UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

SYLLABUS DEL CORSO

Data Mining M

2223-2-F8204B014

Learning objectives

The course aims to provide data analysis and data mining tecniques and to improve predictive modelling skills by using the R software environment for statistical computing.

Contents

Study the set of exploitable data mining and supervised learning methods.
Elucidate a framework for constructing models that generate accurate predictions by means of R. This framework includes pre-processing the data, splitting the data into training and testing sets, selecting an approach for identifying optimal tuning parameters, building models, and estimating predictive performance.

Detailed program

- Important concepts: overfitting, bias and variance tradeoff, optimism
- Regression splines and additive models
- Features selection
- Dimensionality reduction
- Ensemble methods
- The model and the modelling process

Prerequisites

Knowledge of topics covered in the courses *Probability and Statistics M* and *Advanced Statistics M* is highly recommended.

Teaching methods

Lessons are held both in classroom and in lab, integrating theoretical principles with practicals of data analysis and programming in R.

Assessment methods

The exam consists in a data analysis and an oral examination

The data analysis, in addition to the predictions, requires to submit a report containing the description of the analysis and the code used must be submitted by the deadline (at least one week before the exam session). It will be possible to submit the predictions only once per Academic Year.

Textbooks and Reading Materials

Required

- Azzalini, Scarpa (2004). *Analisi dei dati e data mining*. Springer-Verlag Italia
- Hastie, Tibshirani, Friedman (2009). The Elements of Statistical Learning. Springer

Optional :

- Kuhn, Silge (2022). *Tidy Modeling with R.* O'Reilly Media, Inc.
- Lewis, Kane, Arnold (2019) *A Computational Approach to Statistical Learning*. Chapman And Hall/Crc.
- Kuhn, Johnson (2019). *Feature Engineering and Selection*. Chapman and Hall/CRC

Semester

First semester, second period.

Teaching language

The lessons are held in Italian, but the textbooks are in English.

Sustainable Development Goals