Module Description, available in: EN

Data Analysis and Classification

General Information

Number of ECTS Credits

3

Module code

TSM_DataAnaCla

Valid for academic year

2020-2021

Last modification

2019-09-25

Responsible of module

Roberto Mastropietro (SUPSI, roberto.mastropietro@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.

	Berne	Lausanne			Lugano	Zurich		
Instruction					**X** E 100%			
Documentation					**X** E 100%			
Examination					**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
- basic mathematical analysis and linear algebra

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 database technologies and data analysis tools and languages to process large data collections.

- They learn the basics of the analysis of large data sets
- They know the main tools to address analysis of large data sets
- They will learn and use the most common classification techniques
- They will 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 (noise and outliers, aggregation, PCA, features selection, etc.)
- Linear Regression, Logistic Regression
- Data Classification and classifier evaluation
- Clustering and cluster validation
- Recommendation Systems
- Complex Networks

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, as required.

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 exercises is 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 repetition exams are to be in written form

Standard final exam for a module and written repetition exam

Kind of exam

written

Duration of exam

120 minutes

Permissible aids

No aids permitted

Special case: Repetition exam as oral exam

Kind of exam

oral

Duration of exam

30 minutes

Permissible aids

No aids permitted