ANX-PR/CL/001-01
LEARNING GUIDE

SUBJECT
103000935 - Data Mining

DEGREE PROGRAMME
10AZ - Master Universitario En Innovación Digital

ACADEMIC YEAR & SEMESTER
2024/25 - Semester 2
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.............................................................................................................3
6. Schedule..................................................................................................................................................................5
7. Activities and assessment criteria............................................................................................................................7
8. Teaching resources..................................................................................................................................................9
9. Other information....................................................................................................................................................10
1. Description

1.1. Subject details

<table>
<thead>
<tr>
<th>Name of the subject</th>
<th>103000935 - Data Mining</th>
</tr>
</thead>
<tbody>
<tr>
<td>No of credits</td>
<td>4 ECTS</td>
</tr>
<tr>
<td>Type</td>
<td>Optional</td>
</tr>
<tr>
<td>Academic year of the programme</td>
<td>First year</td>
</tr>
<tr>
<td>Semester of tuition</td>
<td>Semester 2</td>
</tr>
<tr>
<td>Tuition period</td>
<td>February-June</td>
</tr>
<tr>
<td>Tuition languages</td>
<td>English</td>
</tr>
<tr>
<td>Degree programme</td>
<td>10AZ - Master Universitario en Innovación Digital</td>
</tr>
<tr>
<td>Centre</td>
<td>10 - Escuela Tecnica Superior De Ingenieros Informaticos</td>
</tr>
<tr>
<td>Academic year</td>
<td>2024-25</td>
</tr>
</tbody>
</table>

2. Faculty

2.1. Faculty members with subject teaching role

<table>
<thead>
<tr>
<th>Name and surname</th>
<th>Office/Room</th>
<th>Email</th>
<th>Tutoring hours *</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fco. Javier Segovia Perez</td>
<td>2305</td>
<td><a href="mailto:javier.segovia@upm.es">javier.segovia@upm.es</a></td>
<td>M - 10:00 - 11:00 Hablar con el profesor</td>
</tr>
<tr>
<td>(Subject coordinator)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ernestina Menasalvas Ruiz</td>
<td>4303</td>
<td><a href="mailto:ernestina.menasalvas@upm.es">ernestina.menasalvas@upm.es</a></td>
<td>M - 10:00 - 11:00 hablar con la profesora</td>
</tr>
</tbody>
</table>

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

- Artificial Intelligence
- Statistics

4. Skills and learning outcomes *

4.1. 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-DIPO03 - Habilidad para hacer conexiones entre los deseos y necesidades del consumidor o cliente y lo que la tecnología puede ofrecer

CG01 - Que los estudiantes sean capaces de predecir y controlar la evolución de situaciones complejas mediante el desarrollo de nuevas e innovadoras metodologías de trabajo adaptadas al ámbito científico/investigador, tecnológico o profesional concreto, en general multidisciplinar, en el que se desarrolle su actividad.

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

CG07 - Capacidad de trabajar y comunicarse también en contextos internacionales.
4.2. Learning outcomes

RA102 - Being able to translate a data insight into a business decision and action.

RA130 - Being able to create data insights using Data Mining

RA100 - Being able to understand how to effectively manage the analytical processes and use the results of these processes (models, clusters, etc.) as the basis for making informed, evidence-based decisions for creating value for a company

RA101 - Being able to reframe a business question as a data question, reasoning about what data might be of assistance and how to obtain it

* 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

The course is mainly dedicated to the improvement of the development of software engineering projects by means of Data Mining.

The course is very interactive, with the development of many short projects and exposition at class. Learning by doing, using the IBM SPSS tool

Students will learn the use of the IBM SPSS tool by themselves, with some teacher support

Topics:

Data Engineering, Data Mining, Business Intelligence

Descriptive, Diagnostic, Predictive and Prescriptive Analysis of data

Techniques:

- Classification
- Regression

- Association

- Clustering

5.2. Syllabus

1. INTRODUCTION TO DATA ANALYSIS AND THE IBM SPSS MODELER
2. BUSINESS VALUE PROPOSITION & DATA MINING
3. DESCRIPTIVE ANALYSIS USING BASIC STATISTICS
4. DESCRIPTIVE ANALYSIS: DATA VISUALIZATION
5. DESCRIPTIVE ANALYSIS: RFM
6. DESCRIPTIVE ANALYSIS: CLUSTERING
7. DIAGNOSTIC ANALYSIS: CORRELATION, ANOVA AND CHI-SQUARED TESTS
8. DIAGNOSTIC ANALYSIS: ASSOCIATION RULES
9. PREDICTIVE ANALYSIS: LINEAR REGRESSION
10. PREDICTIVE ANALYSIS: LOGISTIC REGRESSION
11. PREDICTIVE ANALYSIS: DECISION TREES
12. PREDICTIVE ANALYSIS: NEAREST NEIGHBOR
13. PREDICTIVE ANALYSIS: NEURAL NETWORKS
14. PREDICTIVE ANALYSIS: PROBABILISTIC MODELS
15. PREDICTIVE ANALYSIS: ENSEMBLE METHODS
16. PREDICTIVE ANALYSIS: DEALING WITH TIME
17. PRESCRIPTIVE ANALYSIS: TARGETING CUSTOMERS
18. PRESCRIPTIVE ANALYSIS: RECOMMENDATION SYSTEMS
19. FINAL PROJECT
6. Schedule

6.1. Subject schedule*

<table>
<thead>
<tr>
<th>Week</th>
<th>Type 1 activities</th>
<th>Type 2 activities</th>
<th>Distant / On-line</th>
<th>Assessment activities</th>
</tr>
</thead>
</table>
| 1    | Introduction to Data Mining                                                      | Exercise about Value Proposition Canvas | Tool practice | Individual presentation  
Duration: 00:30  
Lecture |  
Duration: 00:30  
Problem-solving class |  
Individual presentation  
Progressive assessment  
Presential  
Duration: 02:30 |
|      | Problem about Data Analysis on a Business Case                                   | Tool practice     | Theoretical Problem | Individual presentation  
Progressive assessment  
Presential  
Duration: 01:00 |
|      |                                                                 | Duration: 02:30  
Problem-solving class | | |
| 2    | Problem about DATA VISUALIZATION using the tool                                 | Problem about STATISTICAL DESCRIPTION using the tool | Practical Problem using the tool | Individual presentation  
Progressive assessment  
Presential  
Duration: 04:00 |
|      | Duration: 02:00  
Problem-solving class | Duration: 02:00  
Problem-solving class | | |
| 3    | Problem about RFM using the tool                                                 | Problem about CLUSTERING using the tool | Practical Problem using the tool | Individual presentation  
Progressive assessment  
Presential  
Duration: 04:00 |
|      | Duration: 02:00  
Problem-solving class | Duration: 02:00  
Problem-solving class | | |
| 4    | Problem about CORRELATION, ANOVA AND CHI-SQUARED TESTS using the tool           | Problem about ASSOCIATION RULES using the tool | Practical Problem using the tool | Individual presentation  
Progressive assessment  
Presential  
Duration: 04:00 |
|      | Duration: 02:00  
Problem-solving class | Duration: 02:00  
Problem-solving class | | |
<table>
<thead>
<tr>
<th></th>
<th>Problem about LINEAR REGRESSION using the tool</th>
<th>Duration: 02:00</th>
<th>Problem-solving class</th>
<th>Practical Problem using the tool</th>
<th>Individual presentation</th>
<th>Progressive assessment</th>
<th>Presential</th>
<th>Duration: 04:00</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Problem about LOGISTIC REGRESSION using the tool</td>
<td>Duration: 02:00</td>
<td>Problem-solving class</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Problem about NEAREST NEIGHBOR AND NEURAL NETWORKS USING THE TOOL</td>
<td>Duration: 02:00</td>
<td>Problem-solving class</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Problem about DECISION TREES using the tool</td>
<td>Duration: 02:00</td>
<td>Problem-solving class</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Problem about TIME RELATED CASES using the tool</td>
<td>Duration: 02:00</td>
<td>Problem-solving class</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Problem about ENSEMBLE METHODS using the tool</td>
<td>Duration: 02:00</td>
<td>Problem-solving class</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Problem about PRESCRIPTIVE ANALYSIS using the tool</td>
<td>Duration: 02:00</td>
<td>Problem-solving class</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Final Project Presentation</td>
<td>Duration: 00:20</td>
<td>Additional activities</td>
<td>Final Project presentation and all assignments uploaded</td>
<td>Individual presentation</td>
<td>Progressive assessment</td>
<td>Presential</td>
<td>Duration: 00:20</td>
</tr>
<tr>
<td>17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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.
# 7. Activities and assessment criteria

## 7.1. Assessment activities

### 7.1.1. Assessment

<table>
<thead>
<tr>
<th>Week</th>
<th>Description</th>
<th>Modality</th>
<th>Type</th>
<th>Duration</th>
<th>Weight</th>
<th>Minimum grade</th>
<th>Evaluated skills</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Tool practice</td>
<td>Individual presentation</td>
<td>Face-to-face</td>
<td>02:30</td>
<td>5%</td>
<td>5 / 10</td>
<td>CB07</td>
</tr>
<tr>
<td>1</td>
<td>Theoretical Problem</td>
<td>Individual presentation</td>
<td>Face-to-face</td>
<td>01:00</td>
<td>5%</td>
<td>5 / 10</td>
<td>CG03, CG07, CE-DIPO03, CB07</td>
</tr>
<tr>
<td>2</td>
<td>Practical Problem using the tool</td>
<td>Individual presentation</td>
<td>Face-to-face</td>
<td>04:00</td>
<td>10%</td>
<td>5 / 10</td>
<td>CG07, CE-DIPO03, CB07, CG03, CG01</td>
</tr>
<tr>
<td>3</td>
<td>Practical Problem using the tool</td>
<td>Individual presentation</td>
<td>Face-to-face</td>
<td>04:00</td>
<td>10%</td>
<td>5 / 10</td>
<td>CG07, CE-DIPO03, CB07, CG03, CG01</td>
</tr>
<tr>
<td>4</td>
<td>Practical Problem using the tool</td>
<td>Individual presentation</td>
<td>Face-to-face</td>
<td>04:00</td>
<td>10%</td>
<td>5 / 10</td>
<td>CG03, CG07, CE-DIPO03, CB07, CG01</td>
</tr>
<tr>
<td>5</td>
<td>Practical Problem using the tool</td>
<td>Individual presentation</td>
<td>Face-to-face</td>
<td>04:00</td>
<td>10%</td>
<td>5 / 10</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Practical Problem using the tool</td>
<td>Individual presentation</td>
<td>Face-to-face</td>
<td>04:00</td>
<td>10%</td>
<td>5 / 10</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Practical Problem using the tool</td>
<td>Individual presentation</td>
<td>Face-to-face</td>
<td>04:00</td>
<td>10%</td>
<td>5 / 10</td>
<td>CG03, CG07, CE-DIPO03, CB07, CG01</td>
</tr>
<tr>
<td>8</td>
<td>Practical Problem using the tool</td>
<td>Individual presentation</td>
<td>Face-to-face</td>
<td>02:00</td>
<td>10%</td>
<td>5 / 10</td>
<td>CG03, CG07, CE-DIPO03, CB07, CG01</td>
</tr>
</tbody>
</table>
### 7.1.2. Global examination

<table>
<thead>
<tr>
<th>Week</th>
<th>Description</th>
<th>Modality</th>
<th>Type</th>
<th>Duration</th>
<th>Weight</th>
<th>Minimum grade</th>
<th>Evaluated skills</th>
</tr>
</thead>
<tbody>
<tr>
<td>17</td>
<td>FINAL PROJECT presentation and all assignments uploaded</td>
<td>Individual</td>
<td>Face-to-face</td>
<td>00:20</td>
<td>20%</td>
<td>5 / 10</td>
<td>CG03 CG07 CE-DIPO03 CB07 CG01</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Week</th>
<th>Description</th>
<th>Modality</th>
<th>Type</th>
<th>Duration</th>
<th>Weight</th>
<th>Minimum grade</th>
<th>Evaluated skills</th>
</tr>
</thead>
<tbody>
<tr>
<td>17</td>
<td>FINAL PROJECT presentation and all assignments uploaded</td>
<td>Individual</td>
<td>Face-to-face</td>
<td>00:20</td>
<td>100%</td>
<td>5 / 10</td>
<td>CG03 CG07 CE-DIPO03 CB07 CG01</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Description</th>
<th>Modality</th>
<th>Type</th>
<th>Duration</th>
<th>Weight</th>
<th>Minimum grade</th>
<th>Evaluated skills</th>
</tr>
</thead>
<tbody>
<tr>
<td>FINAL PROJECT presentation and all assignments uploaded</td>
<td>Individual</td>
<td>Face-to-face</td>
<td>00:20</td>
<td>100%</td>
<td>5 / 10</td>
<td>CG03 CG07 CE-DIPO03 CB07 CG01</td>
</tr>
</tbody>
</table>
7.2. Assessment criteria

At each session we will solve a problem/assignment at class. The evaluation is based on each session problem/assignment and the final project. But solving, uploading and passing (mark 5 minimum) all problems/assignments is mandatory.

For the continuous or progressive evaluation, it is allowed one resubmission for each assignment.

For the global evaluation, you can only submit once each assignment.

For the extraordinary evaluation, only failed assignments submitted in the continuous or global examinations can be resubmitted.

8. Teaching resources

8.1. Teaching resources for the subject

<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jiawei Han, Micheline Kamber, Data Mining : Concepts and Techniques, 2nd edition, Morgan Kaufmann, ISBN 1558609016, 2006.</td>
<td>Bibliography</td>
<td></td>
</tr>
</tbody>
</table>
### 9. Other information

#### 9.1. Other information about the subject

We will use the Problem-based learning technique in class. Before each session you will have to read some materials and watch some videos around a topic. Then at class I will answer questions regarding those materials and videos, and I will propose a problem related to the topic that you have to solve at class. The whole session will be for you to solve the problem, with me around to help you. It is, therefore, very important that you have in mind that this course requires your presence at class, and a computer with the tool.