



INTERNATIONAL
CAMPUS OF
EXCELLENCE

COORDINATION PROCESS OF
LEARNING ACTIVITIES
PR/CL/001



E.T.S. de Ingenieros
Informaticos

ANX-PR/CL/001-01

LEARNING GUIDE

SUBJECT

103000389 - Concepts, Practice And Research Challenges In Data Mining

DEGREE PROGRAMME

10AK - Master Universitario En Software Y Sistemas

ACADEMIC YEAR & SEMESTER

2021/22 - 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.....	11
9. Other information.....	12

1. Description

1.1. Subject details

Name of the subject	103000389 - Concepts, Practice And Research Challenges In 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	2021-22

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

- Descubrimiento De Conocimiento En Bases De Datos

3.2. Other recommended learning outcomes

- Statistics

- Artificial Intelligence

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

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.

CG10 - Capacidad de pensamiento creativa con el objetivo de desarrollar enfoques y métodos nuevos y originales.

CG16 - Capacidad de trabajar de forma independiente en su campo profesional.

CG18 - Capacidad de trabajar y comunicarse también en contextos internacionales.

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

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

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

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 Modeler tool

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. LINER 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	Face-to-face classroom activities	Face-to-face laboratory activities	Distant / On-line	Assessment activities
1	INTRODUCTION TO DATA ENGINEERING Duration: 02:00 Lecture		INTRODUCTION TO DATA ENGINEERING Duration: 02:00 Lecture	
2	THE PROCESS OF DATA MINING Duration: 01:00 Lecture		Tema 2 Duration: 01:00 Problem-solving class THE PROCESS OF DATA MINING Duration: 01:00 Lecture	
3	Data Understanding and Data Preparation with IBM SPSS Modeler I Duration: 01:00 Problem-solving class		Data Understanding and Data Preparation with IBM SPSS Modeler I Duration: 02:00 Problem-solving class	
4	Data Understanding and Data Preparation with IBM SPSS Modeler II Duration: 01:00 Problem-solving class		Data Understanding and Data Preparation with IBM SPSS Modeler II Duration: 02:00 Problem-solving class	
5	LINEAR REGRESSION Duration: 01:00 Lecture		LINEAR REGRESSION Duration: 02:00 Problem-solving class LINEAR REGRESSION Duration: 01:00 Lecture	
6	Modeling with IBM SPSS Modeler I Duration: 01:00 Problem-solving class		Modeling with IBM SPSS Modeler I Duration: 02:00 Problem-solving class	ASSIGNMENT 1 Group work Continuous assessment and final examination Not Presential Duration: 02:00
7	LOGISTIC REGRESSION Duration: 01:00 Lecture		LOGISTIC REGRESSION Duration: 02:00 Problem-solving class LOGISTIC REGRESSION Duration: 01:00 Lecture	
8	Modeling with IBM SPSS Modeler II Duration: 01:00 Problem-solving class		Modeling with IBM SPSS Modeler II Duration: 02:00 Problem-solving class	ASSIGNMENT 2 Group work Continuous assessment and final examination Presential Duration: 02:00

9	DECISION TREES Duration: 01:00 Lecture		DECISION TREES Duration: 02:00 Problem-solving class DECISION TREES Duration: 01:00 Lecture	
10	RFM ANALYSIS Duration: 01:00 Lecture		RFM ANALYSIS Duration: 02:00 Problem-solving class RFM ANALYSIS Duration: 01:00 Lecture	
11	NEURAL NETWORKS Duration: 01:00 Lecture		NEURAL NETWORKS Duration: 02:00 Problem-solving class NEURAL NETWORKS Duration: 01:00 Lecture	ASSIGNMENT 3 Group work Continuous assessment and final examination Presential Duration: 02:00
12	CLUSTERING Duration: 01:00 Lecture		CLUSTERING Duration: 02:00 Problem-solving class CLUSTERING Duration: 01:00 Lecture	
13	NEAREST NEIGHBOR Duration: 01:00 Lecture		NEAREST NEIGHBOR Duration: 02:00 Problem-solving class NEAREST NEIGHBOR Duration: 01:00 Lecture	
14	ASSOCIATION RULES Duration: 01:00 Lecture		ASSOCIATION RULES Duration: 02:00 Problem-solving class ASSOCIATION RULES Duration: 01:00 Lecture	ASSIGNMENT 4 Group work Continuous assessment and final examination Presential Duration: 02:00
15			EXERCISES Duration: 01:00 Problem-solving class	ASSIGNMENT 5 Group work Continuous assessment and final examination Presential Duration: 02:00
16				
17				FINAL PROJECT Individual presentation Continuous assessment and 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. Continuous assessment

Week	Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
6	ASSIGNMENT 1	Group work	No Presential	02:00	10%	5 / 10	CG1 CG3 CG10 CG16 CG18 CEM8 CEM7
8	ASSIGNMENT 2	Group work	Face-to-face	02:00	10%	5 / 10	CG16 CG1 CG10 CG18 CG3 CEM8 CEM7
11	ASSIGNMENT 3	Group work	Face-to-face	02:00	10%	5 / 10	CG1 CG3 CG10 CG16 CG18 CEM8 CEM7
14	ASSIGNMENT 4	Group work	Face-to-face	02:00	10%	5 / 10	CG3 CG10 CG16 CG18 CG1 CEM8 CEM7
15	ASSIGNMENT 5	Group work	Face-to-face	02:00	10%	5 / 10	CG1 CG3 CG10 CG16 CG18 CEM8 CEM7

17	FINAL PROJECT	Individual presentation	Face-to-face	02:00	50%	5 / 10	CG10 CG16 CG18 CEM8 CG1 CG3 CEM7
----	---------------	-------------------------	--------------	-------	-----	--------	--

7.1.2. Final examination

Week	Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
6	ASSIGNMENT 1	Group work	No Presential	02:00	10%	5 / 10	CG1 CG3 CG10 CG16 CG18 CEM8 CEM7
8	ASSIGNMENT 2	Group work	Face-to-face	02:00	10%	5 / 10	CG16 CG1 CG10 CG18 CG3 CEM8 CEM7
11	ASSIGNMENT 3	Group work	Face-to-face	02:00	10%	5 / 10	CG1 CG3 CG10 CG16 CG18 CEM8 CEM7
14	ASSIGNMENT 4	Group work	Face-to-face	02:00	10%	5 / 10	CG3 CG10 CG16 CG18 CG1 CEM8 CEM7
15	ASSIGNMENT 5	Group work	Face-to-face	02:00	10%	5 / 10	CG1 CG3 CG10 CG16 CG18 CEM8 CEM7

17	FINAL PROJECT	Individual presentation	Face-to-face	02:00	50%	5 / 10	CG10 CG16 CG18 CEM8 CG1 CG3 CEM7
----	---------------	-------------------------	--------------	-------	-----	--------	--

7.1.3. Referred (re-sit) examination

Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
FINAL PROJECT AND ASSIGNMENTS	Individual presentation	Face-to-face	02:02	100%	5 / 10	CG10 CG16 CG18 CEM8 CEM7 CG1 CG3

7.2. Assessment criteria

The evaluation is based on the assignments and the final project.

Assignments and projects will be performed individually or by groups, depending on the size of the course

To pass the course it is mandatory to present all the assignments and the final project, in any modality of evaluation

Participation in class would give a 10% increase in the final score.

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
Sala de trabajo en grupo con ordenadores	Equipment	

aula	Equipment	
------	-----------	--

9. Other information

9.1. Other information about the subject

We will use Teams for communication and collaboration