

**Module Description, available in: EN**

## *Data Analysis and Classification*

**General Information****Number of ECTS Credits**

3

**Module code**

TSM\_DataAnaCla

**Valid for academic year**

2022-2023

**Last modification**

2022-01-10

**Coordinator of the module**

Michela Papandrea (SUPSI, michela.papandrea@supsi.ch)

**Explanations regarding the language definitions for each location:**

- Instruction is given in the language defined below for each location/each time the module is held.
- Documentation is available in the languages defined below. Where documents are in several languages, the percentage distribution is shown (100% = all the documentation).
- The examination is available 100% in the languages shown for each location/each time it is held.

	Lausanne			Lugano	Zurich		
<b>Instruction</b>				X E 100%			
<b>Documentation</b>				X E 100%			
<b>Examination</b>				X E 100%			

**Module Category**

TSM Technical scientific module

**Lessons**

2 lecture periods and 1 tutorial period per week

**Entry level competences****Prerequisites, previous knowledge**

- basic python scripting and SQL
- basic calculus, linear algebra and statistics concepts

## Brief course description of module objectives and content

The module is organised around 4 core subject areas:

- Data Preprocessing
- Data Classification
- Clustering
- Complex Networks

## Aims, content, methods

### Learning objectives and acquired competencies

Students understand how to use data analysis tools to process large, structured and heterogeneous data collections.

- They learn the basics of the data analysis
- They know the main tools and techniques to address the analysis of large data sets
- They learn and use the most common classification techniques
- They learn how to exploit the networking structure of the data to handle the complexity and dynamicity of large set of data
- They learn the main tools for data and results visualization
- They learn methods for processing and clustering with the purpose of effective analysis
- They can reuse the material acquired in this course in their own working environment and apply them to solve their specific problems
- They know the current research directions within these domains.

### Contents of module with emphasis on teaching content

The content of the module is the following:

- Introduction to data analysis
- Data Preprocessing (univariate and bivariate analysis, features selection, dimensionality reduction)
- Linear Regression, Logistic Regression
- Data Classification, Bagging and Boosting, classifiers evaluation
- Clustering and clustering validation
- Recommendation Systems
- Complex Networks Theory
- Network measures and Models

### Teaching and learning methods

Problem based learning. During the lesson the lecturer will introduce real world problems and the class will try to solve them together.

The lecturer will support the problem solving process, introducing new concepts and tools, as required.

Practical work will complement the theory, so that students can put in practice the studied arguments.

### Literature

Lecture slides, references to internet resources and books

## Assessment

### Certification requirements

Module uses certification requirements

### Certification requirements for final examinations (conditions for attestation)

The successful delivery of solved practical projects is a condition for entering the examination, and will contribute to the final mark.

### Basic principle for exams

**As a rule, all the standard final exams for modules and also all resit exams are to be in written form**

### Standard final exam for a module and written resit exam

Kind of exam

written

Duration of exam

120 minutes

Permissible aids

No aids permitted

### Special case: Resit exam as oral exam

Kind of exam

oral

Duration of exam

30 minutes

Permissible aids

No aids permitted