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

2020/21 - Semester 1

Index

Learning guide

1. Description

1.1. Subject details

Name of the subject	103000865 - Data Mining And Time Series
No of credits	3 ECTS
Type	Optional
Academic year ot the programme	Second year
Semester of tuition	Semester 3
Tuition period	September-January
Tuition languages	English
Degree programme	10AZ - Master Universitario en Innovación Digital
Centre	10 - Escuela Tecnica Superior de Ingenieros Informaticos
Academic year	2020-21

2. Faculty

2.1. Faculty members with subject teaching role

Name and surname	Office/Room	Email	Tutoring hours *
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

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

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

* 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 and Complex data

5. Schedule

5.1. Subject schedule*

Week	Face-to-face classroom activities	Face-to-face laboratory activities	Distant / On-line	Assessment activities
1	**1. Introduction** Duration: 02:00		**1. Introduction** Duration: 02:00	
2	**2.1 Knowledge Discovery Process** Duration: 01:45 **Case Study: Knowledge Discovery Process** Duration: 00:15		**2.1 Knowledge Discovery Process** Duration: 01:45 **Case Study: Knowledge Discovery Process** Duration: 00:15	
3	**2.2 Data Preprocessiong** Duration: 01:45 **Case Study: Data Preprocessing** Duration: 00:15		**2.2 Data Preprocessiong** Duration: 01:45 **Case Study: Data Preprocessing** Duration: 00:15	
4	**3 KDD Tools** Duration: 00:20 **3.1 Background and 3.2 WEKA** Duration: 01:40		**3 KDD Tools** Duration: 00:20 **3.1 Background and 3.2 WEKA** Duration: 01:40	
5	**3.2 Case Study: WEKA** Duration: 01:20 **Domain Analysis and KDD Process** Duration: 00:40		**3.2 Case Study: WEKA** Duration: 01:20 **Domain Analysis and KDD Process** Duration: 00:40	**Project Stage 1: Domain Analysys, Data study, Objective definition** Continuous assessment Not Presential Duration: 00:20
6	**4.1 Classification Techniques** Duration: 01:45 **Case Study: Classification Techniques** Duration: 00:15		**4.1 Classification Techniques** Duration: 01:45 **Case Study: Classification Techniques** Duration: 00:15	
7	**4.1 Classification Techniques** Duration: 01:45 **Case Study: Classification Techniques** Duration: 00:15		**4.1 Classification Techniques** Duration: 01:45 **Case Study: Classification Techniques** Duration: 00:15	

8	**4.2 ClusteringTechniques** Duration: 01:45 **Case Study: Clustering Techniques** Duration: 00:15		**4.2 ClusteringTechniques** Duration: 01:45 **Case Study: Clustering Techniques** Duration: 00:15	
9	**4.2 Clustering Techniques** Duration: 01:45 **Case Study: Clustering Techniques** Duration: 00:15		**4.2 Clustering Techniques** Duration: 01:45 **Case Study: Clustering Techniques** Duration: 00:15	
10	**4.3 Genetic Algorithms** Duration: 01:45 **Case Study: Genetic Algorithms** Duration: 00:15		**4.3 Genetic Algorithms** Duration: 01:45 **Case Study: Genetic Algorithms** Duration: 00:15	
11	**4.3 Genetic Algorithms** Duration: 01:45 **Case Study: Genetic Algorithms** Duration: 00:15		**Case Study: Genetic Algorithms** Duration: 00:15 **Case Study: Genetic Algorithms** Duration: 00:15	**Project Stage 2: Application of Data Mining Techniques** Continuous assessment Not Presential Duration: 00:20
12	**4.4 Time Series Data Mining** Duration: 01:45 **Case Study: Time Series Data Mining** Duration: 00:15		**4.4 Time Series Data Mining** Duration: 01:45 **Case Study: Time Series Data Mining** Duration: 00:15	
13	**4.4 Time Series Data Mining** Duration: 01:45 **Case Study: Time Series Data Mining** Duration: 00:15		**4.4 Time Series Data Mining** Duration: 01:45 **Case Study: Time Series Data Mining** Duration: 00:15	
14	**4.4 Time Series Data Mining** Duration: 01:45 **Case Study: Time Series Data Mining** Duration: 00:15		**Case Study: Time Series Data Mining** Duration: 00:15 **Case Study: Time Series Data Mining** Duration: 00:15	
15	**5 Evaluation** Duration: 01:45 **Case Study: Data Mining Project Evaluation** Duration: 00:15		**5 Evaluation** Duration: 01:45 **Case Study: Data Mining Project Evaluation** Duration: 00:15	

16				
17				**Project Stage 3: Evaluation** Continuous assessment Not Presential Duration: 00:20 **Project complete** Final examination Not Presential Duration: 01:00 **Project Presentation** Continuous assessment and final examination Not Presential Duration: 02:00 **Attendance to class, participation and evaluable exercises** 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
5	Project Stage 1: Domain Analysys, Data study, Objective definition		No Presential	00:20	10%	/ 10	CG02
11	Project Stage 2: Application of Data Mining Techniques		No Presential	00:20	10%	/ 10	CB07 CG02 CE-CD04
17	Project Stage 3: Evaluation		No Presential	00:20	10%	/ 10	CB07 CG02 CG03
17	Project Presentation		No Presential	02:00	40%	/ 10	CB07 CG02 CG03 CE-CD04
17	Attendance to class, participation and evaluable exercises		No Presential	00:00	30%	/ 10	CB07 CG02 CE-CD04

6.1.2. Final examination

Week	Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
17	Project complete		No Presential	01:00	60%	/ 10	CB07 CG02 CG03 CE-CD04
17	Project Presentation		No Presential	02:00	40%	/ 10	CB07 CG02 CG03 CE-CD04

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	CB07 CG02 CG03 CE-CD04

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, 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 three phases described below and the 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.

· 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 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 beggining 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 http://www.cs.waikato.ac.nz/ml/weka/
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 fundational works on nowadays Data Mining Techniques U. Fayyad, G. Piatetsky-Shapiro y P. Smyth, 1996
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 (Spanish only if every student speaks spanish correctly). 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