

ANX-PR/CL/001-01

LEARNING GUIDE

SUBJECT

103000389 - Data Mining

DEGREE PROGRAMME

10AK - Master Universitario En Software Y Sistemas

ACADEMIC YEAR & SEMESTER

2024/25 - 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.....	9
9. Other information.....	10

1. Description

1.1. Subject details

Name of the subject	103000389 - Data Mining
No of credits	4 ECTS
Type	Optional
Academic year of the programme	First year
Semester of tuition	Semester 2
Tuition period	February-June
Tuition languages	English
Degree programme	10AK - Master Universitario en Software y Sistemas
Centre	10 - Escuela Tecnica Superior De Ingenieros Informaticos
Academic year	2024-25

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

CEM7 - Evaluar y aplicar las diversas teorías matemáticas y estadísticas, y los procesos, métodos y técnicas disponibles para la extracción y descubrimiento de conocimiento a partir de grandes volúmenes de datos

CG1 - 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.

CG3 - Que los estudiantes sepan comunicar sus conclusiones y los conocimientos y razones últimas que las sustentan a públicos especializados y no especializados de un modo claro y sin ambigüedades.

4.2. Learning outcomes

RA19 - Ser capaz de realizar data mining a través de un proceso, demostrando su competencia en un estándar incluyendo las fases de conocimiento del negocio, conocimiento de los datos, análisis exploratorio de los datos, modelado, evaluación y explotación

RA21 - Comprender los fundamentos y aplicar un amplio y variado repertorio de algoritmos de clustering, estimación, predicción y clasificación.

* 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

The course is mainly dedicated to the improvement of the development of software engineering projects by means of Data Mining.

The course is very interactive, with the development of many short projects and exposition at class. Learning by doing, using the IBM SPSS tool

Students will learn the use of the IBM SPSS tool by themselves, with some teacher support

Topics:

Data Engineering, Data Mining, Business Intelligence

Descriptive, Diagnostic, Predictive and Prescriptive Analysis of data

Techniques:

- Classification

- Regression

- Association

- Clustering

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: RECOMMENDATION 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	5%	5 / 10	CG1
1	Theoretical Problem	Individual presentation	Face-to-face	01:00	5%	5 / 10	CG1 CG3
2	Practical Problem using the tool	Individual presentation	Face-to-face	04:00	10%	5 / 10	CG1 CG3 CEM7
3	Practical Problem using the tool	Individual presentation	Face-to-face	04:00	10%	5 / 10	CG1 CG3 CEM7
4	Practical Problem using the tool	Individual presentation	Face-to-face	04:00	10%	5 / 10	CEM7 CG1 CG3
5	Practical Problem using the tool	Individual presentation	Face-to-face	04:00	10%	5 / 10	CG1 CG3 CEM7
6	Practical Problem using the tool	Individual presentation	Face-to-face	04:00	10%	5 / 10	CG1 CG3 CEM7
7	Practical Problem using the tool	Individual presentation	Face-to-face	04:00	10%	5 / 10	CG1 CG3 CEM7
8	Practical Problem using the tool	Individual presentation	Face-to-face	02:00	10%	5 / 10	CG1 CG3 CEM7
17	FINAL PROJECT presentation and all assignments uploaded	Individual presentation	Face-to-face	00:20	20%	5 / 10	CG1 CG3 CEM7

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	CG1 CG3 CEM7
----	---	-------------------------	--------------	-------	------	--------	--------------------

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	CG1 CG3 CEM7

7.2. Assessment criteria

At each session we will solve a problem/assignment at class. The evaluation is based on each session problem/assignment and the final project. But solving, uploading and passing (mark 5 minimum) all problems/assignments is mandatory.

For the continuous or progressive evaluation, it is allowed one resubmission for each assignment.

For the global evaluation, you can only submit once each assignment.

For the extraordinary evaluation, only failed assignments submitted in the continuous or global examinations can be resubmitted.

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, 3rd 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

We will use the Problem-based learning technique in class. Before each session you will have to read some materials and watch some videos around a topic. Then at class I will answer questions regarding those materials and videos, and I will propose a problem related to the topic that you have to solve at class. The whole session will be for you to solve the problem, with me around to help you. It is, therefore, very important that you have in mind that this course requires your presence at class, and a computer with the tool.