

Academic Year: ( 2022 / 2023 )

Review date: 17-05-2022

Department assigned to the subject: Statistics Department

Coordinating teacher: MINGUEZ SOLANA, ROBERTO

Type: Basic Core ECTS Credits : 6.0

Year : 1 Semester : 1

Branch of knowledge: Engineering and Architecture

## OBJECTIVES

At the end of the course students will be able:

To understand the importance of data science in today's knowledge society.

Use data visualization techniques to understand the problems faced by a data scientist and to report the results obtained.

To know when to use a supervised or an unsupervised data analysis technique.

To know the main data analysis techniques and applications where they have been used successfully.

To know the main problems a data scientist may encounter and how to deal with them

To know the different actual data analysis tools.

To perform an basic data analysis using R-Studio.

## DESCRIPTION OF CONTENTS: PROGRAMME

1. The importance of Data Science
2. Introduction to R-Studio
3. Understanding the data: Case studies of exploratory data analysis and visualization techniques I
4. Understanding the data: Case studies of exploratory data analysis and visualization techniques II
5. Importance of a good design of experiments and choice of performance measures: precision, sensitivity, specificity. Over-fitting
6. Introduction to supervised classification: case studies on decision trees and random forests
7. Introduction to unsupervised techniques: case studies of clustering methods

## LEARNING ACTIVITIES AND METHODOLOGY

The course is taught in 14 theoretic-practical lessons and 14 practical lessons.

The subject is mostly practical, and for this reason in the master classes the main theoretical concepts of the subject will be explained, but they will also be put into practice with computer exercises. These concepts will be further elaborated in the practical classes in which various computer-based data analyses will be carried out.

The students will also have office hours where they will have the opportunity to resolve any doubts they may have about the theoretical and practical classes or about the assignments they have to carry out.

## ASSESSMENT SYSTEM

The final grade of the course will be obtained by averaging the grades of the continuous evaluation (40%) and the final exam (60%).

The continuous evaluation will consist of two assignments in which the student must apply the knowledge acquired during the course. These assignments will consists of an exploratory data analysis (15%) and a data classification task (25%). Assignments can be done in pairs and must be presented orally during class.

In order to pass the subject, a 5 out of 10 must be obtained in the final exam.

**% end-of-term-examination:** 60

**% of continuous assessment (assignments, laboratory, practicals...):** 40

## BASIC BIBLIOGRAPHY

- PATHAK, Manas A. "Beginning Data Science with R", Springer, 2014

#### ADDITIONAL BIBLIOGRAPHY

- Bruce, P. C. & Bruce, A. Practical statistics for data scientists: 50 essential concepts., O'Reilly, 2017
- Irizarry, R. A. Introduction to data science: data analysis and prediction algorithms with R., CRC Press, 2020
- Peng, R. D. R programming for data science., Leanpub, 2016