1. OVERVIEW

Subject area	Data Mining
Degree	Bachelor's Degree in Physics (Computing and Data Analysis Speciality)
School/Faculty	Architecture, Engineering and Design
Year	Third
ECTS	6 ECTS
Type	Optional
Language(s)	Spanish
Delivery mode	On campus
Semester	Second semester

2. INTRODUCTION

Data Mining is an elective 6 ECTS subject (compulsory for the Computing and Data Analysis Speciality) which is studied in the third year of the Bachelor's Degree in Physics.

The aim of this subject area is to teach students about key data mining techniques and areas of application. Students will build a solid foundation in the key areas of application of data mining. They will also be able to apply data mining techniques to explain and work on the principles of machine learning, linking them to the needs of real projects.

3. SKILLS AND LEARNING OUTCOMES

Key skills (CB, by the acronym in Spanish):

- CB2: Students can apply their knowledge to their work or vocation in a professional manner and possess the skills which are usually evident through the forming and defending of opinions and resolving problems within their study area.
- CB5: Students have developed the learning skills necessary to undertake further study in a much more independent manner.
- CG2: Ability to plan and perform independent work when managing projects associated with different areas of physics.
- CG3 - To understand and express oneself in a language of science other than Spanish in a professional setting.
- CG4: To convey knowledge, procedures, results and scientific ideas in the field of physics, both orally and in writing.
- CG5: To understand diverse phenomena that, despite being physically different, share certain similarities, allowing known solutions to be applied to new problems.

Transversal skills (CT, by the acronym in Spanish):

- CT2 - Independent learning: A range of skills in order to choose research, analysis, evaluation and information management strategies from different sources, as well as to learn and put into practice what has been learnt independently.
- CT3 - Teamwork: Ability to integrate and collaborate actively with other people, areas and/or organisations to reach common goals.
- CT5 - Problem solving: Be able to critically evaluate information, separate complex situations into their constituent parts, recognise patterns, and consider alternatives, different approaches and perspectives in order to find optimal solutions and negotiate efficiently.
- CT7 - Leadership: To be able to direct, motivate and guide others, recognising their skills and abilities in order to effectively manage their development and common interests.

Specific skills (CE, by the acronym in Spanish):

- CE06 - To understand key experimental models and to perform experiments independently, describing, analysing and critically assessing experimental data.
- CE09 - To use the most suitable electronic instruments and IT tools to study physical problems and search for solutions.

Learning outcomes (RA, by the acronym in Spanish):

- RA1 - To identify the areas of application of data mining.
- RA2 - To explain the processes of knowledge extraction and preparation and to apply data mining techniques.
- RA3 - To explain the principles of machine learning and link them to the needs of a project.

The following table shows how the skills developed in the subject area match up with the intended learning outcomes:

Skills	Learning outcomes
CB2, CG2, CT2, CT5	RA1 - To identify the areas of application of data mining.
CE2, CG2, CG3, CG4, CT5, CT7, CE06, CE07	RA2 - To explain the processes of knowledge extraction and preparation and to apply data mining techniques.
CB5, CG2, CG3, CG4, CG5, CT2, CT3, CT5, CT7, CE06, CE07	RA3 - To explain the principles of machine learning and link them to the needs of a project.

4. CONTENTS

The subject is organised into five learning units that, in turn, are divided into topics:

Unit 1. Data extraction processes 1.1.
Introduction.
1.2. Data sources.
1.3. Phases in the data extraction process.
1.4. Data analysis and interpretation.

Unit 2. Data preparation
2.1. The data transformation process.
2.2. Data validation.

2.3. Data cleaning.
2.4. Transformation.
2.5. Enrichment.

Unit 3. Data mining techniques
3.1. Basic concepts.
3.2. Big Data.
3.3. Algorithms.

Unit 4. Machine learning
4.1. Fundamentals of machine learning.
4.2. Machine learning algorithms.
4.3. Model selection.
4.4. Training.
4.5. Prediction.

Unit 5. Model evaluation, diffusion and use
5.1. Model evaluation.
5.2. Interpretation of results.
5.3. Model fitting.
5.4. Interpretability and machine learning.

5. TEACHING/LEARNING METHODS

The types of teaching/learning methods are as follows:

- Case studies.
- Collaborative learning.
- Problem-based learning.
- Lectures.
- Project-based learning.

6. LEARNING ACTIVITIES

The types of learning activities, plus the amount of time spent on each activity, are as follows:

On campus:

Learning activity	Number of hours
Lectures	50
Individual and/or collaborative practical activities	35
Independent working	40
Tutorials, academic monitoring and assessment	25
TOTAL	**150**

7. ASSESSMENT

The assessment systems, plus their weighting in the final grade for the subject area, are as follows:

On campus:

Assessment system	Weighting
Individual on-campus knowledge tests (theory and/or practice)	50%
Submission of group and/or individual reports, written work, projects or exercises	35%
Oral defence	5%
Performance observation	10%

On the Virtual Campus, when you open the subject area, you'll find details of your assessable tasks, including the submission dates and assessment procedures for each task.

8. BIBLIOGRAPHY

The reference material for the subject area is as follows:

- Big Data con Python, Rafael Caballero – Editorial Alfaomega RC Libros 2019.
- Aprende Machine Learning, Juan Ignacio Bagnato – Editorial Agencia del ISBN en España, 2021.
- Interpretable Machine Learning, Christoph Molnar, 2021.