UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

COURSE SYLLABUS

Data Analytics

2526-2-F1801Q104

Aims

The main objectives of the course concern the acquisition of knowledge related to the main techniques of data analysis, considering both structured and unstructured data, developing specific skills regarding data, networks and text analytics. Students will also be able to design and deploy applications for the development of specific analytics functionalities (Python)

The specific objectives with respect to the Dublin Descriptors (DdD) are the following:

1. Knowledge and understanding
The student will acquire:

- Theoretical knowledge on the main techniques of data analysis, both structured and unstructured;
- Understanding of the fundamentals of Graph Theory and Natural Language Processing (NLP) theory;
- Knowledge about visualization techniques.

2. Applied knowledge and understanding
The student will be able to:

- design and develop analytics applications in Python;
- apply learned methods and models to solve concrete data analytics problems;
- integrate visualization techniques to represent results effectively.

3. Autonomy of judgment
The student will develop the ability to:

- critically evaluate the effectiveness of the analytics techniques adopted based on the type of data and application context;
- consciously choose models and tools appropriate to the analytical objective;
- interpret experimental results, analyzing implications and limits.

4. Communication skills
The student will be able to:

- present the results of their technical analyses in a clear and structured way;
- present their design solution during the oral exam and justify the methodological choices adopted;
- work in a group in the realization of the project, contributing to the documentation and presentation.

5. Learning ability

- The student will be able to develop autonomous study skills through the completion of optional assignments resulting from laboratory activities focused on the use of network analytics and Natural Language Processing models and tools.

Contents

Introduction to data analytics, with particular reference to different types of data and different analysis techniques. Focuses on specific data analytics contexts such as network analytics and text analytics.

Detailed program

Graph Teory

1. Static and dynamic networks: statistical properties for structural analysis
2. Measures of centrality
3. Graph clustering: algorithms and applications (community detection)

Natural Language Processing

1. Basics of Natural Language Processing
2. Statistical and Neural Language Models
3. Word and Sentence Embeddings
4. Tranformers and Attention Mechanisms
5. Large Language Models (ELMO, BERT, GPT e LLAMA)
6. Explainability methods for language models

Prerequisites

No essential prerequisite. Basic knowledge of linear algebra and graph theory is helpful.

Teaching form

- 12 lessons of 2 hours delivered in presence mode;
- 12 laboratory activities of 2 hours delivered in presence mode;

The course will be delivered in Italian.

Textbook and teaching resource

Albert-László BARABÁSI. Network science. Cambridge University Press.
Cristopher MANNING and Hinrich SCHÜTZE. Foundations of Statistical Natural Language Processing. MIT Press.
Daniel Jurafsky and James Martin, "Speech and Language Processing, 2nd Edition", Prentice Hall, 2008.

Semester

Second semester.

Assessment method

Group/individual project (with oral presentation) and oral exam. No mid-term tests.

The project will consist of the development of an analytics tool based on methods and models presented in class. The project evaluation consists of a numerical evaluation expressed in a range of 0-24. For the evaluation of each project, the following points will be evaluated:

- Adopted methodology (models and methods): max 7 points
- Demonstrator and visualization techniques adopted: max 5 points
- Analysis of experimental results: max 7 points
- Presentation: max 5 points

The oral exam includes 4 theoretical questions among the course topics listed in the detailed program. For each question, a score equal to -2 will be given to a wrong response or a missing answer, and a score equal to +2 for a correct answer.

Office hours

On appointment.

Sustainable Development Goals