N38 (Velodromo)
4	Thu 15/02/2024	18.15 - 19.45	N38 (Velodromo)
5	Fri 16/02/2024	14.45 - 16.15	N38 (Velodromo)
6	Fri 16/02/2024	16.30 - 18.00	N38 (Velodromo)
7	Thu 22/02/2024	18.15 - 19.45	N38 (Velodromo)
8	Fri 23/02/2024	14.45 - 16.15	N38 (Velodromo)
9	Fri 23/02/2024	16.30 - 18.00	N38 (Velodromo)
10	Mon 26/02/2024	18.15 - 19.45	N38 (Velodromo)
11	Thu 29/02/2024	18.15 - 19.45	N38 (Velodromo)
12	Mon 04/03/2024	18.15 - 19.45	InfoAS04/05

## Syllabus of the course

Lecture	Topics	Book reference
1	<b>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>	

---

## 7 Introduction to Natural language processing

- Preprocessing with unstructured data
  - Working with Embeddings
  - Standard NLP Pipeline with scikit-learn
  - Lab: Text Classification
- 

### 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 to **students of the Master of Science Programs**. Registrations cannot be carried out once this number has been reached or after closing of the registration period.