Subject

Data Science I

Year: 1   **Credits**: 9 ECTS   **Language**: Spanish

Competencies

Core competences:

CB1. Students have demonstrated knowledge and understanding in an area of study that builds on the foundation of general secondary education and is usually at a level that, while relying on advanced textbooks, also includes certain elements involving cutting-edge knowledge in their field of study

CB2. Students are able to apply their knowledge to their work or vocation in a professional manner and possess the skills that are typically demonstrated through the development and defence of arguments and problem solving within their field of study.

CB5. Students have developed those learning skills that are necessary to undertake further studies with a high degree of autonomy.

General competences:

CG4. Flexible and broad-minded Outlook

Specific competences:

CE1. Understand, know and apply the theories, tools and processes for capturing data sources from different types of data sources and for data warehousing

CE4. Be able to formulate a problem related to the world of data in terms of modelling

CE7. Planning, designing and executing projects linked to the world of data, based on the data life-cycle

Learning outcomes

RA4. Ability to think and devise solutions and responses beyond the common and that stipulated by convention. Ability to respond to unexpected current and possible future circumstances by thinking outside the box, allowing them to understand the situation in which they find themselves.

RA7. Understand, know and apply the theories, tools and processes for capturing data sources of a different type.

RA8. Understand, know and apply the theories, tools and processes for the storage of data of a different type.

RA11. Be able to formulate a problem related to the world of data in terms of modelling.

RA15. Planning, designing and executing projects based on the data life-cycle.

Syllabus

Introduction to data science
- Advanced Excel
- What is data science?
- Phases of the data life cycle
- Methodology of working in a Data Science project: CRISP-DM
- The data discovery process

Data cleaning in R
- Data cleaning process: detecting biases and errors in the data
- Detection and treatment of outliers
- Detection, typologies and treatment of missing values

Data sources & data pre-processing
- Data types: Structured and unstructured, open source and proprietary, Web Scrapping
- Data format: CSV, TXT, files, Json, relational databases, APIs, etc.
- Normalisation and discretisation: Chi-Merge, MDLP and CAIM
- Anonymisation and pseudo-anonymisation: hash functions, synthetic data and differential privacy
- Dummy variables

Design of an API
- What is an API? REST protocol
- API queries
- Using plumber in R

Repositories: Git and GitHub
- Introduction to SCV: what is Git
- Basic commands in Git: git init, git config, git add, git status, git commit.
- Basic work in GitHub: clone repositories, git push, git pull

Data governance
- Roles in data governance and science
- Data governance

Data Science Experiences (company visits)

Training activities

The training activities planned for this module are the following:

- Challenge-based learning (3 ECTS)
- Teamwork (2 ECTS)
- Workshops (1 ECTS)
- Online resources (0.5 ECTS)
- Reflection (0.5 ECTS)
- Individual work (1 ECTS)
- Carrying out projects with real companies (1 ECTS)

Assessment system

Assessment will be by means of the continuous assessment system, providing constant feedback to both teachers and students on the learning process throughout the academic period:

- Learning activities involving the presentation of knowledge and individual study may be assessed by means of oral and/or written tests, which will account for a maximum of 60% of the final mark.
- The training activities aimed at acquiring the practical skills of the subjects will be assessed through the completion of various activities (assignments, case studies, challenges, etc.) accounting for at least 40% of the final mark.

Details of the assessment and marking will be made explicit in the annual academic planning of the subjects, in accordance with the teachers responsible and the determining factors of each course.

Bibliography

- Cady, F (2017).The Data Science Handbook. USA. Wiley