



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

**103000853 - Data Processes**

### DEGREE PROGRAMME

10AZ - Master Universitario En Innovación Digital

### ACADEMIC YEAR & SEMESTER

2023/24 - Semester 1

## Index

---

### Learning guide

1. Description.....	1
2. Faculty.....	1
3. Prior knowledge recommended to take the subject.....	2
4. Skills and learning outcomes .....	3
5. Brief description of the subject and syllabus.....	4
6. Schedule.....	7
7. Activities and assessment criteria.....	9
8. Teaching resources.....	12
9. Other information.....	13

## 1. Description

---

### 1.1. Subject details

<b>Name of the subject</b>	103000853 - Data Processes
<b>No of credits</b>	4.5 ECTS
<b>Type</b>	Optional
<b>Academic year of the programme</b>	First year
<b>Semester of tuition</b>	Semester 1
<b>Tuition period</b>	September-January
<b>Tuition languages</b>	English
<b>Degree programme</b>	10AZ - Master Universitario en Innovación Digital
<b>Centre</b>	10 - Escuela Tecnica Superior De Ingenieros Informaticos
<b>Academic year</b>	2023-24

## 2. Faculty

---

### 2.1. Faculty members with subject teaching role

<b>Name and surname</b>	<b>Office/Room</b>	<b>Email</b>	<b>Tutoring hours *</b>
Luis Mengual Galan		luis.mengual@upm.es	Sin horario. contact the professor
Antonio Jesus Diaz Honrubia	4303	antoniojesus.diaz@upm.es	Sin horario. contact the professor

Maria Covadonga Fernandez Baizan		mariacovadonga.fernandez @upm.es	Sin horario. contact the professor
Ernestina Menasalvas Ruiz (Subject coordinator)	4303	ernestina.menasalvas@upm. es	Sin horario. contact the professor
Alejandro Rodriguez Gonzalez	4302	alejandro.rg@upm.es	Sin horario. contact the professor

\* 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

- SQL
- Databases
- relational data model

## 4. Skills and learning outcomes \*

---

### 4.1. Skills to be learned

CB06 - Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en el desarrollo y/o aplicación de ideas, a menudo en un contexto de investigación

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

CB08 - Que los estudiantes sean capaces de integrar conocimientos y enfrentarse a la complejidad de formular juicios a partir de una información que, siendo incompleta o limitada, incluya reflexiones sobre las responsabilidades sociales y éticas vinculadas a la aplicación de sus conocimientos y juicios

CB09 - 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

CB10 - 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.

CE-CD04 - Capacidad para aplicar métodos avanzados para clasificación, modelado, segmentación y predicción a partir de un conjunto de datos

CE-CD07 - Capacidad para diseñar y gestionar proyectos de ciencia de datos

CG06 - Capacidad para gestionar la información.

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

CG09 - La capacidad de transformar las experiencias prácticas en problemas y desafíos de investigación.

## 4.2. Learning outcomes

RA56 - develop data science projects following a methodology

RA49 - apply IA techniques in real data scenarios

\* 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

In this course we will deepen on the importance of data for an organization. In fact, the course is centered on the process of extraction of knowledge from databases as a support for decision making.

The planning and management of a project of data science project is the main goal of this course.

Consequently the course will start with the analysis of data sources in an organization and very briefly analyzed data base management systems. Emphasis will be put on the ethical aspects regarding data management and knowledge extraction.

Later students will understand the data value chain and will go deep into the process of knowledge extraction. At this stage CRISP-Dm methodology will be used.

The course will follow on the different phases of the process: i) business understanding, ii) data understanding, iii) data preparation, iv) modeling v) evaluation and vi) deployment.

Through all the phases the main emphasis will be on students getting hands on the different steps, techniques, algorithms and tools.

The course will end with use cases in different domains.

## 5.2. Syllabus

### 1. Introduction

- 1.1. Course description.
- 1.2. Data Science and Data Scientist Skills.
- 1.3. The Value hidden in data.

### 2. Operational Data bases Vs Decisional databases

- 2.1. The BIG Data Value Chain.
- 2.2. Data Warehouse.
- 2.3. Data Lakes

### 3. The process of Knowledge Discovery in Databases

- 3.1. CRISP-DM

### 4. Business Understanding

- 4.1. Goal of BU.
- 4.2. Planning of a DataScience project.

### 5. Data Understanding

- 5.1. Understanding data.
- 5.2. Nulls and outliers detection.
- 5.3. Correlation analysis

### 6. Data Preparation

- 6.1. Preparing data for mining: dealing with problems encountered in understanding, transforming data, discretization, data reduction, agregation, ?.

### 7. Data mining/data modeling

- 7.1. Type of problems. Data nature , data problems and possible algorithms.
- 7.2. Classification, association and clustering
- 7.3. Complex Networks

### 8. Evaluation and Deployment

- 8.1. Evaluation of the models.
- 8.2. Deployment of the models

## 9. Ethics

### 9.1. GDPR and implications in Data Science



## 6. Schedule

### 6.1. Subject schedule\*

Week	Classroom activities	Laboratory activities	Distant / On-line	Assessment activities
1	<b>Unit 1</b> Duration: 03:00 Lecture			
2	<b>Unit 2</b> Duration: 03:00 Lecture			
3	<b>Unit 3</b> Duration: 03:00 Lecture			
4	<b>Unit 4</b> Duration: 03:00 Lecture			
5	<b>Unit 5</b> Duration: 04:00 Lecture			
6	<b>Unit 6</b> Duration: 04:00 Lecture			
7	<b>Unit 6</b> Duration: 02:00 Lecture			
8	<b>Unit 6</b> Duration: 01:00 Lecture			
9	<b>Unit 7</b> Duration: 02:00 Lecture			
10	<b>Unit 7</b> Duration: 02:00 Lecture			<b>Evaluation First Assigment</b> Group presentation Continuous assessment Presential Duration: 02:00
11	<b>Unit 7</b> Duration: 02:00 Lecture			
12	<b>Unit 8</b> Duration: 02:00 Lecture			
13	<b>Unit 8</b> Duration: 02:00 Lecture			

14	<b>Unit 9</b> Duration: 02:00 Lecture			
15				<b>Evaluation Second Assigment</b> Group presentation Continuous assessment Presential Duration: 02:00
16				<b>Exam</b> Written test Continuous assessment Presential Duration: 02:00  <b>Exam. This will include the recoverable part of the assigments.</b> Written test Final examination Presential Duration: 02:00
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
10	Evaluation First Assigment	Group presentation	Face-to-face	02:00	25%	4 / 10	CB06 CB07 CB08 CB09 CB10 CG06 CG07 CG09 CE-CD04 CE-CD07
15	Evaluation Second Assigment	Group presentation	Face-to-face	02:00	25%	4 / 10	CB06 CB07 CB08 CB09 CB10 CG06 CG07 CG09 CE-CD04 CE-CD07
16	Exam	Written test	Face-to-face	02:00	50%	4 / 10	CB06 CB07 CB08 CB09 CB10 CG06 CG07 CG09 CE-CD04 CE-CD07

#### 7.1.2. Global examination

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

16	Exam. This will include the recoverable part of the assignments.	Written test	Face-to-face	02:00	100%	5 / 10	CB06 CB07 CB08 CB09 CB10 CG06 CG07 CG09 CE-CD04 CE-CD07
----	--	--------------	--------------	-------	------	--------	--

### 7.1.3. Referred (re-sit) examination

Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
Exam in July	Written test	Face-to-face	02:00	100%	5 / 10	CB06 CB07 CB08 CB09 CB10 CG06 CG07 CG09 CE-CD04 CE-CD07

## 7.2. Assessment criteria

The course will be evaluated by:

- 2 practical assignments: They are recoverable. This means that the student will have an exam of the assignment in the exam in case they did not follow them along the course.
- 1 exam

Practical assignments will be done in groups among those enrolled in the course at the beginning of the academic year (nature and number of components will be established at the beginning of the course, depending on the number of students enrolled).

In order to pass the course in the fall semester the requirements are:

- 1. To obtain a minimum of 50 points out of 100 in the added evaluation.
- 2. It is MANDATORY to do the exam and do the practical assignment.
- 3. In the exam and on the practical assignments students must obtain a minimum of 40%.

Final score will be calculated as follows:

- 50% Practical assignment (divided between the two assignments)
- 50% Final exam

ts it is appreciated a breach in the duties as college student, the subject coordinator may communicate the headmaster as established in the article 74 (n) of EUPM to have the competences to "propose the initiation of a disciplinary procedure to any College member, by its own initiative or as instance from the "Comisión de Gobierno" to the Rector, pursuant to the statutes and rules of application.

### Action procedures against fraudulent behavior

All exams and lab deliverables must be done individually, unless specified otherwise by the teaching staff. If any non-conformity with this criteria is detected (copy, plagiarism, etc.), according to article 13 of the assessment regulations, it will be considered as academic fraud.

According to this same article, all students involved in the fraud will be graded with 0 points (out of 10) in the corresponding call and professors may prepare a special exam only for them in the next official call to assess the learning results of the course.

To this end, all students involved in the fraud are considered accountable, and the above actions will be taken against both active and passive agents.

If a student is involved in a fraudulent behaviour, their marks will not be kept for the extraordinary call. Additionally, no marks will be kept between calls of future academic years.

The above actions do not exclude other actions covered by the normative and laws.

## 8. Teaching resources

---

### 8.1. Teaching resources for the subject

Name	Type	Notes
Moodle	Web resource	<a href="http://moodle.upm.es">http://moodle.upm.es</a>
Data Mining book	Bibliography	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.
Smart Machines book	Bibliography	Smart Machines: IBM's Watson and the Era of Cognitive Computing. Columbia University Press (October 15, 2013)
Database Systems: The Complete Book (DS:CB), by Hector Garcia-Molina, Jeff Ullman, and Jennifer Widom	Bibliography	
Pang-Ning Tan, Michael Steinbach, Vipin Kumar, Introduction to Data Mining, Pearson Addison Wesley (May, 2005).	Bibliography	

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

## 9. Other information

---

### 9.1. Other information about the subject