

Scientific Programming and Visualization with Python

Alessio Botta, Stefania Zinno, Giovanni Stanco

University of Napoli Federico II

Department of Electrical Engineering and Information Technologies

a.botta@unina.it stefania.zinno@unina.it, giovanni.stanco@unina.it

Introduction to Programming, Data Analysis, and Machine Learning with Python

This course provides a comprehensive and progressive overview of using Python for programming, data analysis, and an introduction to machine learning. Structured into three modules—BASICS, INTERMEDIATE, and ADVANCED—the course guides students from the fundamentals of the Python language to advanced applications in data analysis and modelling. Students will acquire practical skills through hands-on exercises with real datasets, using visualization tools and machine learning frameworks commonly applied in the field.

Module 1: BASICS - Introduction to Python - 10h

The first module introduces the fundamentals of Python and the basics of programming, essential for developing data science applications. Covered topics include:

1. **Python Syntax Basics**
Date: Tuesday, January 28, 2025, 14:30-17:30
Content: Introduction to Python syntax, understanding basic concepts, and writing foundational code.
2. **Loops**
Date: Thursday, January 30, 2025, 14:30-16:30
Content: Using for and while loops to automate repetitive tasks and efficiently manage iterations.
3. **Data Structures**
Date: Tuesday, February 4, 2025, 14:30-17:30
Content: Working with lists, tuples, dictionaries, and sets to structure and manage data effectively.
4. **Numpy Arrays and Mathematical Operators**
Date: Thursday, February 6, 2025, 14:30-16:30
Content: Utilizing Numpy arrays for numerical data operations and applying mathematical operators and basic functions for numerical computation.

Module 2: INTERMEDIATE – Plotting Graphs - 10h

In the second module, students learn how to use Python for data visualization and analysis, in addition to exploring object-oriented programming and basic statistical functions. Topics covered include:

5. **Classes and Objects**
 Date: Tuesday, February 11, 2025, 14:30-17:30
 Content: Introduction to object-oriented programming, focusing on using classes and objects to handle complex data and behaviors.
6. **Basic Statistical Functions**
 Date: Thursday, February 13, 2025, 14:30-16:30
 Content: Applying statistical functions for descriptive data analysis, covering foundational methods for summarizing data.
7. **Plotting with Matplotlib**
 Date: Tuesday, February 18, 2025, 14:30-17:30
 Content: Creating visualizations with Matplotlib, including one-dimensional, two-dimensional, and three-dimensional plots.
8. **Dataset Elaboration with Matplotlib**
 Date: Thursday, February 20, 2025, 14:30-16:30
 Content: Practical exercise focused on using Matplotlib to create and interpret visualizations based on real-world datasets.

Module 3: ADVANCED – Data Handling and Machine Learning - 10h

The advanced module introduces data manipulation tools and machine learning with scikit-learn, preparing students for real-world applications in data analysis and predictive modeling. Topics covered include:

9. **Pandas: Dataset Elaboration**
 Date: Tuesday, February 25, 2025, 14:30-17:30
 Content: Managing, cleaning, and transforming structured data using Pandas, including practical dataset exercises.
10. **Seaborn: Dataset Elaboration**
 Date: Thursday, February 27, 2025, 14:30-16:30
 Content: Advanced data visualization techniques using Seaborn, with a focus on creating detailed and visually appealing plots.
11. **Scikit-learn: Clustering and Binary Classification**
 Date: Thursday, March 6, 2025, 14:30-17:30
 Content: Introduction to machine learning with Scikit-learn, covering clustering techniques to identify patterns in data and binary classification models for predictive tasks.
12. **Recap and Test**
 Date: Tuesday, March 11, 2025, 14:30-16:30
 Content: A review of all covered material, followed by a test to assess understanding and practical application skills.

Course Objectives

By the end of the course, students will have acquired the following skills:

- Programming in Python and using its main data structures.
- Using Numpy, Matplotlib, Pandas, and Seaborn for data analysis and visualization.
- Fundamentals of machine learning with scikit-learn, focusing on clustering and binary classification techniques.
- Practical application of data science tools and methodologies on real datasets.

This course is ideal for anyone looking to get started in data science and machine learning, providing a solid theoretical and practical foundation.

The course lasts **30 hours**, with **2- or 3- hour lessons**.

The course will involve practical assignments and a final test on the last day of the course. Students are required to bring their own laptops.