



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

2022/23 - 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

Name of the subject	103000907 - Time Series Data Mining
No of credits	3 ECTS
Type	Optional
Academic year of the programme	First year
Semester of tuition	Semester 2
Tuition period	February-June
Tuition languages	English
Degree programme	10BA - Master Universitario en Ciencia de Datos
Centre	10 - Escuela Tecnica Superior De Ingenieros Informaticos
Academic year	2022-23

2. Faculty

2.1. Faculty members with subject teaching role

Name and surname	Office/Room	Email	Tutoring hours *
Aurora Perez Perez	D4301	aurora.perez@upm.es	M - 10:30 - 13:30 Th - 10:30 - 13:30
Juan Pedro Caraca-Valente Hernandez (Subject coordinator)	D4301	juanpedro.caracavalente@upm.es	Tu - 09:00 - 12:00 Th - 10:00 - 13:00

* 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 the most important techniques for each stage will be listed.

In this subject we want to explore areas of Knowledge Discovery less known, but are becoming increasingly relevant. 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. We will discuss traditional numeric Time Series Techniques, novel approaches and will pay special attention to Symbolic approaches.

4.2. Syllabus

1. Introduction
 - 1.1. Time Series
 - 1.2. Basic Concepts
2. Numeric Techniques: Classic Approach
 - 2.1. Introduction to Time Series Distances
 - 2.2. Fourier Transform Approach
 - 2.3. Pattern Discovery
3. Numeric Techniques: Forecasting
 - 3.1. Introduction to Autocorrelation
 - 3.2. Basic ARIMA Models
 - 3.3. Seasonal ARIMA Models

3.4. Multivariate ARIMA Models

3.5. Other Forecasting Techniques

4. Symbolic Techniques for Time Series

4.1. Domain Independent Symbolic Techniques

4.2. Domain Dependent Temporal Abstraction

4.3. Case Studies

5. Schedule

5.1. Subject schedule*

Week	Classroom activities	Laboratory activities	Distant / On-line	Assessment activities
1	<p>1. Introduction Duration: 01:00</p> <p>2.Numeric Techniques: Introduction Duration: 01:00</p>			<p>Time Series Generation: Understanding the underlying components of a real world time series</p> <p>Continuous assessment Presential Duration: 01:00</p>
2	<p>2.1 Time Series Distances Duration: 00:45</p> <p>2.2 Fourier Transform Duration: 00:30</p> <p>2.3 Pattern Discovery Duration: 00:45</p>			<p>Time Series Clustering</p> <p>Continuous assessment Presential Duration: 01:00</p>
3	<p>Domain Analysis and KDD Process Duration: 00:25</p> <p>3.1 Introduction to Autocorrelation Duration: 00:20</p> <p>3.2 Basic ARIMA Models Duration: 01:15</p>			<p>Project Stage 1: Domain Analysis, Data study, Objective definition</p> <p>Continuous assessment Not Presential Duration: 00:20</p> <p>Applying Basic ARIMA Models with R</p> <p>Continuous assessment Presential Duration: 01:00</p>
4	<p>3.3 Seasonal ARIMA Models Duration: 01:00</p> <p>3.4 Multivariate ARIMA Models Duration: 01:00</p>			<p>Applying Seasonal and Multivariate ARIMA Models with R</p> <p>Continuous assessment Presential Duration: 01:00</p>
5	<p>Group Discussion Duration: 01:00</p>			<p>Techniques for Time Series Forecasting</p> <p>Continuous assessment Presential Duration: 02:00</p>

6	4.1 Symbolic Techniques for Time Series: Domain Independent Duration: 02:00			Applying SAX or other Symbolic Technique Continuous assessment Presential Duration: 01:00
7	4.2 Symbolic techniques: Temporal Abstraction Duration: 00:30 Case Study: Temporal Abstraction Duration: 01:30			Project Stage 2: Application of Data Mining Techniques Continuous assessment Not Presential Duration: 00:20 Applying Temporal Abstraction Continuous assessment Presential Duration: 01:00
8				Project Presentations Continuous assessment Presential Duration: 03:00
9				
10				
11				
12				
13				
14				
15				
16				
17				Project Delivery Final examination Not 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.

6. Activities and assessment criteria

6.1. Assessment activities

6.1.1. Assessment

Week	Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
1	Time Series Generation: Understanding the underlying components of a real world time series		Face-to-face	01:00	5%	/ 10	CECD01 CECD03 CG11
2	Time Series Clustering		Face-to-face	01:00	5%	/ 10	CECD01 CECD03 CG08 CG11
3	Project Stage 1: Domain Analysis, Data study, Objective definition		No Presential	00:20	10%	/ 10	CECD01 CECD03
3	Applying Basic ARIMA Models with R		Face-to-face	01:00	5%	/ 10	CECD01 CECD03 CG08 CG11
4	Applying Seasonal and Multivariate ARIMA Models with R		Face-to-face	01:00	5%	/ 10	CECD01 CECD03 CG08 CG11
5	Techniques for Time Series Forecasting		Face-to-face	02:00	10%	/ 10	CECD01 CECD03 CG08 CG11
6	Applying SAX or other Symbolic Technique		Face-to-face	01:00	5%	/ 10	CECD01 CECD03 CG08 CG11
7	Project Stage 2: Application of Data Mining Techniques		No Presential	00:20	10%	/ 10	CECD03 CG08 CG11
7	Applying Temporal Abstraction		Face-to-face	01:00	5%	/ 10	CG08 CG11 CECD01 CECD03

8	Project Presentations		Face-to-face	03:00	50%	/ 10	CECD03 CG08 CG11 CECD01
---	-----------------------	--	--------------	-------	-----	------	----------------------------------

6.1.2. Global examination

Week	Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
17	Project Delivery		No Presential	02:00	100%	5 / 10	CECD01 CECD03 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 Progressive Evaluation Activities and, on the other hand, the Data Mining Project.

For the **Progressive Evaluation** we will consider:

- the attendance to class
- the active participation of the student
- the tasks that will be done in each class applying the concepts explained previously. This will be the most important part of the Progressive Evaluation

As they are carried out during the course, progressive evaluation activities are only evaluable for the

progressive evaluation, and can not be repeated for the Final Evaluation or the extra call.

The **Data Mining Project** will be evaluated according to the phases described below and their 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 describing the domain they have selected, analysing the data and making an initial proposal of the objectives of the project.

- Phase 2: through the use of a Knowledge Discovery software tool or just a programming language, Data Mining algorithms will be applied to the data of each domain. The results should be evaluated and analyzed. In addition, the student will analyze the limitations of the algorithms used and possible improvements.

The two 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 approximately 15 minutes for the oral presentation plus 5 minutes of questions.

Qualification standards

- Progressive Evaluation: The subject will be evaluated on 10 points, divided into 4 points for Progressive Evaluation Activities and 7 for the Data Mining Project (yes, more than 10 points are available to encourage consistency in doing the daily activities). The dates for the delivery of each part of the Data Mining Project will be published at the beginning of the course.

- Final Evaluation and the extra call: Only the Data Mining Project will be considered. 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.

7. Teaching resources

7.1. Teaching resources for the subject

Name	Type	Notes
Time Series Analysis and Its Applications With R Examples - 4th Edition	Bibliography	Book on Time Series Analysis by Shumway and Stoffer
A review on time series data mining	Bibliography	Engineering Applications of Artificial Intelligence 24 (2011) 164?181
Data Mining: Concepts and Techniques	Bibliography	Book about Data Mining Techniques. J.Han y M. Kamber. Ed. Morgan Kauffman, 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
WEKA	Web resource	Official webpage of the Data Mining Tool WEKA, with tutorials and free download http://www.cs.waikato.ac.nz/ml/weka/

Subject webpage	Web resource	http://www.dlsiis.fi.upm.es/master_muss/asigDCBD.html
-----------------	--------------	---

8. Other information

8.1. Other information about the subject

Classes will be taught in English with occasional on demand 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