UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

COURSE SYLLABUS

Data Analytics

2223-2-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 (for example, python and R scripts).

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

1. Static and dynamic networks: statistical properties for structural analysis
2. Measures of centrality
3. Graph clustering: algorithms and applications (community detection)
4. Basics of Natural Language Processing
5. Affective Computing: lexicons and neural models of language (word embeddings, sentence embeddings)
6. Named-Entity Recognition: Conditional Random Fields and neural extensions
7. Topic Extraction: Latent Dirichlet Allocation, Neural Topic Models
8. Visualization techniques and tools

Prerequisites

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

Teaching form

Classroom lectures, exercises and laboratory activities.

The course will be given 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.
Fabio TAMBURINI. Neural Models for the Automatic Processing of Italian. Patron Editore.

Semester

Second semester.

Assessment method

Team project (with oral presentation) and oral exam. No intermediate 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): 7 points

- Demonstrator and visualization techniques adopted: 5 points

- Analysis of experimental results: 7 points

- Presentation: 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