# Data Analysis Techniques and Tools

## Prof. Daniele Tessera

***COURSE AIMS AND EXPECTED LEARNING RESULTS***

Quantitative data analysis of complex systems requires a multidisciplinary methodological training. The aim of this course is to provide students modern computational capabilities for quantitative data analyses. The course focuses on Exploratory Data Analysis (EDA) with both visual and numerical tools, basic clustering and fitting techniques and an introduction to data classification and data explanation techniques with hand-on experience on simple datasets

At the end of the course students will be able to take advantage of data analysis techniques to derive simple models able to describe and explain the principal data properties.

***COURSE CONTENT***

– Introduction to Business Intelligence, OLAP and Data Mining

– Finding data patterns: multi-dimensional analysis and modeling

* Principal Component Analysis

– Hierarchical/prototype/density-based Cluster Analysis

– Frequent Pattern Mining and Association Rule

– Finding data explanation:

 – Decision Tree: Information Entropy and Information Gain

 – Linear Regression, goodness of fit testing and optimal nested model identification

– Data Imputation with missing values: the k-Nearest Neighbors algorithm.

***TEACHING METHOD***

The course will include lectures and class exercises based on traditional teaching and teach by example principles. It is strongly advised to attend to lectures for working on case studies and examples, and for revising materials.

The course also involves lectures and exercise sessions using the integrated development environment RStudio. Active participation, and ongoing personal study are required.

***ASSESSMENT METHOD AND CRITERIA***

An oral interview aimed at assessing students' understanding of data analysis techniques and tools.

The assessment will also consider the active participation in the course, as well as a brief dissertation about an assigned project that accounts for 6/30 of the final grade.

The project topic has to discussed with the instructor in the second half of the course and can be elaborated boh indivudually or by teams of two or three students. Before taking the exam, studends are requested to send to the instructor a copy of the project report discussing achieved results and the methodological approach.

The project assessment will take into account:

* originality of the project;
* methodological approch to the project;
* students' ability to work in a team;
* project dissertation.

***NOTES AND PREREQUISITES***

Basic knowledge on statistical techniques and elementary skills on computer programming and recomened.

Attendance is strongly recommended.