



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

**103000865 - Data Mining And Time Series**

### DEGREE PROGRAMME

10AZ - Master Universitario En Innovación Digital

### ACADEMIC YEAR & SEMESTER

2022/23 - Semester 1

## Index

---

### Learning guide

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

## 1. Description

---

### 1.1. Subject details

<b>Name of the subject</b>	103000865 - Data Mining And Time Series
<b>No of credits</b>	3 ECTS
<b>Type</b>	Optional
<b>Academic year of the programme</b>	Second year
<b>Semester of tuition</b>	Semester 3
<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>	2022-23

## 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>
Juan Pedro Caraca-Valente Hernandez (Subject coordinator)	D4301	juanpedro.caracavalente@upm.es	Tu - 09:00 - 12:00 Th - 10:00 - 13:00
Aurora Perez Perez	D4301	aurora.perez@upm.es	M - 10:30 - 13:30 Th - 10:30 - 13:30

\* The tutoring schedule is indicative and subject to possible changes. Please check tutoring times with the faculty member in charge.

## 3. Skills and learning outcomes \*

---

### 3.1. Skills to be learned

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

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

CG02 - Que los estudiantes desarrollen la autonomía suficiente para participar en proyectos de investigación y colaboraciones científicas o tecnológicas dentro su ámbito temático explorando y generando nuevas ideas sistemáticamente, en contextos interdisciplinares y, en su caso, con una alta componente de transferencia del conocimiento.

CG03 - La capacidad de usar la lengua inglesa de manera competente, es decir, con capacitación para tareas complejas de trabajo y estudio.

### 3.2. Learning outcomes

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

RA76 - Know how to select and apply the best candidate techniques to Time Series Data Mining project

\* 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.

## 4. Brief description of the subject and syllabus

---

### 4.1. Brief description of the subject

This course will be taught in English primarily, but can be changed to Spanish if all the students prefer that option.

Knowledge Discovery techniques (or Data Mining) in large volumes of information are widely used today in different domains such as medicine, banking environments, industrial systems, etc. with a wide variety of applications such as data analysis, fraud detection, risk analysis, marketing campaigns, etc.

In this course all the stages of the Knowledge Discovery process will be reviewed and the most important techniques for each stage will be listed. Emphasis will be placed on techniques for data cleaning and preprocessing that, despite their importance, are often forgotten.

Next, the main techniques of Data Mining including Classification and Clustering techniques will be addressed. Some more recent methods for Data Analysis, including Deep Learning Techniques will also be covered.

In this subject we also want to explore areas of Knowledge Discovery less known, but equally important. There are domains where information is presented mostly in the form of Time Series which require a very specialized treatment. Examples of these are medical domains such as Electrocardiography or Audiometry, financial domains, etc. Time series are a challenge to the traditional techniques of Data Mining and often require the use of novel solutions. Special emphasis will be made on Temporal Abstraction techniques.

### 4.2. Syllabus

#### 1. Introduction

##### 1.1. Data Types, Time Series

##### 1.2. Basic Concepts

#### 2. Knowledge Discovery Process

##### 2.1. Knowledge Discovery Process Stages

##### 2.2. Data Preprocessing for basic data types and time series

#### 3. KDD Tools

##### 3.1. Background

##### 3.2. A KDD Tool: WEKA

#### 4. Data Mining Techniques

##### 4.1. Classification

##### 4.2. Advanced Methods for Data Analysis

##### 4.3. Clustering

##### 4.4. Time Series Techniques

#### 5. Evaluation

##### 5.1. Objectives

##### 5.2. Evaluation Techniques

## 5. Schedule

### 5.1. Subject schedule\*

Week	Classroom activities	Laboratory activities	Distant / On-line	Assessment activities
1	<b>1. Introduction</b> Duration: 02:00			
2	<b>2.1 Knowledge Discovery Process</b> Duration: 00:45  <b>2.2 Data Preprocessing for basic data types and time series</b> Duration: 01:00			<b>Progressive Evaluation Activity</b>  Continuous assessment Presential Duration: 00:15
3	<b>3 KDD Tools</b> Duration: 00:20  <b>3.1 Background and 3.2 WEKA</b> Duration: 01:25			<b>Progressive Evaluation Activity</b>  Continuous assessment Presential Duration: 00:15
4	<b>3.2 Case Study: WEKA</b> Duration: 01:00  <b>Domain Analysis and KDD Process</b> Duration: 00:45			<b>Progressive Evaluation Activity</b>  Continuous assessment Presential Duration: 00:15
5	<b>4.1 Classification Techniques</b> Duration: 01:45			<b>Progressive Evaluation Activity</b>  Continuous assessment Presential Duration: 00:15
6	<b>4.1 Classification Techniques</b> Duration: 01:45			<b>Progressive Evaluation Activity</b>  Continuous assessment Presential Duration: 00:15
7	<b>4.2 Advanced Methods for Data Analysis</b> Duration: 01:45  <b>Case Study: Advanced Methods</b> Duration: 00:15			<b>Project Stage 1: Domain Analysis, Data study, Objective definition</b>  Continuous assessment Not Presential Duration: 00:20

8	4.2 Clustering Techniques Duration: 01:45			<b>Progressive Evaluation Activity</b>  Continuous assessment Presential Duration: 00:15
9	4.2 Clustering Techniques Duration: 01:45			<b>Progressive Evaluation Activity</b>  Continuous assessment Presential Duration: 00:15
10	4.3 Time Series Data Mining Duration: 01:45			<b>Progressive Evaluation Activity</b>  Continuous assessment Presential Duration: 00:15
11	4.3 Time Series Data Mining Duration: 01:45			<b>Progressive Evaluation Activity</b>  Continuous assessment Presential Duration: 00:15
12	4.3 Time Series Data Mining Duration: 01:45  <b>Case Study: Time Series Data Mining</b> Duration: 00:15			<b>Project Stage 2: Application of Data Mining Techniques</b>  Continuous assessment Not Presential Duration: 00:20
13	4.3 Time Series Data Mining Duration: 01:45			<b>Progressive Evaluation Activity</b>  Continuous assessment Presential Duration: 00:15
14	5 Evaluation Duration: 01:45  <b>Group Discussion</b> Duration: 00:15			
15				<b>Project Stage 3: Evaluation</b>  Continuous assessment Not Presential Duration: 00:20  <b>Project Presentation</b>  Continuous assessment Presential Duration: 02:00
16				
17				<b>Project complete</b>  Final examination Not Presential Duration: 01:00

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.

## 6. Activities and assessment criteria

### 6.1. Assessment activities

#### 6.1.1. Assessment

Week	Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
2	Progressive Evaluation Activity		Face-to-face	00:15	3%	/ 10	CE-CD04 CB07 CG03
3	Progressive Evaluation Activity		Face-to-face	00:15	3%	/ 10	CE-CD04 CB07 CG03
4	Progressive Evaluation Activity		Face-to-face	00:15	3%	/ 10	CE-CD04 CB07 CG03
5	Progressive Evaluation Activity		Face-to-face	00:15	3%	/ 10	CE-CD04 CB07 CG02 CG03
6	Progressive Evaluation Activity		Face-to-face	00:15	3%	/ 10	CG02 CG03 CE-CD04 CB07
7	Project Stage 1: Domain Analysys, Data study, Objective definition		No Presential	00:20	10%	/ 10	CG02 CG03 CE-CD04 CB07
8	Progressive Evaluation Activity		Face-to-face	00:15	3%	/ 10	CE-CD04 CB07 CG02 CG03
9	Progressive Evaluation Activity		Face-to-face	00:15	3%	/ 10	CE-CD04 CB07 CG02 CG03
10	Progressive Evaluation Activity		Face-to-face	00:15	3%	/ 10	CE-CD04 CB07 CG02 CG03

11	Progressive Evaluation Activity		Face-to-face	00:15	3%	/ 10	CE-CD04 CB07 CG02 CG03
12	Project Stage 2: Application of Data Mining Techniques		No Presential	00:20	10%	/ 10	CE-CD04 CB07 CG02 CG03
13	Progressive Evaluation Activity		Face-to-face	00:15	3%	/ 10	CE-CD04 CB07 CG02 CG03
15	Project Stage 3: Evaluation		No Presential	00:20	10%	/ 10	CE-CD04 CB07 CG02 CG03
15	Project Presentation		Face-to-face	02:00	40%	/ 10	CB07 CG02 CG03 CE-CD04

### 6.1.2. Global examination

Week	Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
17	Project complete		No Presential	01:00	100%	5 / 10	CE-CD04 CB07 CG02 CG03

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

Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
Project complete		Face-to-face	00:00	100%	5 / 10	

## 6.2. Assessment criteria

For the evaluation of this subject we will take into account, on the one hand, Progressive Evaluation Activities (PEA) that will be carried out daily in class and, on the other hand, the Data Mining Project.

### PEA

The PEA consist on small daily activities carried out in class, so they can only be evaluated through out the course and can not be re-evaluated for the Final Evaluation mode or in the extra call. In these activities we will also account for the attendance to class, active participation of the student and the evaluable exercises that will be performed almost every class.

### Data Mining Project

The Data Mining Project will be evaluated according to the three phases described below and the corresponding weights.

This project will be done individually or in groups of 2 people. The work will be done incrementally and will be presented in the following phases:

- Phase 1: students will choose a domain to which data they have access, analyze their characteristics and establish the objectives to be achieved through the Data Mining Project. They will write a report indicating the different tasks that would be carried out in each stage of the Knowledge Discovery process according to the specific needs of the domain and the objectives.
- Phase 2: through the use of a Knowledge Discovery software tool, Data Mining algorithms will be applied to the data of each domain. In addition, the student will analyze the limitations of the algorithms available in the tool and possible improvements.
- Phase 3: an evaluation plan will be made to assess the results obtained and the plan will be executed.

The 3 deliveries of the Data Mining Project are mandatory and will be evaluated according to the weights assigned in the table in the previous section (summative evaluation).

The Data Mining Project will be presented in class. Each group will have 15 minutes for the oral presentation plus 5 minutes of questions.

## Qualification standards

The subject will be evaluated on 10 points, divided into 3 points for continuous assessment (this part can only be done during the course) and 7 for the Data Mining Project.

The dates for the delivery of each part of the Data Mining Project will be published at the beginning of the course.

In the Final Examination and in the extra call, those parts of the Data Mining Project that are pending may be delivered. Continuous assessment will not be repeated, but will be taken into account if done during the course.



INTERNATIONAL  
CAMPUS OF  
EXCELLENCE

PR/CL/001  
COORDINATION PROCESS OF  
LEARNING ACTIVITIES

ANX-PR/CL/001-01  
LEARNING GUIDE



E.T.S. de Ingenieros  
Informáticos

## 7. Teaching resources

---

### 7.1. Teaching resources for the subject

Name	Type	Notes
WEKA	Web resource	Official webpage of the Data Mining Tool WEKA, with tutorials and free download <a href="http://www.cs.waikato.ac.nz/ml/weka/">http://www.cs.waikato.ac.nz/ml/weka/</a>
Data Mining: Concepts and Techniques	Bibliography	Book about Data Mining Techniques. J.Han y M. Kamber. Ed. Morgan Kaufman, 2006.
Data Mining: Concepts, Models, Methods, and Algorithms	Bibliography	Book about Data Mining Techniques. M. Kantardzic (eds.), John Wiley & Sons, 2003
From Data Mining to Knowledge Discovery in Databases	Bibliography	Book about some foundational works on nowadays Data Mining Techniques U. Fayyad, G. Piatetsky-Shapiro y P. Smyth, 1996
Subject webpage	Web resource	<a href="http://www.dlsiis.fi.upm.es/master_muss/asig DCBD.html">http://www.dlsiis.fi.upm.es/master_muss/asig DCBD.html</a>

## 8. Other information

---

### 8.1. Other information about the subject

Classes will be taught in Spanish or English depending on the students. Documentation (including transparencies used in class) will be provided in English.

During the course, we will try to use as many the data files related to Sustainable Development Goals of UN as possible, specially number 13 Climate Action

Las clases se impartirán en Castellano o Ingles dependiendo del alumnado. La documentación (incluido las transparencias usadas en clase) se proporcionará en Inglés

Durante el curso, se utilizarán tantas ficheros de datos relacionados con Objetivos de Desarrollo Sostenible (ODS) de Naciones Unidas como sea posible, especialmente con el número 13 Acción por el Clima