Applied Machine learning with Python

Lecturer: Andrea Giussani

Course language

English

Course description and objectives

The purpose of this course is to give students a solid introduction to modern applied Machine Learning (ML) methods and pipelines that are available for practitioners in the field of machine learning and statistical learning. In particular, the student 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 reader is expected to have a solid knowledge of Probability and Statistical Inference, and great familiarity with the Python language (or any other high-level, programming language, such as R or Julia). 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.

Duration

24 hours

Teaching mode

It will be possible to join the course in presence and/or in distance, by connecting remotely and following the streaming of the lesson held in the classroom.

Calendar

Lecture	Date	Time	Room
1	Tue 15/02/2022	18.40 – 20.10	Info6
2	Thu 17/02/2022	18.40 – 20.10	Info6
3	Tue 22/02/2022	18.40 – 20.10	Info6
4	Thu 24/02/2022	18.40 – 20.10	Info6
5	Fri 25/02/2022	15.00 – 16.30	Info6
6	Fri 25/02/2022	16.50 – 18.20	Info6
7	Tue 01/03/2022	18.40 – 20.10	Info6
8	Thu 03/03/2022	18.40 – 20.10	Info6
9	Fri 04/03/2022	15.00 – 16.30	Info6
10	Fri 04/03/2022	16.50 – 18.20	Info6
11	Tue 08/03/2022	18.40 – 20.10	Info6
12	Thu 10/03/2022	18.40 – 20.10	InfoAS04/05

Syllabus of the course

Lecture	Topics	Book reference
1	**Introduction to Machine Learning** - Objectives of the course - Introduction to Machine Learning (ML) - The Python ML toolkit: a gentle introduction	
2	**Modern Machine Learning Pipeline: Preprocessing** - Feature Distribution and Scaling - Normalization	Ch. 1
3	**Modern Machine Learning Pipeline: Preprocessing (2)** - Imputation - Dealing with Categorical Variables - Model Selection	Ch. 1
4	**Modern Machine Learning Pipeline: Modeling** - A simple ML Pipeline	Ch. 2

	- Overfitting and Underfitting - Model Selection	
5	**Modern Machine Learning Pipeline: Modeling (2)** - Linear vs Logistic Regression - Shrinkage Methods - Lab: Classification - Lab: Regression	**Ch. 2**
6	**Modern Machine Learning Pipeline: Modeling (3)** - Nonlinear Models: why do we need them - Gradient Boosting - Random Forest - XGBoost - Explain Machine Learning Pipeline using Shap - Fraud Detection	**Ch. 3**
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* – BUP (2020)
- Hastie, Tibshirani, Friedman - *The elements of statistical learning* – Springer (2009)

Available seats

This activity is limited to **96** 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.