



TECHNICAL SHEET OF THE SUBJECT

Data of the subject	
Subject name	Machine Learning III: Unsupervised Learning Techniques and Network Analysis
Subject code	FCEE-BA-422
Main program	Grado en Análisis de Negocios/Business Analytics
Involved programs	Grado en Análisis de Negocios/Business Analytics y Grado en Derecho [Fourth year] Grado en Administración y Dirección de Empresas y Grado en Análisis de Negocios/Business Analytics [Fourth year] Grado en Ingeniería en Tecnologías de Telecom. y Grado en Análisis de Negocios/Business Analytics [Fourth year]
Level	Reglada Grado Europeo
Quarter	Semestral
Credits	3,0 ECTS
Type	Compulsory
Department	Departamento de Métodos Cuantitativos
Coordinator	Lucía Barcos Redín
Office hours	Request an appointment. Tutoring schedules will be available when the timetables are finalized.

Teacher Information	
Teacher	
Name	Carlos Álvarez Fernández
Department	Departamento de Métodos Cuantitativos
Office	Alberto Aguilera 23
E-Mail	calvarez@icade.comillas.edu
Teacher	
Name	Lucía Barcos Redín
Department	Departamento de Gestión Empresarial
Office	Alberto Aguilera 23 [C-429]
E-Mail	lbarcos@icade.comillas.edu

SPECIFIC DATA OF THE SUBJECT

Contextualization of the subject
Contribution to the professional profile of the degree
Machine Learning provides techniques that help extract relevant information from data. With this information, companies can make decisions and carry out actions to improve different aspects of the business. This course deals with a part of Machine Learning, specifically unsupervised learning.



In addition, and given that most of the data available today comes in the form of text, we will study some tools needed to represent and analyze text, gaining insights from it.

Furthermore, social networks are currently a very important source of information. Therefore, this course will also examine basic network analysis techniques.

Prerequisites

Fundamentals of statistical data analysis

Algebra Basics

Programming Fundamentals

Business Analytics Fundamentals

Competencies - Objectives

Competences

GENERALES

CG02	Capacidad de análisis de datos masivos procedentes de diversas fuentes: texto, audio, numérica e imagen	
CG03	Resolución de problemas y toma de decisiones en un entorno de datos masivos tanto cuantitativos como cualitativos	
	RA1	Saber seleccionar para cada problema la técnica o técnicas de análisis de datos más adecuada para poder convertir los datos <i>en bruto</i> en información y ésta en conocimiento que ayude a la toma de decisiones y a mejorar la gestión.
CG04	Capacidad para elaborar proyectos e informes de manera oral y escrita, difundiendo estas ideas a través de canales digitales	
	RA1	Ser capaz de resumir, sintetizar y comunicar de una forma atractiva y eficaz los resultados de la aplicación de las técnicas de análisis de datos, incluso de las más sofisticadas, de manera que resulten comprensibles a destinatarios no técnicos y ayuden de forma eficiente a la toma de decisiones empresariales.
CG09	Compromiso ético en la sociedad de la información	
	RA1	Persigue la excelencia en las actuaciones profesionales
	RA2	Se preocupa por las consecuencias que su actividad y su conducta pueden tener para los demás
	RA3	Incorpora en su discurso y en sus propuestas de actuaciones, las consecuencias que las mismas pueden tener para los distintos stakeholders de una organización global

ESPECÍFICAS

CE19	Conocer los fundamentos de las principales técnicas tanto de la estadística clásica (descriptiva e inferencial) como del data mining	
-------------	--	--



	RA4	Conocer los fundamentos de las principales técnicas de data mining NO supervisado
CE20	Saber modelizar un problema empresarial real que precise análisis de datos y seleccionar críticamente la técnica o combinación de técnicas más adecuada	
	RA3	Saber modelizar un problema de aprendizaje no supervisado
	RA4	Conocer las ventajas y limitaciones de los procedimientos de data mining y saber cómo enfocar un problema desde diferentes técnicas complementarias
CE21	Saber interpretar, evaluar y comunicar resultados derivados de las técnicas de análisis de datos así como usarlos para la ayuda en la gestión y la toma de decisiones empresariales	
	RA3	Saber interpretar, evaluar y comunicar los resultados derivados de un análisis que emplee técnicas de data mining no supervisado
CE22	Saber aplicar las técnicas de análisis de datos (tanto las de la estadística clásica como las técnicas de data mining) a un conjunto de datos reales, mediante el empleo de algún software apropiado para tal fin	
	RA1	Saber realizar un análisis descriptivo básico a un conjunto de datos reales usando software apropiado para tal fin
	RA4	Saber aplicar técnicas de data mining de aprendizaje NO supervisado a un conjunto de datos reales usando software apropiado para tal fin.

THEMATIC BLOCKS AND CONTENTS

Contents - Thematic Blocks

Unit 1. Introduction to unsupervised learning, Text Mining and social network analysis. Review of programming fundamentals.

Unit 2: Dimension Reduction Techniques, Principal component analysis (PCA)

Unit 3. Clustering Techniques: K-means and hierarchical clustering

Unit 4. Association rules and Recommendation Systems

Unit 5. Text representation and Text mining.

Unit 6. Network Analysis.

TEACHING METHODOLOGY

General methodological aspects of the subject

After the explanation of the main theoretical concepts of each technique, we will work with different data sets in order to show the practical application of these techniques in a business or social context. This activity will be carried out using a programming language.



In-class Methodology: Activities

Presentation of the main theoretical concepts	CG02, CG03, CG04, CG09, CE19, CE20, CE21, CE22
Resolution and discussion of different practical examples/cases (using some programming language)	
General tutoring of practical application work	
Evaluation activities	

Non-Presential Methodology: Activities

Personal study	CG02, CG03, CG04, CG09, CE19, CE20, CE21, CE22
Resolution of practical cases using different data sets and a programming language	

SUMMARY STUDENT WORKING HOURS

CLASSROOM HOURS		
Lecciones de Carácter expositivo	Ejercicios y resolución de casos y de problemas	
15.00	15.00	
NON-PRESENTIAL HOURS		
Ejercicios y resolución de casos y de problemas	Estudios individual y/o en grupo, y lectura organizada	Trabajos monográficos y de investigación, individuales o colectivos
15.00	15.00	15.00
ECTS CREDITS: 3,0 (75,00 hours)		

EVALUATION AND CRITERIA

Evaluation activities	Evaluation criteria	Weight
Final Exam	Scores rated from 0 to 10 It is necessary to obtain a minimum grade of 5 points in the final exam in order to pass the course	50 %
Teamwork (final project with a practical application)	Scores rated from 0 to 10 A minimum level of quality is required in order to pass the course (applicable to both, ordinary and extraordinary calls)	20 %
Intermediate tests and other items of continuous evaluation (weekly practices, kahoots, etc.)	Scores rated from 0 to 10	30 %

Ratings

For the evaluation of both calls, ordinary and extraordinary, the same grading system will be followed (same weightings as previously shown). The philosophy that continuous work is part of the learning of the subject underlies.

It is necessary to obtain a **minimum grade of 5 points** (out of 10) in the final exam in order to pass the course (applicable to both, ordinary and extraordinary calls).

It is necessary a minimum level of quality in the final project in order to pass the course (applicable to both, ordinary and extraordinary calls). The final project is a way of demonstrating that the student has acquired the skills to apply theoretical concepts to real data by means of the appropriate analysis techniques and the necessary programming skills.

EXTRAORDINARY CALL (2nd): the same grading system as for the ordinary call will be applied, with the same weightings. Students who have not submit or have failed the final project will have to submit it again in this call.

THIRD AND SUBSEQUENT CALLS: it will be necessary to submit a final project (25% of the grade) and take a final exam (75%). The submission of the final project is mandatory.

EXCHANGE STUDENTS (IN): same regime as regular students.

EXCHANGE STUDENTS (OUT): same system as for third and subsequent calls.

BIBLIOGRAPHY AND RESOURCES

Basic Bibliography

Material (presentations, code, etc.) available in Moodle Rooms

Shmueli, G., Bruce, G., Yahav, I., Patel, N.R., Lichtendahl, K. (2018). Data Mining for Business Analytics. Concepts, Techniques and Applications in R. John Wiley & Sons, Inc

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2021). An Introduction to Statistical Learning: with Applications in R. New York, Springer.

Complementary Bibliography

Provost, F., & Fawcett, T. (2013). Data Science for Business: What you need to know about data mining and data-analytic thinking. O'Reilly Media, Inc.

Aggarwal C.C. (2016) Recommender Systems. The textbook. Springer

Kassambara, A. (2017). Practical guide to cluster analysis in R. Published by STHDA

In compliance with current regulations on the **protection of personal data**, we would like to inform you that you may consult the aspects related to privacy and data that you have accepted on your registration form by entering this website and clicking on "download"

<https://servicios.upcomillas.es/sedelectronica/inicio.aspx?csv=02E4557CAA66F4A81663AD10CED66792>