IProML: Introduction to Programming and Machine Learning in Python, 2021

Syllabus

Course responsible:

Andrea Vandin, andrea.vandin@santannapisa.it https://www.santannapisa.it/en/andrea-vandin
Daniele Licari, daniele.licari@santannapisa.it https://www.linkedin.com/in/daniele-licari/

Language: English

Duration:
Module 1 16h, From 18/06/2021 to 02/07/2021
Module 2 14h, From 05/07/2021 to 19/07/2021

Description:
The course introduces students to programming and data analysis, using python as a reference language.

- Module 1 introduces students to the fundamental principles of structured programming with basic applications to data processing. The module starts from basic notions of programming (variables, data types, collections, control & repetition structures, functions & modules), up to basic data processing functionalities (loading, manipulation, and visualization of CSV data).
- Module 2 introduces the students to the components of typical data analysis processes and machine learning pipelines. It first builds the necessary toolset by introducing popular Python libraries for data manipulation/visualization (NumPy, Pandas, Seaborn, scikit-learn), applied to simple applications. The toolset is then applied to a more complex case study on the classification of benign and malignant breast cancer, including aspects of data preprocessing, dimensionality reduction, clustering, and classification. The module concludes by presenting KNIME, a popular python-integrated workflow-based language for data analysis.

A student who has met the objectives of the course will acquire an understanding of the issues and tasks involved in structured computer programming and data analysis, to be able to make informed decisions. The student will be able to write python programs of various nature, with a focus on complex data analysis and predictive tasks.

Prerequisites: No prerequisites for Module 1, while Module 2 requires knowledge of computer programming (possibly obtained attending Module 1).

Materials:
The course makes extensive use of online repositories and game-based e-learning platforms to

- GitHub Wiki (website): collect slides, coding examples, datasets, and further course material
- Colab: distribute and automatically provide feedback for weekly coding assignments
- Kahoot: perform online quizzes to monitor the learning process

Suggested books are:

- Learning Python, M. Lutz
- Python for Data Analysis, W. McKinney
- Statistics and Machine Learning in Python, E.Duchesnay, T.Löfstedt, F.Younes

We will use **python** as the programming language and statistical software of choice for the course.

Evaluation:
Students can attend single modules. These are 'attività trasversali', hence there will not be an exam, but an attendance certificate (attestazione di presenza) with mandatory attendance of at least 80%.

Attendance:
Due to restriction imposed by the COVID-19 epidemics, the course will likely be conducted remotely.

Schedule:

Module 1 – 16 hours

Class	Topic	Date	Time
1	Course Introduction & Console I/O & Variables	18/06	15:00-17:00
2	Data types & Operations	21/06	15:00-17:00
3	Collections & First plots	23/06	15:00-18:00
4	Control statements CSV manipulation on COVID19 data	25/06	15:00-18:00
5	Functions Application to epidemiological models Creation of word clouds from online news	28/06	15:00-18:00
6	Modules & Exceptions & OOP	02/07	15:00-18:00

Module 2 – 14 hours

Class	Topic	Date	Time
1	Course & Project Introduction Advanced libraries for data manipulation (Pandas & NumPy) 1 Application to official Italian COVID'19 data Application to Yahoo! Finance stock prices	05/07	15:00-18:00
2	Advanced libraries for data manipulation (Pandas & NumPy) 2 Application to official Italian COVID'19 data Application to Yahoo! Finance stock prices	09/07	15:00-18:00
3	Introduction to ML & Data pre-processing. Unsupervised ML Application to breast cancer diagnosis	12/07	15:00-18:00
4	Supervised ML & Project supervision Application to breast cancer diagnosis	16/07	15:00-18:00
5	KNIME a graphical language for complex data analysis	19/07	15:00-17:00