



INTRODUZIONE AL DATA MINING

INF/01 - 9 CFU - 1° Semester

Teaching Staff

ALFREDO FERRO

Email: ferro@dmi.unict.it

Office: Stanza 40, Blocco III, Dipartimento di Matematica e Informatica, Viale Andrea Doria 6, 95125 Catania (CT)

Phone: 0957383071

Office Hours: Su appuntamento

GIOVANNI MICALE

Email: gmicale@dmi.unict.it

Office: Dipartimento di Matematica e Informatica, Blocco III, Stanza 40

Phone: 0957383071

Office Hours: Mercoledì - 11-13

LEARNING OBJECTIVES

General teaching training objectives in terms of expected learning outcomes.

Knowledge and understanding: The course aims to give the knowledge and basic and advanced skills to the analysis of data.

Applying knowledge and understanding: the student will acquire knowledge about the models and algorithms for analyzing data such as: mining high support, recommendation systems, search for similarities in high dimension, networks analysis, text mining, classification and clustering.

Making judgments: Through concrete examples and case studies, the student will be able to independently develop solutions to specific problems related to data analysis.

Communication skills: the student will acquire the necessary communication skills and expressive appropriateness in the use of technical language in the general area of data analysis.

Learning skills: The course aims to provide students with the necessary theoretical and practical methods to deal independently and solve new problems that may arise during a work activity. For this purpose, different topics will be covered in class by involving students in the search for possible solutions to real problems, using benchmarks available in the literature.

COURSE STRUCTURE

Lectures

Should teaching be carried out in mixed mode or remotely, it may be necessary to introduce changes with respect to previous statements, in line with the programme planned and outlined in the syllabus.

DETAILED COURSE CONTENT

- Background
 - Probability and statistics
 - Entropy
- Introduction to R
- High Support Data Mining
- Recommendation Systems
- Clustering (hierarchical, k-means, density-based)
- Classification
- Bayesian Classifiers
- Probabilistic Graphical Models
- Networks (Centrality, Clustering Coefficient)
- Introduction to Artificial Neural Networks and Deep Learning

TEXTBOOK INFORMATION

- Mining of Massive Datasets, Jure Leskovec, Anand Rajaraman, Jeff Ullman, <http://www.mmms.org>
 - Data Mining: Concepts and Techniques, Jiawei Han and Micheline Kamber, The Morgan Kaufmann Series in Data Management Systems
 - The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Trevor Hastie, Robert Tibshirani, Jerome Friedman, Springer
-