

ANX-PR/CL/001-01

LEARNING GUIDE

SUBJECT

103000935 - Data Mining

DEGREE PROGRAMME

10AZ - Master Universitario En Innovación Digital

ACADEMIC YEAR & SEMESTER

2023/24 - 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.....	8
8. Teaching resources.....	12

1. Description

1.1. Subject details

Name of the subject	103000935 - 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	10AZ - Master Universitario en Innovación Digital
Centre	10 - Escuela Técnica Superior De Ingenieros Informáticos
Academic year	2023-24

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

CE-DIPO02 - Capacidad para evaluar la interacción persona-ordenador de productos y servicios de alto valor innovador

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

CE-FT01 - Capacidad para seleccionar las soluciones de almacenamiento, manipulación, análisis y visualización para datos estructurados y no estructurados financieros de fuentes heterogéneas adecuadas en función del problema a resolver y realizar una correcta comunicación del análisis

CE-FT04 - Capacidad para diseñar proyectos robustos relacionados con las finanzas y la tecnología aplicando las últimas tecnologías software y de inteligencia artificial

CE-FT05 - Capacidad para aplicar técnicas para la generación de visualizaciones adecuadas para el análisis y la exploración de datos para resolver un determinado problema asociado a tecnologías financieras

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

RA56 - develop data science projects following a methodology

RA91 - Produce a professional speech and writing on a business analysis topic

RA60 - To know and apply the main techniques to explore, describe and analyse multivariate data.

RA68 - Identify areas of application where techniques of intelligent systems can be used

RA88 - Choose and apply relevant concepts/methods and/or tools and collect relevant data for conducting a business research analysis in a real environment

RA49 - apply IA techniques in real data scenarios

RA75 - Understand the basics of the Knowledge Discovery Process, and its application to time series and complex data

* 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 Modeler tool

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

Topics:

Data Engineering, Data Mining, Business Intelligence

CRISP-DM, or the Data Mining Process

Techniques:

- Classification
- Regression
- Association
- Clustering

5.2. Syllabus

1. INTRODUCTION TO DATA ENGINEERING
2. THE TOOL: IBM SPSS MODELER
3. THE PROCESS CRISP-DM
4. LINEAR REGRESSION
5. LOGISTIC REGRESSION
6. RFM ANALYSIS
7. DECISION TREES
8. NEURAL NETWORKS
9. CLUSTERING
10. NEAREST NEIGHBOR
11. ASSOCIATION RULES

6. Schedule

6.1. Subject schedule*

Week	Classroom activities	Laboratory activities	Distant / On-line	Assessment activities
1	INTRODUCTION TO DATA ANALYTICS Duration: 02:00 Additional activities			
2				Supermarket Assignment Individual work Continuous assessment and final examination Presential Duration: 02:00
3	IBM SPSS: INTRO AND PRACTICE Duration: 02:00 Problem-solving class			
4				IBM SPSS EXERCICES Individual work Continuous assessment and final examination Presential Duration: 02:00
5	A PRACTICAL SESSION ON DESCRIPTIVE ANALYSIS Duration: 01:00 Problem-solving class			Descriptive Analysis Questions Individual work Continuous assessment and final examination Presential Duration: 01:00
6	Some Data Visualization Tools of IBM SPSS Duration: 01:00 Lecture			Visual Descriptive Analysis of Supermarket data Individual work Continuous assessment and final examination Presential Duration: 01:00
7	DESCRIPTIVE ANALYSIS: RFM Duration: 01:00 Lecture			Segmentation of supermarket customers Individual work Continuous assessment and final examination Presential Duration: 01:00
8	DESCRIPTIVE ANALYSIS: CLUSTERING Duration: 01:00 Lecture			Clustering of Supermarket Customers Individual work Continuous assessment and final examination Presential Duration: 01:00

9	DIAGNOSTIC ANALYTICS: CORRELATION, ANOVA AND CHI- SQUARED TESTS Duration: 01:00 Lecture			Titanic Survival Individual work Continuous assessment and final examination Presential Duration: 01:00 Basket Analysis Individual work Continuous assessment and final examination Presential Duration: 01:00
10	LINEAR REGRESSION Duration: 01:00 Lecture			Predicting purchases Individual work Continuous assessment and final examination Presential Duration: 01:00
11	LOGISTIC REGRESSION Duration: 01:00 Lecture			Touchpoints Individual work Continuous assessment and final examination Presential Duration: 01:00
12	DECISION TREES Duration: 01:00 Lecture Instance-Based Classifiers: NEAREST NEIGHBOR Duration: 01:00 Lecture			
13	NEURAL NETWORKS Duration: 01:00 Lecture ENSEMBLE METHODS Duration: 01:00 Lecture			
14				Competition: Predictive Analysis Group work Continuous assessment and final examination Presential Duration: 04:00
15				Competition: Predictive Analysis Group work Continuous assessment and final examination Presential Duration: 04:00
16	DEALING WITH TIME Duration: 01:00 Lecture ASSOCIATION RULES Duration: 01:00 Lecture			

17				
----	--	--	--	--

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.

* The schedule is based on an a priori planning of the subject; it might be modified during the academic year, especially considering the COVID19 evolution.

7. Activities and assessment criteria

7.1. Assessment activities

7.1.1. Assessment

Week	Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
2	Supermarket Assignment	Individual work	Face-to-face	02:00	5%	5 / 10	CG03 CG07 CE-FT01 CE-FT04 CE-FT05 CE-DIPO03 CE-DIPO02
4	IBM SPSS EXERCICES	Individual work	Face-to-face	02:00	5%	5 / 10	CG03 CG07 CE-FT01 CE-FT04 CE-FT05 CE-DIPO03 CE-DIPO02
5	Descriptive Analysis Questions	Individual work	Face-to-face	01:00	5%	5 / 10	CE-FT01 CE-FT04 CE-FT05 CE-DIPO03 CE-DIPO02 CG03 CG07
6	Visual Descriptive Analysis of Supermarket data	Individual work	Face-to-face	01:00	5%	5 / 10	CG03 CG07 CE-FT01 CE-FT04 CE-FT05 CE-DIPO03 CE-DIPO02
7	Segmentation of supermarket customers	Individual work	Face-to-face	01:00	10%	5 / 10	CG03 CG07 CE-FT01 CE-FT04 CE-FT05 CE-DIPO03 CE-DIPO02

8	Clustering of Supermarket Customers	Individual work	Face-to-face	01:00	10%	5 / 10	CG03 CG07 CE-FT01 CE-FT04 CE-FT05 CE-DIPO03 CE-DIPO02
9	Titanic Survival	Individual work	Face-to-face	01:00	10%	5 / 10	CG03 CG07 CE-FT01 CE-FT04 CE-FT05 CE-DIPO03 CE-DIPO02
9	Basket Analysis	Individual work	Face-to-face	01:00	10%	5 / 10	CG07 CE-FT01 CE-FT04 CE-FT05 CG03 CE-DIPO03 CE-DIPO02
10	Predicting purchases	Individual work	Face-to-face	01:00	10%	5 / 10	CG03 CG07 CE-FT01 CE-FT04 CE-FT05 CE-DIPO03 CE-DIPO02
11	Touchpoints	Individual work	Face-to-face	01:00	10%	5 / 10	CG03 CG07 CE-FT01 CE-FT04 CE-FT05 CE-DIPO03 CE-DIPO02
14	Competition: Predictive Analysis	Group work	Face-to-face	04:00	10%	5 / 10	CE-FT01 CE-FT04 CE-FT05 CE-DIPO03 CG03 CG07 CE-DIPO02
15	Competition: Predictive Analysis	Group work	Face-to-face	04:00	10%	5 / 10	CE-FT01 CE-FT04 CE-FT05 CE-DIPO03 CG03 CG07 CE-DIPO02

7.1.2. Global examination

Week	Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
2	Supermarket Assignment	Individual work	Face-to-face	02:00	5%	5 / 10	CG03 CG07 CE-FT01 CE-FT04 CE-FT05 CE-DIPO03 CE-DIPO02
4	IBM SPSS EXERCICES	Individual work	Face-to-face	02:00	5%	5 / 10	CG03 CG07 CE-FT01 CE-FT04 CE-FT05 CE-DIPO03 CE-DIPO02
5	Descriptive Analysis Questions	Individual work	Face-to-face	01:00	5%	5 / 10	CE-FT01 CE-FT04 CE-FT05 CE-DIPO03 CE-DIPO02 CG03 CG07
6	Visual Descriptive Analysis of Supermarket data	Individual work	Face-to-face	01:00	5%	5 / 10	CG03 CG07 CE-FT01 CE-FT04 CE-FT05 CE-DIPO03 CE-DIPO02
7	Segmentation of supermarket customers	Individual work	Face-to-face	01:00	10%	5 / 10	CG03 CG07 CE-FT01 CE-FT04 CE-FT05 CE-DIPO03 CE-DIPO02
8	Clustering of Supermarket Customers	Individual work	Face-to-face	01:00	10%	5 / 10	CG03 CG07 CE-FT01 CE-FT04 CE-FT05 CE-DIPO03 CE-DIPO02
9	Titanic Survival	Individual work	Face-to-face	01:00	10%	5 / 10	CG03 CG07 CE-FT01 CE-FT04 CE-FT05 CE-DIPO03 CE-DIPO02

9	Basket Analysis	Individual work	Face-to-face	01:00	10%	5 / 10	CG07 CE-FT01 CE-FT04 CE-FT05 CG03 CE-DIPO03 CE-DIPO02
10	Predicting purchases	Individual work	Face-to-face	01:00	10%	5 / 10	CG03 CG07 CE-FT01 CE-FT04 CE-FT05 CE-DIPO03 CE-DIPO02
11	Touchpoints	Individual work	Face-to-face	01:00	10%	5 / 10	CG03 CG07 CE-FT01 CE-FT04 CE-FT05 CE-DIPO03 CE-DIPO02
14	Competition: Predictive Analysis	Group work	Face-to-face	04:00	10%	5 / 10	CE-FT01 CE-FT04 CE-FT05 CE-DIPO03 CG03 CG07 CE-DIPO02
15	Competition: Predictive Analysis	Group work	Face-to-face	04:00	10%	5 / 10	CE-FT01 CE-FT04 CE-FT05 CE-DIPO03 CG03 CG07 CE-DIPO02

7.1.3. Referred (re-sit) examination

Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
-------------	----------	------	----------	--------	---------------	------------------

ALL ASSIGNMENTS	Individual presentation	Face-to-face	02:02	100%	5 / 10	CE-FT05 CE-FT04 CG03 CG07 CE-FT01
-----------------	-------------------------	--------------	-------	------	--------	---

7.2. Assessment criteria

The evaluation is based on the final project BUT ALL assignments are mandatory for all three types of examinations.

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	Others	THE TOOL WE WILL USE
Sala de trabajo en grupo con ordenadores	Equipment	
aula	Equipment	