UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

SYLLABUS DEL CORSO

Data Mining

2122-3-E4101B026

Learning objectives

The course aims at introducing the main concepts behind the Data Mining world, from data pre-processing to model selection.
At the end of the course, students will be able to compare and select the best Data Mining method for the problem under study. Students will also be able to solve the main issue related to data and, autonomously, tackle complex real problem .

Contents

The course deals with techniques for handling specific data's issue and supervised and unsupervised statistical methods. Additionally, the course deals also with Text Mining techniques.

Detailed program

1. Introduction to Data Mining. Main concepts and examples.
2. Data pre-processing: how to deal with missing values
3. Introduction to classification. Main concepts and examples. Classification methods: logistics regression, linear and quadratic discriminant analysis and k-nearest neighborhood classifier.
4. Overfitting and related techniques.
5. Introduction to clustering. Main concepts and examples. Clustering methods: hierarchical and partitional clustering.
6. Text Mining. Main concepts and examples. Pre-processing (stop words, stem words, ...), visual representations and clustering for Text Mining.

Prerequisites

Multivariate Statistical Analysis and R language.

Teaching methods

Lectures and computer lab.

Assessment methods

Project work and oral exam.

Written exam

Written exam aimed at assessing the competence acquired during the course.

Project work

Project work (also in group) related to the analysis of a real data problem. The problem is chosen by the students or assigned by the professor. The project should be done in R and it aims at demonstrating the ability in dealing with real application applying what has been studied during the course.

Oral exam

Project work's presentation and discussion. Questions about theory related to course's subjects.

During Covid-19, oral exam will be done using the WebEx platform.

Note

No middle exam are expected.

Student workers (non-attending students) are kindly invited to contact the professor at least 15 days before the exam date.

Textbooks and Reading Materials

Gareth J., Witten D., Hastie T., Tibshirani R., *An Introduction to statistical learning with application in R*, springer (2013).

Materials will be also provided during the course.

Semester

I Semester

Teaching language

Italian