SUBJECT TEACHING GUIDE

261 - Data Mining

University Master's Degree in Data Science

Academic year 2026 - 2027

1. IDENTIFYING DATA					
Degree	University Master's Degree in Data Science			Type and Year	Compulsory. Year 1
Faculty	Faculty of Sciences				
Discipline	METHODS IN SCIENCE DATA				
Course unit title and code	261 - Data Mining				
Number of ECTS credits allocated	6	Term		Semester based (1)	
Web					
Language of instruction	Spanish	English Friendly	Yes	Mode of delivery	Face-to-face

Department	DPTO. MATEMATICA APLICADA Y CIENCIAS DE LA COMPUTACION
Name of lecturer	SIXTO HERRERA GARCIA
E-mail	sixto.herrera@unican.es
Office	E.T.S. de Ingenieros de Caminos, Canales y Puertos. Planta: + 1. DESPACHO PROFESORES (1034)
Other lecturers	STEVEN JOHAN MARIA VAN VAERENBERGH JOAQUIN BEDIA JIMENEZ RODRIGO GARCIA MANZANAS MIRIAM FUENTE GONZALEZ

3.1 LEARNING OUTCOMES

- Be able to identify different data mining problems (association, segmentation, classification, and prediction) and know which techniques to apply in each case.
- Understand the main machine learning methodologies (supervised and unsupervised).
- Understand how data mining works and know how to apply the main techniques. Know how the concept of regularization, seen in other subjects, is extended to data mining problems.
- Understand the concept of generalization and the problem of overfitting.
- Be able to critically validate learned models and identify the most appropriate model for a specific practical case.
- Be able to use software packages to solve practical data mining cases and critically analyze the solutions obtained.

4. OBJECTIVES

To gain knowledge on the statistical methods for data description, presentation and analysis. To learn how to apply the machine learning techniques to big data problems.

6. COURSE ORGANIZATION

CONTENTS	
1	Introduction and historical perspective Canonical problems and Learning Paradigms Non Supervised Learning with Association Rules Cross-validation, learning and over-fitting
2	Supervised Learning with K-Nearest Neighbors
3	Classification and regression trees Ensemble methods: Bagging and Random Forests
4	Dimension reduction with non-linear techniques
5	Non-supervised learning. Segmentation.

7. ASSESSMENT METHODS AND CRITERIA

Description	Type	Final Eval.	Reassessment	%
Evaluation of the reports of practices.	Activity evaluation with Virtual Media	Yes	Yes	60,00
Exam covering the concepts explained in the subject.	Activity evaluation with Virtual Media	Yes	Yes	40,00
TOTAL				100,00

Observations

To recover the course, each practice failed should be repeated.
The minimum marks apply to the two assessment blocks, the continuous assessment and the final exam. Based on article 35 of the Regulations, in the event that a student does not obtain the minimum grade required in any of the evaluation tests, the overall grade for the subject will be the lowest value between 4.9 and the weighted average of the tests. Assessment.All the tests will be carried out or delivered using the virtual platform of the subject. Therefore, it is the student's responsibility to ensure that they can access the virtual platform of the subject, before the beginning of the sessions in which the tests are carried out.

Observations for part-time students

Same evaluation will be applied for both full and partial time students

8. BIBLIOGRAPHY AND TEACHING MATERIALS

BASIC

An Introduction to Statistical Learning with Applications in R, Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani, Editorial Springer. Available at http://www-bcf.usc.edu/~gareth/ISL/
Dimension Reduction: A Guided Tour, Christopher J. C. Burges, Foundations and Trends® in Machine Learning: Vol. 2: No. 4, pp 275-365. http://dx.doi.org/10.1561/2200000002
Machine learning: a probabilistic perspective, Kevin P. Murphy. ISBN 978-0-262-01802-9, MIT Press Cambridge, Massachusetts, London, England