

COURSE: Data Mining (ING-INF/05 – CFU: 4)
TEACHER: Marta Cimitile (marta.cimitile@unitelmasapienza.it)
TUTOR: <i>da definire</i>

1. KNOWLEDGE AND SKILLS TO BE ACHIEVED DURING THE COURSE

Since data are becoming an ever-increasing part of modern life, data analytics is listed as one of the highly desirable skills employers are looking for while the talent to extract information and value from complex data is scarce.

The course is designed to improve the skill of data analytics and provides an overview on modern data mining methods together with the associated computational skills that are essential for handling large quantities of unstructured data.

The aim of the course is to produce graduates that:

- are equipped with the knowledge of the main data mining techniques and the associated computational skills for handling large quantities of unstructured data;
- have developed a critical awareness of the underlying needs of a wide variety of fields through relevant case studies;
- are able to analyse real-world data;
- have the necessary computational skills to build and analyse simple/appropriate solutions using Big Data technologies.

By the end of the course the student should be able:

- to apply and interpret methods of dimension reduction including principal component analysis;
- to apply and interpret methods for cluster analysis;
- to use R to perform data analysis tasks.

2. PROGRAM/ CONTENTS

Data Mining: basic notions
Data comprehension and setting Data Mining techniques
Text Mining and Analytics
Data interpretation and validation Weka
Exercises

3. TEXT BOOKS

Pang-Ning Tan; Michael Steinbach; Vipin Kumar; Anuj Karpatne, Introduction to Data Mining, Global Edition, Pearson (Intl), ISBN: 9780273769224, 0273769227, 2020

4. EDUCATIONAL METHOD AND TOOLS

On line lessons, web seminars, forum, quiz.

Tools: Weka, R

5. SELF-ASSESSMENT PROCEDURES

Students answer some questions about the course subjects orally. Eventually, they are also required to discuss their proposed resolution to the e-tivity proposed during the course.

6. EVALUATION METHODS (FINAL EXAM)

On line test in Moodle environment .

7. AREAS OF APPLICATION OF ACQUIRED KNOWLEDGE

The data mining skill are strongly required by all the organizations that are able to capture and have necessity to use a huge amount of data. The data need to be opportunely filtered and analyzed on the base of the organization goals.