ANX-PR/CL/001-01
LEARNING GUIDE

SUBJECT

103000935 - Data Mining

DEGREE PROGRAMME

10AZ - Master Universitario En Innovación Digital

ACADEMIC YEAR & SEMESTER

2025/26 - Semester 2

Index

Learning guide

1. Description....1
2. Faculty....1
3. Prior knowledge recommended to take the subject....2
4. Skills and learning outcomes....2
5. Brief description of the subject and syllabus....3
6. Schedule....5
7. Activities and assessment criteria....7
8. Teaching resources....10
9. Other information....11

1. Description

1.1. Subject details

Name of the subject	103000935 - Data Mining
No of credits	4 ECTS
Type	Optional/elective
Academic year ot the programme	First year
Semester of tuition	Semester 2
Tuition period	February-June
Tuition languages	English
Degree programme	10AZ - Master Universitario en Innovación Digital
Centre	10 - E.T.S. De Ingenieros Informáticos
Academic year	2025-26

2. Faculty

2.1. Faculty members with subject teaching role

Name and surname	Office/Room	Email	Tutoring hours *
Fco.javier Segovia Perez (Subject coordinator)	2305	javier.segovia@upm.es	M - 10:00 - 11:00 Hablar con el profesor
Ernestina Menasalvas Ruiz	4303	ernestina.menasalvas@upm.es	M - 10:00 - 11:00 hablar con la profesora

* The tutoring schedule is indicative and subject to possible changes. Please check tutoring times with the faculty member in charge.

3. Prior knowledge recommended to take the subject

3.1. Recommended (passed) subjects

The subject - recommended (passed), are not defined.

3.2. Other recommended learning outcomes

- Artificial Intelligence

- Statistics

4. Skills and learning outcomes *

4.1. Skills to be learned

CB07 - Que los estudiantes sepan aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio

CE-DIPO03 - Habilidad para hacer conexiones entre los deseos y necesidades del consumidor o cliente y lo que la tecnología puede ofrecer

CG01 - Que los estudiantes sean capaces de predecir y controlar la evolución de situaciones complejas mediante el desarrollo de nuevas e innovadoras metodologías de trabajo adaptadas al ámbito científico/investigador, tecnológico o profesional concreto, en general multidisciplinar, en el que se desarrolle su actividad.

CG03 - La capacidad de usar la lengua inglesa de manera competente, es decir, con capacitación para tareas complejas de trabajo y estudio.

CG07 - Capacidad de trabajar y comunicarse también en contextos internacionales.

4.2. Learning outcomes

RA140 - Obtain data mining results from descriptive, diagnostic, predictive, and prescriptive analysis by applying a broad and varied repertoire of classification, regression, association, segmentation, and clustering algorithms.

RA139 - Turn a business question into a data question, and use what is learned from applying a data mining process to make smart decisions that help the company or institution succeed.

* The Learning Guides should reflect the Skills and Learning Outcomes in the same way as indicated in the Degree Verification Memory. For this reason, they have not been translated into English and appear in Spanish.

5. Brief description of the subject and syllabus

5.1. Brief description of the subject

At the end of this course we will make you able to:

- Turn a business question into a data question, and use what is learned from applying a data mining process to make smart decisions that help the company or institution succeed.

- Obtain data mining results from descriptive, diagnostic, predictive, and prescriptive analysis by applying a broad and varied repertoire of classification, regression, association, segmentation, and clustering algorithms.

We will use a professional data mining tool called IBM SPSS Modeler, which is designed to be user-friendly and does not require any prior programming experience.

You will learn how to use this tool by following an IBM tutorial, along with interactive class exercises and assignments.

This course is very hands-on and practical?you will mainly learn by doing. Through real examples and guided practice, you'll gain the skills needed to analyze data effectively.

By the end of the course, you will be able to solve a real-world problem of your choice using data mining techniques. Throughout the course, we will go through all the key phases of data analysis, step by step, while

practicing with real-life examples.

5.2. Syllabus

1. INTRODUCTION TO DATA ANALYSIS AND THE IBM SPSS MODELER
2. BUSINESS VALUE PROPOSITION & DATA MINING
3. DESCRIPTIVE ANALYSIS USING BASIC STATISTICS
4. DESCRIPTIVE ANALYSIS: DATA VISUALIZATION
5. DESCRIPTIVE ANALYSIS: RFM
6. DESCRIPTIVE ANALYSIS: CLUSTERING
7. DIAGNOSTIC ANALYSIS: CORRELATION, ANOVA AND CHI-SQUARED TESTS
8. DIAGNOSTIC ANALYSIS: ASSOCIATION RULES
9. PREDICTIVE ANALYSIS: LINEAR REGRESSION
10. PREDICTIVE ANALYSIS: LOGISTIC REGRESSION
11. PREDICTIVE ANALYSIS: DECISION TREES
12. PREDICTIVE ANALYSIS: NEAREST NEIGHBOR
13. PREDICTIVE ANALYSIS: NEURAL NETWORKS
14. PREDICTIVE ANALYSIS: PROBABILISTIC MODELS
15. PREDICTIVE ANALYSIS: ENSEMBLE METHODS
16. PREDICTIVE ANALYSIS: DEALING WITH TIME
17. PRESCRIPTIVE ANALYSIS: TARGETING CUSTOMERS
18. PRESCRIPTIVE ANALYSIS: RECOMMENTATION SYSTEMS
19. FINAL PROJECT

6. Schedule

6.1. Subject schedule*

Week	Type 1 activities	Type 2 activities	Distant / On-line	Assessment activities
1	**Introduction to Data Mining** Duration: 00:30 Lecture **Exercise about Value Proposition Canvas** Duration: 00:30 Problem-solving class **Problem about Data Analysis on a Business Case** Duration: 00:30 Problem-solving class **Tool practice** Duration: 02:30 Problem-solving class			**Tool practice** Individual presentation Progressive assessment Presential Duration: 02:30 **Theoretical Problem** Individual presentation Progressive assessment Presential Duration: 01:00
2	**Problem about DATA VISUALIZATION using the tool** Duration: 02:00 Problem-solving class **Problem about STATISTICAL DESCRIPTION using the tool** Duration: 02:00 Problem-solving class			**Practical Problem using the tool** Individual presentation Progressive assessment Presential Duration: 04:00
3	**Problem about RFM using the tool** Duration: 02:00 Problem-solving class **Problem about CLUSTERING using the tool** Duration: 02:00 Problem-solving class			**Practical Problem using the tool** Individual presentation Progressive assessment Presential Duration: 04:00
4	**Problem about CORRELATION, ANOVA AND CHI-SQUARED TESTS using the tool** Duration: 02:00 Problem-solving class **Problem about ASSOCIATION RULES using the tool** Duration: 02:00 Problem-solving class			**Practical Problem using the tool** Individual presentation Progressive assessment Presential Duration: 04:00

5	**Problem about LINEAR REGRESSION using the tool** Duration: 02:00 Problem-solving class **Problem about LOGISTIC REGRESSION using the tool** Duration: 02:00 Problem-solving class			**Practical Problem using the tool** Individual presentation Progressive assessment Presential Duration: 04:00
6	**Problem about NEAREST NEIGHBOR AND NEURAL NETWORKS USING THE TOOL** Duration: 02:00 Problem-solving class **Problem about DECISION TREES using the tool** Duration: 02:00 Problem-solving class			**Practical Problem using the tool** Individual presentation Progressive assessment Presential Duration: 04:00
7	**Problem about TIME RELATED CASES using the tool** Duration: 02:00 Problem-solving class **Problem about ENSEMBLE METHODS using the tool** Duration: 02:00 Problem-solving class			**Practical Problem using the tool** Individual presentation Progressive assessment Presential Duration: 04:00
8	**Problem about PRESCRIPTIVE ANALYSIS using the tool** Duration: 02:00 Problem-solving class			**Practical Problem using the tool** Individual presentation Progressive assessment Presential Duration: 02:00
9				
10				
11				
12				
13				
14				
15				
16				
17	**FINAL PROJECT PRESENTATION** Duration: 00:20 Additional activities			**FINAL PROJECT presentation and all assignments uploaded** Individual presentation Progressive assessment Presential Duration: 00:20 **FINAL PROJECT presentation and all assignments uploaded** Individual presentation Global examination Presential Duration: 00:20

Depending on the programme study plan, total values will be calculated according to the ECTS credit unit as 26/27 hours of student face-to-face contact and independent study time.

7. Activities and assessment criteria

7.1. Assessment activities

7.1.1. Assessment

Week	Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
1	Tool practice	Individual presentation	Face-to-face	02:30	1%	5 / 10	CB07
1	Theoretical Problem	Individual presentation	Face-to-face	01:00	1%	5 / 10	CB07 CG03 CG07 CE-DIPO03
2	Practical Problem using the tool	Individual presentation	Face-to-face	04:00	1%	5 / 10	CB07 CG01 CG03 CG07 CE-DIPO03
3	Practical Problem using the tool	Individual presentation	Face-to-face	04:00	1%	5 / 10	CB07 CG01 CG03 CG07 CE-DIPO03
4	Practical Problem using the tool	Individual presentation	Face-to-face	04:00	1%	5 / 10	CB07 CG01 CG03 CG07 CE-DIPO03
5	Practical Problem using the tool	Individual presentation	Face-to-face	04:00	1%	5 / 10	
6	Practical Problem using the tool	Individual presentation	Face-to-face	04:00	1%	5 / 10	
7	Practical Problem using the tool	Individual presentation	Face-to-face	04:00	1%	5 / 10	CB07 CG01 CG03 CG07 CE-DIPO03
8	Practical Problem using the tool	Individual presentation	Face-to-face	02:00	1%	5 / 10	CB07 CG01 CG03 CG07 CE-DIPO03

17	FINAL PROJECT presentation and all assignments uploaded	Individual presentation	Face-to-face	00:20	91%	5 / 10	CB07 CG01 CG03 CG07 CE-DIPO03

7.1.2. Global examination

Week	Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
17	FINAL PROJECT presentation and all assignments uploaded	Individual presentation	Face-to-face	00:20	100%	5 / 10	CB07 CG01 CG03 CG07 CE-DIPO03

7.1.3. Referred (re-sit) examination

Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
FINAL PROJECT presentation and all assignments uploaded	Individual presentation	Face-to-face	00:20	100%	5 / 10	CB07 CG01 CG03 CG07 CE-DIPO03

7.2. Assessment criteria

Course Evaluation Overview:

- In every session, we will work on a problem or assignment in class. Your final grade is based on the assignments from each session, and mainly, your final project.

- Important: You must complete, upload, and pass (minimum grade of 5 out of 10) all assignments to pass the course.

Types of Evaluation

1. Continuous (Progressive) Evaluation

You can resubmit each assignment once if needed.

2. Global Evaluation

You can only submit each assignment once. No resubmissions allowed.

3. Extraordinary Evaluation

You may resubmit only the assignments you failed during the continuous or global evaluations.

8. Teaching resources

8.1. Teaching resources for the subject

Name	Type	Notes
Principles of Data Mining (Adaptive Computation and Machine Learning), D Hand, MIT Press, 2001.	Bibliography	
Jiawei Han, Micheline Kamber, Data Mining : Concepts and Techniques, 2nd edition, Morgan Kaufmann, ISBN 1558609016, 2006.	Bibliography	
Data Mining Techniques: Marketing, Sales and Customer Support, Michael J. A. Berry, Gordon Linoff, John Wiley & Sons, 1997.	Bibliography	
Pang-Ning Tan, Michael Steinbach, Vipin Kumar, Introduction to Data Mining, Pearson Addison Wesley (May, 2005). Hardcover: 769 pages. ISBN: 0321321367	Bibliography	MOST RECOMMENDED BOOK
Ian Witten, Eibe Frank, Mark Hall, Data Mining: Practical Machine Learning Tools and Techniques, 3nd Edition, Morgan Kaufmann, ISBN 978-0-12-374856-0, 2011.	Bibliography	
Página web de la asignatura en moodle	Web resource	
IBM SPSS MODELER	Others	THE TOOL WE WILL USE
aula	Equipment	

9. Other information

9.1. Other information about the subject

Several innovative teaching methodologies are implemented in the course (https://innovacioneducativa.upm.es/guias-pdi) with the aim of motivating and reinforcing student learning:
? Learning by Doing: I use this technique, first, to help students learn how to use the IBM SPSS Modeler data mining tool. They follow a tutorial at home to learn the basics, and I continue with my own tutorial during some class sessions. Second, instead of just reading about data mining or doing exercises on paper, in each class session?after I give a theoretical introduction to a concept or technique?students use IBM SPSS Modeler to analyze real data and solve problems related to that concept or algorithm.

? Project based Learning: A large part of the students' final grade comes from completing a data mining project of their choice. Around the middle of the course, students must find a dataset related to a real-world problem that interests them and begin applying a data mining process to solve it, with a final report submitted at the end. I supervise and monitor their progress, answering questions and guiding them when they encounter difficulties. Projects can be of any type. The only requirement I impose is that students must first inform me about the dataset they want to use and the type of problem they aim to solve with it. I then check whether the dataset contains enough information and data to make the project feasible. If I determine that the dataset lacks the necessary information to successfully apply data mining techniques and algorithms, I ask them to choose a different problem and/or dataset.