UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

SYLLABUS DEL CORSO

Laboratorio di Data Analytics

2627-2-F8205B014-F8205B014-2

Learning objectives

The laboratory aims to provide theoretical and, above all, practical skills for understanding the methods and tools of artificial intelligence applied to data analysis, with particular attention to the relationship between Big Data and AI. The fundamentals of programming in Python and of data querying with SQL are consolidated as preparatory groundwork for the analytical activities. A substantial part of the course is devoted to text mining and Natural Language Processing (NLP), from classical text representations (Bag of Words) through to embeddings, deep learning models, and Large Language Models (LLMs). The laboratory also introduces the principles of Explainable AI (XAI), with the aim of developing the ability to critically interpret models and their results in support of decision-making processes, including in the healthcare and epidemiological domains. Finally, through data visualisation techniques, the course seeks to make information and analytical results readily accessible and intelligible. The path concludes with the development of a group application project.

Contents

Introduction to AI and Big Data. Python and SQL recap. Text mining and NLP (Bag of Words, embeddings, deep learning, LLMs). Explainable AI. Data visualisation.

Detailed program

Introduction to AI and Big Data Analytics: objectives and rationale of artificial intelligence; the relationship between Big Data and AI. Python recap: core elements of the language for data analysis. SQL recap: query theory and hands-on exercises on databases. Text representation with the Bag-of-Words (BoW) model. NLP laboratory: text pre-processing and analysis. Deep Learning and embeddings: neural networks and distributed text representations. NLP laboratory: building application pipelines. Large Language Models (LLMs): operating

principles and applications. Introduction to Explainable AI (XAI): model interpretability and explainability. XAI laboratory. Data Visualisation laboratory: techniques for the visual representation and storytelling of data.

Prerequisites

Basic knowledge of Machine Learning.

Teaching methods

In-person lectures (DE). Laboratories (DI) using software tools. Up to 30% of hours may be delivered online (synchronous). 24h lectures (delivery); 18h interactive (laboratories); 6h exercises + laboratory.

Assessment methods

Assessment is based on the development of a group project, carried out in part during laboratory hours. The project requires the integrated application of the tools and methods covered in the laboratory (data analysis, NLP, machine and deep learning models, XAI, and data visualisation). The evaluation takes into account: the methodological and technical correctness of the work carried out, the ability to critically interpret the results, and the quality of the data communication. Evaluation includes the submission of the deliverables produced and a discussion/presentation of the project, during which the contribution and theoretical knowledge of each individual group member on the course topics are also assessed.

Textbooks and Reading Materials

Lectures supported by slides, laboratory sessions, and application case studies. Reference scientific articles will be provided by the instructor. The software used will be open-source.

Semester

April–June.

Teaching language

Italian

Sustainable Development Goals

GOOD HEALTH AND WELL-BEING | QUALITY EDUCATION