



POLITÉCNICA

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

**103000895 - Data Processes**

### DEGREE PROGRAMME

10BA - Master Universitario En Ciencia De Datos

### ACADEMIC YEAR & SEMESTER

2021/22 - 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 .....	2
5. Brief description of the subject and syllabus.....	4
6. Schedule.....	6
7. Activities and assessment criteria.....	9
8. Teaching resources.....	11

## 1. Description

---

### 1.1. Subject details

<b>Name of the subject</b>	103000895 - Data Processes
<b>No of credits</b>	4.5 ECTS
<b>Type</b>	Compulsory
<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>	10BA - Master Universitario en Ciencia de Datos
<b>Centre</b>	10 - Escuela Tecnica Superior De Ingenieros Informaticos
<b>Academic year</b>	2021-22

## 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>
Antonio Jesus Diaz Honrubia	4302	antoniojesus.diaz@upm.es	Sin horario. contact the professor
Alejandro Rodriguez Gonzalez (Subject coordinator)	4302	alejandro.rg@upm.es	Sin horario. contact the professor

Ernestina Menasalvas Ruiz	4303	ernestina.menasalvas@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.

### 2.3. External faculty

Name and surname	Email	Institution
Lucia Prieto Santamaría	lucia.prieto.santamaria@upm.es	Centro de Tecnología Biomédica

## 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

- Databases
- relational data model
- SQL

## 4. Skills and learning outcomes \*

---

### 4.1. Skills to be learned

CECD01 - Conocer los procesos de captura, extracción, manipulación y conversión de datos en diferentes entornos.

CECD02 - Conocer los principales sistemas de almacenamiento de la información estructurada y no estructurada de fuentes heterogéneas.

CECD03 - Manejar las herramientas informáticas para Big Data

CECD04 - Capacidad para aplicar técnicas para la generación de visualizaciones adecuadas a cada problema

para el análisis y la exploración de datos, y para la correcta comunicación de los resultados del análisis.

CECD06 - Conocer y tener capacidad para aplicar métodos de minería de datos para clasificación, modelado, segmentación y predicción a partir de un conjunto de datos

CECD08 - Capacidad para diseñar y gestionar proyectos en el ámbito del análisis de datos

CECD09 - Capacidad para actuar con los principios éticos y legales relacionados con la manipulación de datos según el ámbito de aplicación

CG07 - Aplicación de los últimos o más novedosos métodos para resolver problemas que, posiblemente, involucren a otras disciplinas

## 4.2. Learning outcomes

RA22 - Poseer destrezas fundamentales de la programación que permitan la implementación de algoritmos y el uso de estructuras de datos típicos en ciencia de datos. e distintos tipos de herramientas (software o metodológicas y conceptuales) necesarias para el correcto y eficaz desarrollo de software, incluyendo entornos y librerías en el contexto de ciencia de datos.

RA23 - Destreza en el uso de distintos tipos de herramientas (software o metodológicas y conceptuales) necesarias para el correcto y eficaz desarrollo de software, incluyendo entornos y librerías en el contexto de ciencia de datos.

RA20 - Ser capaz de procesar datos masivos

\* 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.

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

#### 10. Extraction of knowledge from data in different domains

10.1. Data Science in a specific domain

## 6. Schedule

### 6.1. Subject schedule\*

Week	Face-to-face classroom activities	Face-to-face laboratory activities	Distant / On-line	Assessment activities
1	<b>Unit 1</b> Duration: 02:00		<b>Unit 1</b> Duration: 02:00	
2	<b>Unit 2</b> Duration: 01:00  <b>Unit 2</b> Duration: 01:00		<b>Unit 2</b> Duration: 01:00  <b>Unit 2</b> Duration: 01:00	
3	<b>Unit 3</b> Duration: 01:00  <b>Unit 3</b> Duration: 01:00		<b>Unit 3</b> Duration: 01:00  <b>Unit 3</b> Duration: 01:00	
4	<b>Unit 4</b> Duration: 01:00  <b>Unit 4</b> Duration: 01:00		<b>Unit 4</b> Duration: 01:00  <b>Unit 4</b> Duration: 01:00	
5	<b>Unit 5</b> Duration: 01:00  <b>Unit 5</b> Duration: 01:00		<b>Unit 5</b> Duration: 01:00  <b>Unit 5</b> Duration: 01:00	
6	<b>Unit 6</b> Duration: 01:00  <b>Unit 6</b> Duration: 01:00		<b>Unit 6</b> Duration: 01:00  <b>Unit 6</b> Duration: 01:00	
7	<b>Unit 6</b> Duration: 01:00  <b>Unit 6</b> Duration: 01:00		<b>Unit 6</b> Duration: 01:00  <b>Unit 6</b> Duration: 01:00	



8	<p><b>Unit 6</b> Duration: 01:00</p> <p><b>Unit 6</b> Duration: 01:00</p>		<p><b>Unit 6</b> Duration: 01:00</p> <p><b>Unit 6</b> Duration: 01:00</p>	
9	<p><b>Unit 7</b> Duration: 01:00</p> <p><b>Unit 7</b> Duration: 01:00</p>		<p><b>Unit 7</b> Duration: 01:00</p> <p><b>Unit 7</b> Duration: 01:00</p>	
10	<p><b>Unit 7</b> Duration: 01:00</p> <p><b>Unit 7</b> Duration: 01:00</p>		<p><b>Unit 7</b> Duration: 01:00</p> <p><b>Unit 7</b> Duration: 01:00</p>	<p><b>Evaluation First Assignment</b></p> <p>Continuous assessment Presential Duration: 02:00</p>
11	<p><b>Unit 7</b> Duration: 01:00</p> <p><b>Unit 7</b> Duration: 01:00</p>		<p><b>Unit 7</b> Duration: 01:00</p> <p><b>Unit 7</b> Duration: 01:00</p>	
12	<p><b>Unit 8</b> Duration: 01:00</p> <p><b>Unit 8</b> Duration: 01:00</p>		<p><b>Unit 8</b> Duration: 01:00</p> <p><b>Unit 8</b> Duration: 01:00</p>	
13	<p><b>Unit 8</b> Duration: 01:00</p> <p><b>Unit 8</b> Duration: 01:00</p>		<p><b>Unit 8</b> Duration: 01:00</p> <p><b>Unit 8</b> Duration: 01:00</p>	
14	<p><b>Unit 9</b> Duration: 02:00</p>		<p><b>Unit 9</b> Duration: 02:00</p>	
15	<p><b>Unit 10</b> Duration: 02:00</p>		<p><b>Unit 10</b> Duration: 02:00</p>	
16				<p><b>Evaluation Second Assignment</b></p> <p>Continuous assessment Presential Duration: 02:00</p>

17				<b>Exam</b>  Continuous assessment Presential Duration: 02:00  <b>Exam</b>  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
10	Evaluation First Assigment		Face-to-face	02:00	25%	4 / 10	CG07 CECD06 CECD02 CECD01 CECD08 CECD09
16	Evaluation Second Assigment		Face-to-face	02:00	25%	4 / 10	CECD03 CECD04
17	Exam		Face-to-face	02:00	50%	4 / 10	CECD03 CG07 CECD06 CECD02 CECD01 CECD04 CECD08 CECD09

#### 7.1.2. Final examination

Week	Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
17	Exam		Face-to-face	02:00	100%	5 / 10	CECD03 CG07 CECD06 CECD02 CECD01 CECD04 CECD08 CECD09

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

Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
Exam in July		Face-to-face	02:00	100%	5 / 10	CECD03 CG07 CECD06 CECD02 CECD01 CECD04 CECD08 CECD09

## 7.2. Assessment criteria

The "only final exam" evaluation will only be offered if the UPM "Normativa Reguladora de los Sistemas de Evaluación" requires so in the academic year 2015-2016, and the procedure to opt-in will be as stated by the head of studies.

**Continuous evaluation:** The course will be evaluated by:

- 2 practical assignments,
- 1 exam

Practical assignments will be done in groups among those enrolled in the course at the beginning of the academic year (nature and the 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%.

The final score will be calculated as follows:

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

## Final exam evaluation

Those students whose extraordinary circumstances cannot perform the continuous evaluation, and having done the final exam evaluation written petition during the first 15 days of the course, will perform the final exam evaluation without having the opportunity to do the continuous evaluation;

In these premises, the final exam evaluation will consist of an exam as stated by the head of studies.

Measures against copies and fraud Rights and duties of college students are gathered on the statutes of the Universidad Politécnica de Madrid (BOCM de 15 de Noviembre de 2010) and in the statutes of the college student (RD 1791/2010 de 30 de Diciembre). Article 124 a) of EUPM fixes the duty of the student... "to follow with responsibility and taking advantage of the learning process, knowledge acquisition correspondent to its condition of college student"... and the article 13 of the statutes of the college student in its point d) also specifies as duty of the college student "abstain from the use or cooperation in fraudulent procedures in the evaluation assessments, in the assignments developed or in the official documents of the university". In the case that in the development of the evaluation assessments 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.

## 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)
IBM Watson - How it works	Web resource	<a href="http://www.youtube.com/watch?v=_Xcmh1LQB9I">http://www.youtube.com/watch?v=_Xcmh1LQB9I</a>
Database Systems: The Complete Book (DS:CB), by Hector Garcia-Molina, Jeff Ullman, and Jennifer Widom	Bibliography	
"MySQL Administrator's Bible". Sheeri K. Cabral and Keith Murphy. Wiley	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	