



POLITÉCNICA

INTERNATIONAL
CAMPUS OF
EXCELLENCE

COORDINATION PROCESS OF
LEARNING ACTIVITIES
PR/CL/001



E.T.S. de Ingenieros
Informáticos

ANX-PR/CL/001-01

LEARNING GUIDE

SUBJECT

103000389 - Data Mining

DEGREE PROGRAMME

10AK - Master Universitario En Software Y Sistemas

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.....	4
6. Schedule.....	6
7. Activities and assessment criteria.....	8
8. Teaching resources.....	11

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	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

CEM2 - Analizar y sintetizar soluciones a problemas que requieran aproximaciones novedosas para la definición de la infraestructura computacional que permita el procesamiento y el análisis de datos de diversa naturaleza

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

CEM8 - Aplicar los fundamentos teóricos y matemáticos adecuados al procesamiento y análisis de funciones y datos de diversa naturaleza, y evaluar y diseñar los métodos relacionados para su aplicación en dominios prácticos

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.

CG12 - Comprensión amplia de las técnicas y métodos aplicables en una especialización concreta, así como de sus límites

CG13 - Apreciación de los límites del conocimiento actual y de la aplicación práctica de la tecnología más reciente.

CG14 - Conocimiento y comprensión de la informática necesaria para la creación de modelos de información, y de los sistemas y procesos complejos

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.

CG4 - Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo.

CG7 - Especificación y realización de tareas informáticas complejas, poco definidas o no familiares

CG8 - Planteamiento y resolución de problemas también en áreas nuevas y emergentes de su disciplina

CG9 - Aplicación de los métodos de resolución de problemas más recientes o innovadores y que puedan implicar el uso de otras disciplinas

CG120 - Adquirir conocimientos científicos avanzados del campo de la informática que le permitan generar nuevas ideas dentro de una línea de investigación.

CG123 - Capacidad de leer y comprender publicaciones dentro de su ámbito de estudio/investigación, así como su catalogación y valor científico

4.2. Learning outcomes

RA22 - Conocer ejemplos de aplicaciones reales y tendencias y líneas de investigación

RA20 - Manejar aplicaciones software para realizar tareas de data mining

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 ENGINEERING
2. THE TOOL: IBM SPSS MODELER
3. Descriptive, Diagnostic, Predictive and Prescriptive Analysis
4. RFM ANALYSIS
5. CLUSTERING
6. LINER REGRESSION
7. LOGISTIC REGRESSION
8. NEAREST NEIGHBOR
9. DECISION TREES
10. NEURAL NETWORKS
11. ENSEMBLE METHODS
12. ASSOCIATION RULES
13. DEALING WITH TIME

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 Presential Duration: 02:00
3	IBM SPSS: INTRO AND PRACTICE Duration: 02:00 Problem-solving class			
4				IBM SPSS EXERCICES Individual work Continuous assessment Presential Duration: 02:00
5	A PRACTICAL SESSION ON DESCRIPTIVE ANALYSIS Duration: 01:00 Problem-solving class			Descriptive Analysis Questions Individual work Continuous assessment 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 Presential Duration: 01:00
7	DESCRIPTIVE ANALYSIS: RFM Duration: 01:00 Lecture			Segmentation of supermarket customers Individual work Continuous assessment Presential Duration: 01:00
8	DESCRIPTIVE ANALYSIS: CLUSTERING Duration: 01:00 Lecture			Clustering of Supermarket Customers Individual work Continuous assessment Presential Duration: 01:00
9	DIAGNOSTIC ANALYTICS: CORRELATION, ANOVA AND CHI-SQUARED TESTS Duration: 01:00 Lecture			Titanic Survival Individual work Continuous assessment Presential Duration: 01:00 Basket Analysis Individual work Continuous assessment Presential Duration: 01:00

10	LINEAR REGRESSION Duration: 01:00 Lecture			Predicting purchases Individual work Continuous assessment Presential Duration: 01:00
11	LOGISTIC REGRESSION Duration: 01:00 Lecture			Touchpoints Individual work Continuous assessment 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 Individual work Continuous assessment Presential Duration: 04:00
15	DEALING WITH TIME Duration: 01:00 Lecture ASSOCIATION RULES Duration: 01:00 Lecture			
16				Competition: Predictive Analysis Individual work Continuous assessment Presential Duration: 04:00
17				ALL ASSIGNMENTS Individual work Final examination Presential Duration: 02:00

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	CEM7
4	IBM SPSS EXERCICES	Individual work	Face-to-face	02:00	5%	5 / 10	CEM7
5	Descriptive Analysis Questions	Individual work	Face-to-face	01:00	5%	5 / 10	CEM7
6	Visual Descriptive Analysis of Supermarket data	Individual work	Face-to-face	01:00	5%	5 / 10	CG7 CG12 CG13 CEM7
7	Segmentation of supermarket customers	Individual work	Face-to-face	01:00	10%	5 / 10	CG7 CG12 CG13 CEM7
8	Clustering of Supermarket Customers	Individual work	Face-to-face	01:00	10%	5 / 10	CG3 CG12 CEM7
9	Titanic Survival	Individual work	Face-to-face	01:00	10%	5 / 10	CG7 CG12 CG13 CEM7
9	Basket Analysis	Individual work	Face-to-face	01:00	10%	5 / 10	CG12 CEM7 CG3
10	Predicting purchases	Individual work	Face-to-face	01:00	10%	5 / 10	CG4 CG3 CG12 CG13 CEM7 CG1 CG7 CEM2

11	Touchpoints	Individual work	Face-to-face	01:00	10%	5 / 10	CG7 CG12 CG13 CEM7
14	Competition: Predictive Analysis	Individual work	Face-to-face	04:00	10%	5 / 10	CG7 CGI20 CEM2 CEM8 CGI23 CG1 CG9 CG4 CG3 CG8 CG12 CG13 CG14 CEM7
16	Competition: Predictive Analysis	Individual work	Face-to-face	04:00	10%	5 / 10	CG7 CGI20 CEM2 CEM8 CGI23 CG1 CG9 CG4 CG3 CG8 CG12 CG13 CG14 CEM7

7.1.2. Global examination

Week	Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
17	ALL ASSIGNMENTS	Individual work	Face-to-face	02:00	100%	5 / 10	CG7 CGI20 CEM2 CEM8 CGI23 CG1 CG9 CG4 CG3 CG8 CG12 CG13 CG14 CEM7

7.1.3. Referred (re-sit) examination

Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
ALL ASSIGNMENTS	Individual presentation	Face-to-face	02:00	100%	5 / 10	CG7 CGI20 CEM2 CEM8 CGI23 CG1 CG9 CG4 CG3 CG8 CG12 CG13 CG14 CEM7

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	
------	-----------	--