UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

COURSE SYLLABUS

Big Data Analytics - 2

2627-3-E3303M016-T2

Aims

The course aims to provide both theoretical and practical skills for understanding methods and tools for analysing large volumes of data, with a particular focus on data wrangling and machine learning techniques using tools such as OpenRefine and KNIME. It also aims to develop the ability to interpret analytical results, thereby supporting effective decision-making processes critically. A significant part of the course is dedicated to communicating data and analytical outcomes through data visualisation techniques, making information easily accessible and understandable. Finally, the course introduces the fundamental principles of generative artificial intelligence, with a focus on its application for the automatic synthesis and description of data.

Contents

Introduction to big data
Data processing, machine learning and text mining
Data visualisation

Detailed program

Introduction to Big Data: fundamental concepts, challenges in managing large volumes of data
Data wrangling with OpenRefine: data cleaning and transformation
Introduction to machine learning and text mining: supervised learning models and an introduction to text processing
Basic ML workflow with python: building machine learning pipelines, with a focus on NLP
Data visualisation: techniques for visual representation of data and storytelling
Generative AI: automatic synthesis and description of data using generative models

Prerequisites

None

Teaching form

Lectures (DE) held in person
Workshops (DI) using software tools
Possibility of delivering up to 30% of hours online synchronously
24h of lectures (theoretical content)
18h of interactive workshops
21h dedicated to exercises and lab work

Textbook and teaching resource

Lectures with the support of slides, laboratory and real-life case studies. Scientific Papers and books indicated by the lecturer. The software used is either available as open source or through academic license

Semester

III ciclo

Assessment method

The verification method is based on a written test whilst the oral examination will be provided on request.

The written test takes place at the computer and it consists of open and closed questions with multiple answers on all course topics.

The evaluation is focused on the student's ability to answer to specific questions by referring both to the theoretical and practical aspects (through examples) connected to the requested topic.

The written test is common for both attending students and non-attending students.

The oral exam is aimed at assessing the theoretical knowledge of the student on the topics of the course. The ability to reason and deepen the issues proposed during the examination and the methodological rigor of their development will be evaluated.

Office hours

By Appointment

Sustainable Development Goals