UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

COURSE SYLLABUS

Machine Learning and Ai-Applications for Data Management

2627-2-F7603Q028-F7603Q02802

Aims

By the end of this module, students will acquire the theoretical foundations and operational tools of the main machine learning approaches for data management, applied to the analysis of macro synthetic indicators for sustainability. In particular, they will be able to:

- master the fundamental elements of multivariate analysis and multicriteria decision making to analyse complex data systems defined by macro indicators of sustainability (knowledge and understanding)
- know the principles and operating conditions of the main techniques and choose and manage the investigation approaches most suited to the purposes of the analysis (applying knowledge and understanding)
- evaluate the characteristics, the fields of application and the advantages and disadvantages of the individual machine learning techniques (making judgements)
- suggest and justify the choice of the approach considered most suitable for a specific problem, communicating the results to an audience that might not be familiar with these methods (communication skills)
- learn independently and explore further methodological extensions of machine learning for data management (learning skills).

Contents

- Data structure and data pre-processing in the framework of machine learning; covariance and correlation
- Strategies for data exploration and the rationalization of complex problems: Principal Component Analysis
- Distances for similarity and diversity measures
- Supervised modelling for the analysis of indicators for sustainability
- Strategies for Multicriteria Decision Making
- Practical experience and case studies

Detailed program

The structure of multivariate data:

- Covariance and correlation
- Data scaling and pre-treatment: centering, auto-scaling, range scaling, variance scaling

Strategies for data exploration and the rationalization of complex problems:

- Analysis of the structure and exploration of data related to complex systems in the framework of macro synthetic indicators for sustainability
- Principal Component Analysis (PCA): scores and loadings; eigenvalues and definition of significant components (rank analysis)
- Examples of PCA and case studies

Analysis of similarity and diversity in complex systems:

- The concepts of analogy, similarity, dissimilarity and distance
- Distance and similarity measures for quantitative and binary data
- Cluster Analysis: examples and case studies

Supervised modelling for the analysis of indicators for sustainability:

- The concept of bias and validation methods; validation techniques of supervised models
- Multivariate regression and classification methods: examples and case studies

Strategies for Multicriteria Decision Making:

- Desirability and utility functions for the rational ranking of multivariate data; Pareto front
- Examples and case studies

Prerequisites

Basic knowledge of univariate statistics

Teaching form

Face-to-face lectures and practical laboratory activities:
10 hours, in-person lectures, delivered didactics
10 hours, in-person lectures, interactive teaching
10 hours, in-person lectures, interactive teaching to produce a report on a case study, to be analysed by the students in groups

Textbook and teaching resource

Slides
Scientific papers made available on the e-learning website of the course
Data and code for practical sessions made available on the e-learning website of the course

Semester

First semester

Assessment method

For the assessment method and evaluation, please refer to the syllabus of the course *Laboratory of Data management and synthetic indicators for sustainability*.

Office hours

By appointment scheduled via email

Sustainable Development Goals

QUALITY EDUCATION