



UNIVERSITÀ  
DEGLI STUDI DI MILANO-BICOCCA

## SYLLABUS DEL CORSO

### Laboratorio 2 - Big data analytics: modelli di machine learning e GenAI per l'analisi dei dati

2425-2-F8802N074

---

#### Learning objectives

Understand the challenges and opportunities of Big Data; Acquire knowledge of European regulations and ready to use tools regarding the collection and use of digital traces; Use machine learning models for text analysis; Integrating Artificial Intelligence into social research.

#### Contents

Definition and uses of Big Data; GDPR and DSA applied to digital data collection in the European context; Introduction to BERT and its applications in textual analysis; Introduction to AI APIs and their applications in different tasks of social research.

#### Detailed program

Lesson 1, November 18, 2024, 2.30-6.30 pm: What is Big Data and what is it used for  
Main topics:

- Definition and characteristics of Big Data (volume, variety, velocity, veracity, value).
  - Opportunities of Big Data in different sectors (social research, marketing, automation).
  - Challenges associated with Big Data: management, quality and privacy.
  - Digital traces as the main source of Big Data in social research.
- Activities:
- Discussion of real cases of Big Data in various sectors.

- Brainstorming on potential uses and problems of Big Data.

Lesson 2, November 25, 2024, 2.30-6.30 pm: Data collection and the Digital Services Act

Main topics:

- Regulations for access and use of data in the European context: from the General Data Protection Regulation (GDPR) to the Digital Services Act (DSA)
  - Methods of collecting digital traces: sources and tools with a focus on new regulations.
- Activities:
- Analysis of how DSA affects access to different types of data.
  - Exercises in collecting digital traces using Zeeschuimer and Hyphe.

Lesson 3, December 2, 2024, 2:30-6:30 PM: Introduction to Python

Main topics:

- Introduction to using Python for data wrangling (numpy, pandas, matplotlib).
  - Introduction to the cloud environment CoLab for collaborative programming.
- Activities:
- Exercises on data pre-processing, cleaning, and visualization with Python.
  - Guided exploration of the Twitter dataset (about 2 million tweets) on floods in Emilia Romagna.

Lesson 4, December 9, 2024, 2:30-6:30 PM: Machine Learning for text analysis

Main topics:

- Architecture of BERT (Bidirectional Encoder Representations from Transformers) and its applications in natural language processing (NLP).
  - Advantages of open source compared to closed commercial products.
- Activities:
- Reconnaissance and presentation of online applications based on BERT such as ClimateBERT.
  - Exercises in text classification and sentiment analysis with BERT.

Lesson 5, December 16, 2024, 9:30am-1:30pm: Using Artificial Intelligence APIs

Main topics:

- Introduction to Artificial Intelligence APIs: operation and potential.
  - Overview of various Artificial Intelligence APIs using Replicate.
- Activities:
- Understanding and using an API documentation.
  - Exercises in transcribing interviews with WhisperAI and classifying images with Memespector or ImageSorter or PixelPlot.

Lesson 6, December 16, 2024, 2.30-6:30pm: Seminar and final project

Main topics:

- Seminar by prof. Guido Anselmi starting from his latest book "Opening the Black Box. Reflective Course in Computational Sociology"
  - Discussion with the author on the implications of Big Data Analytics in platform capitalism (e.g. the Airbnb case).
  - Review of the topics covered during the lab.
  - Implementation of a final project in small groups in a hackathon style.
- Activities:
- Group work for the design and implementation of the final project.
  - Presentation of the project to the other participants and final evaluation.

## **Prerequisites**

Basic programming concepts and programming logic.

## **Teaching methods**

The course is designed as a hands-on learning experience based on the 'learning by doing' principle. The teaching method will follow a 'flipped classroom' approach: readings will be carried out independently, while in class, we will discuss key concepts and guided exercises will be carried out.

## **Assessment methods**

The evaluation will be based on the quality of the final project and active participation during lessons.

## **Textbooks and Reading Materials**

Handout provided by the teacher at the beginning of the course.

## **Sustainable Development Goals**

QUALITY EDUCATION

---