



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

**103000907 - Time Series Data Mining**

### DEGREE PROGRAMME

10BA - Master Universitario en Ciencia de Datos

### ACADEMIC YEAR & SEMESTER

2020/21 - Semester 2

## 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.....	7
7. Teaching resources.....	10
8. Other information.....	11

## 1. Description

---

### 1.1. Subject details

<b>Name of the subject</b>	103000907 - Time Series Data Mining
<b>No of credits</b>	3 ECTS
<b>Type</b>	Optional
<b>Academic year of the programme</b>	First year
<b>Semester of tuition</b>	Semester 2
<b>Tuition period</b>	February-June
<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>	2020-21

## 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

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

CECD03 - Manejar las herramientas informáticas para Big Data

CG08 - Capacidad de pensamiento de forma creativa para desarrollar aproximaciones y métodos nuevos y originales

CG11 - Conocimiento y comprensión de la informática para crear modelos, así como sistemas y procesos de información complejos

### 3.2. Learning outcomes

RA14 - Presentar en público los resultados de sus trabajos de investigación

RA31 - Understand the basics of the Knowledge Discovery Process, and its application to time series and complex data

RA30 - Know how to select and apply the best candidate techniques to Time Series Data Mining projects

\* 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

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, focusing on time series and complex data formats, 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.

In this subject we 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. On the other hand, in recent years social networks have become immensely popular and have generated a large amount of data, structured in the form of complex data networks, which are very interesting from the point of view of the discovery of hidden knowledge in them.

### 4.2. Syllabus

1. Introduction
  - 1.1. Data Types
  - 1.2. Basic Concepts
2. Knowledge Discovery Process
  - 2.1. Knowledge Discovery Process Stages
  - 2.2. Data Preprocessing
3. KDD Tools
  - 3.1. Background
  - 3.2. A KDD Tool: WEKA
4. Data Mining Techniques for Time Series

- 4.1. Classification
- 4.2. Time Series Analysis and Prediction (with R)
- 4.3. Time Series Specific Techniques
- 4.4. Temporal Abstraction

## 5. Schedule

### 5.1. Subject schedule\*

Week	Face-to-face classroom activities	Face-to-face laboratory activities	Distant / On-line	Assessment activities
1	<p><b>1. Introduction</b> Duration: 01:30</p> <p><b>2.1 Knowledge Discovery Process</b> Duration: 01:30</p>		<p><b>1. Introduction</b> Duration: 01:30</p> <p><b>2.1 Knowledge Discovery Process</b> Duration: 01:30</p>	
2	<p><b>2.2 Data Preprocessing</b> Duration: 01:15</p> <p><b>Case Study: Knowledge Discovery Process and Data Preprocessing</b> Duration: 00:15</p> <p><b>3. KDD Tools (WEKA)</b> Duration: 01:30</p>		<p><b>2.2 Data Preprocessing</b> Duration: 01:15</p> <p><b>Case Study: Knowledge Discovery Process and Data Preprocessing</b> Duration: 00:15</p> <p><b>3. KDD Tools (WEKA)</b> Duration: 01:30</p>	
3	<p><b>Domain Analysis and KDD Process</b> Duration: 00:40</p> <p><b>3.2 Case Study: WEKA</b> Duration: 01:00</p> <p><b>4.1 Classification Techniques for Time Series</b> Duration: 01:20</p>		<p><b>Domain Analysis and KDD Process</b> Duration: 00:40</p> <p><b>3.2 Case Study: WEKA</b> Duration: 01:00</p> <p><b>4.1 Classification Techniques for Time Series</b> Duration: 01:20</p>	<p><b>Project Stage 1: Domain Analysis, Data study, Objective definition</b></p> <p>Continuous assessment Not Presential Duration: 00:20</p>
4	<p><b>Group Discussion</b> Duration: 01:00</p> <p><b>4.2 Time Series Analysis and Prediction (with R)</b> Duration: 02:00</p>		<p><b>Group Discussion</b> Duration: 01:00</p> <p><b>4.2 Time Series Analysis and Prediction (with R)</b> Duration: 02:00</p>	
5	<p><b>4.2 Time Series Multivariate Analysis and Prediction (with R)</b> Duration: 02:00</p> <p><b>Case Study: Multivariate Analysis</b> Duration: 01:00</p>		<p><b>4.2 Time Series Multivariate Analysis and Prediction (with R)</b> Duration: 02:00</p> <p><b>Case Study: Multivariate Analysis</b> Duration: 01:00</p>	

6	<b>4.3 Time Series Data Mining Specific Techniques</b> Duration: 02:15  <b>Case Study: Time Series Data Mining</b> Duration: 00:45		<b>4.3 Time Series Data Mining Specific Techniques</b> Duration: 02:15  <b>Case Study: Time Series Data Mining</b> Duration: 00:45	
7	<b>4.4 Temporal Abstraction</b> Duration: 01:30  <b>Case Study: Temporal Abstraction</b> Duration: 01:30		<b>4.4 Temporal Abstraction</b> Duration: 01:30  <b>Case Study: Temporal Abstraction</b> Duration: 01:30	<b>Project Stage 2: Application of Data Mining Techniques</b>  Continuous assessment Not Presential Duration: 00:20
8	<b>Project Presentations</b> Duration: 03:00		<b>Project Presentations</b> Duration: 03:00	
9				
10				
11				
12				
13				
14				
15				
16				
17				<b>Project complete</b>  Final examination Not Presential Duration: 01:00  <b>Project Presentation</b>  Continuous assessment and final examination Not Presential Duration: 02:00  <b>Attendance to class, participation and evaluable exercises</b>  Continuous assessment Not Presential Duration: 00:00

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. Continuous assessment

Week	Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
3	Project Stage 1: Domain Analysis, Data study, Objective definition		No Presential	00:20	10%	/ 10	CECD01 CECD03
7	Project Stage 2: Application of Data Mining Techniques		No Presential	00:20	10%	/ 10	CECD03 CG08 CG11
17	Project Presentation		No Presential	02:00	50%	/ 10	CG08 CG11
17	Attendance to class, participation and evaluable exercises		No Presential	00:00	30%	/ 10	CECD01 CECD03 CG08 CG11

#### 6.1.2. Final examination

Week	Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
17	Project complete		No Presential	01:00	50%	/ 10	CECD01 CECD03 CG08 CG11
17	Project Presentation		No Presential	02:00	50%	/ 10	CG08 CG11

#### 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	CECD01 CECD03 CG08 CG11

## 6.2. Assessment criteria

For the evaluation of this subject we will take into account, on the one hand, the attendance and participation in class and, on the other hand, the Data Mining Project.

For the Continuous Evaluation the attendance to class (online or in class), the active participation of the student and the evaluable exercises that are raised in class will be considered.

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

### Data Mining Project

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.

The 2 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 (online). Each group will have approximately 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 and 7 for the Data Mining Project. To pass the subject it will be necessary to attend at least 70% of the classes and obtain a final grade of no less than 5 points.

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

In the extra call, those parts of the Data Mining Project that are pending may be delivered. Continuous assessment will not be repeated, so the grade of the subject will be obtained exclusively from the Data Mining Project.

## 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>
A review on time series data mining	Bibliography	Engineering Applications of Artificial Intelligence 24 (2011) 164?181
Time Series Analysis and Its Applications With R Examples - 4th Edition	Bibliography	Book on Time Series Analysis by Shumway and Stoffer

## 8. Other information

---

### 8.1. Other information about the subject

Classes will be taught in English with occasional parallel explanations in Spanish. 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 serán impartidas en inglés con explicaciones paralelas ocasionales en español. La documentación (incluido las transparencias usadas en clase) se proporcionará en Inglés

Durante el curso, se utilizarán tantos 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