



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

103000875 - Data Engineering

DEGREE PROGRAMME

10AZ - Master Universitario en Innovación Digital

ACADEMIC YEAR & SEMESTER

2020/21 - 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..... | 8 |
| 8. Teaching resources..... | 10 |
| 9. Other information..... | 11 |

1. Description

1.1. Subject details

| | |
|---------------------------------------|--|
| Name of the subject | 103000875 - Data Engineering |
| 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 | 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 * |
|---|--------------------|-----------------------------|--|
| Fco.javier Segovia Perez (Subject coordinator) | 2305 | javier.segovia@upm.es | M - 10:00 - 11:00 Hablar con el profesor |
| Ernestina Menasalvas Ruiz | 4303 | ernestina.menasalvas@upm.es | M - 10:00 - 11:00 hablar con la profesora |

* 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

CB08 - Que los estudiantes sean capaces de integrar conocimientos y enfrentarse a la complejidad de formular juicios a partir de una información que, siendo incompleta o limitada, incluya reflexiones sobre las responsabilidades sociales y éticas vinculadas a la aplicación de sus conocimientos y juicios

CE-EIT03 - Capacidad para identificar el nivel de madurez de una tecnología y desarrollar e interpretar un roadmap tecnológico seleccionando la mejor manera de proteger esa tecnología dependiendo de su tipo, nivel de madurez y las restricciones geográficas, y entendiendo las consecuencias de acceder a ella y comercializarla.

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.

CG09 - La capacidad de transformar las experiencias prácticas en problemas y desafíos de investigación.

4.2. Learning outcomes

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

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

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

RA103 - Being able to understand the data science's implications for management and decision making in a datarich environment.

* 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 Modeler tool

Topics:

Data Engineering, Data Mining, Business Intelligence

CRISP-DM, or the Data Mining Process

Techniques:

- Classification

- Regression

- Association

- Clustering

5.2. Syllabus

1. INTRODUCTION TO DATA ENGINEERING
2. THE TOOL: IBM SPSS MODELER
3. THE PROCESS CRISP-DM
4. LINER REGRESSION
5. LOGISTIC REGRESSION
6. RFM ANALYSIS
7. DECISION TREES
8. NEURAL NETWORKS
9. CLUSTERING
10. NEAREST NEIGHBOR
11. ASSOCIATION RULES

6. Schedule

6.1. Subject schedule*

| Week | Face-to-face classroom activities | Face-to-face laboratory activities | Distant / On-line | Assessment activities |
|------|--|------------------------------------|--|---|
| 1 | INTRODUCTION TO DATA ENGINEERING Duration: 02:00 | | INTRODUCTION TO DATA ENGINEERING Duration: 02:00 | |
| 2 | THE PROCESS OF DATA MINING Duration: 01:00 | | Tema 2 Duration: 01:00 THE PROCESS OF DATA MINING Duration: 01:00 | |
| 3 | Data Understanding and Data Preparation with IBM SPSS Modeler I Duration: 01:00 | | Data Understanding and Data Preparation with IBM SPSS Modeler I Duration: 02:00 | |
| 4 | Data Understanding and Data Preparation with IBM SPSS Modeler II Duration: 01:00 | | Data Understanding and Data Preparation with IBM SPSS Modeler II Duration: 02:00 | |
| 5 | LINEAR REGRESSION Duration: 01:00 | | LINEAR REGRESSION Duration: 02:00 LINEAR REGRESSION Duration: 01:00 | |
| 6 | Modeling with IBM SPSS Modeler I Duration: 01:00 | | Modeling with IBM SPSS Modeler I Duration: 02:00 | ASSIGNMENT 1 Continuous assessment and final examination Not Presential Duration: 02:00 |
| 7 | LOGISTIC REGRESSION Duration: 01:00 | | LOGISTIC REGRESSION Duration: 02:00 LOGISTIC REGRESSION Duration: 01:00 | |
| 8 | Modeling with IBM SPSS Modeler II Duration: 01:00 | | Modeling with IBM SPSS Modeler II Duration: 02:00 | ASSIGNMENT 2 Continuous assessment and final examination Not Presential Duration: 02:00 |

| | | | | |
|----|---|--|--|--|
| 9 | DECISION TREES Duration: 01:00 | | DECISION TREES Duration: 02:00 DECISION TREES Duration: 01:00 | |
| 10 | RFM ANALYSIS Duration: 01:00 | | RFM ANALYSIS Duration: 02:00 RFM ANALYSIS Duration: 01:00 | |
| 11 | NEURAL NETWORKS Duration: 01:00 | | NEURAL NETWORKS Duration: 02:00 NEURAL NETWORKS Duration: 01:00 | ASSIGNMENT 3 Continuous assessment and final examination Not Presential Duration: 02:00 |
| 12 | CLUSTERING Duration: 01:00 | | CLUSTERING Duration: 02:00 CLUSTERING Duration: 01:00 | |
| 13 | NEAREST NEIGHBOR Duration: 01:00 | | NEAREST NEIGHBOR Duration: 02:00 NEAREST NEIGHBOR Duration: 01:00 | |
| 14 | ASSOCIATION RULES Duration: 01:00 | | ASSOCIATION RULES Duration: 02:00 ASSOCIATION RULES Duration: 01:00 | ASSIGNMENT 4 Continuous assessment and final examination Not Presential Duration: 02:00 |
| 15 | | | EXERCISES Duration: 01:00 | ASSIGNMENT 5 Continuous assessment and final examination Not Presential Duration: 02:00 |
| 16 | | | | |
| 17 | | | | FINAL PROJECT Continuous assessment and 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.

7. Activities and assessment criteria

7.1. Assessment activities

7.1.1. Continuous assessment

| Week | Description | Modality | Type | Duration | Weight | Minimum grade | Evaluated skills |
|------|---------------|----------|---------------|----------|--------|---------------|--|
| 6 | ASSIGNMENT 1 | | No Presential | 02:00 | 10% | 5 / 10 | CB08 CE-EIT03 CG03 CG07 CB07 CG09 |
| 8 | ASSIGNMENT 2 | | No Presential | 02:00 | 10% | 5 / 10 | CG03 CG07 CB08 CE-EIT03 CB07 |
| 11 | ASSIGNMENT 3 | | No Presential | 02:00 | 10% | 5 / 10 | CG07 CB08 CE-EIT03 CB07 CG09 CG03 |
| 14 | ASSIGNMENT 4 | | No Presential | 02:00 | 10% | 5 / 10 | CG03 CG07 CB08 CE-EIT03 CB07 CG09 |
| 15 | ASSIGNMENT 5 | | No Presential | 02:00 | 10% | 5 / 10 | CG03 CG07 CB08 CE-EIT03 CB07 CG09 |
| 17 | FINAL PROJECT | | No Presential | 02:00 | 50% | 5 / 10 | CB08 CE-EIT03 CB07 CG09 CG03 CG07 |

7.1.2. Final examination

| Week | Description | Modality | Type | Duration | Weight | Minimum grade | Evaluated skills |
|------|---------------|----------|---------------|----------|--------|---------------|--|
| 6 | ASSIGNMENT 1 | | No Presential | 02:00 | 10% | 5 / 10 | CB08 CE-EIT03 CG03 CG07 CB07 CG09 |
| 8 | ASSIGNMENT 2 | | No Presential | 02:00 | 10% | 5 / 10 | CG03 CG07 CB08 CE-EIT03 CB07 |
| 11 | ASSIGNMENT 3 | | No Presential | 02:00 | 10% | 5 / 10 | CG07 CB08 CE-EIT03 CB07 CG09 CG03 |
| 14 | ASSIGNMENT 4 | | No Presential | 02:00 | 10% | 5 / 10 | CG03 CG07 CB08 CE-EIT03 CB07 CG09 |
| 15 | ASSIGNMENT 5 | | No Presential | 02:00 | 10% | 5 / 10 | CG03 CG07 CB08 CE-EIT03 CB07 CG09 |
| 17 | FINAL PROJECT | | No Presential | 02:00 | 50% | 5 / 10 | CB08 CE-EIT03 CB07 CG09 CG03 CG07 |

7.1.3. Referred (re-sit) examination

| Description | Modality | Type | Duration | Weight | Minimum grade | Evaluated skills |
|-------------|----------|------|----------|--------|---------------|------------------|
|-------------|----------|------|----------|--------|---------------|------------------|

| | | | | | | |
|-------------------------------|--|--------------|-------|------|--------|--|
| FINAL PROJECT AND ASSIGNMENTS | | Face-to-face | 02:02 | 100% | 5 / 10 | |
|-------------------------------|--|--------------|-------|------|--------|--|

7.2. Assessment criteria

The evaluation is based on the assignments and the final project.

Assignments and projects will be performed individually or by groups, depending on the size of the course

To pass the course it is mandatory to present all the assignments and the final project, in any modality of evaluation

Participation in class would give a 10% increase in the final score.

8. Teaching resources

8.1. Teaching resources for the subject

| Name | Type | Notes |
|--|--------------|-----------------------|
| Principles of Data Mining (Adaptive Computation and Machine Learning), D Hand, MIT Press, 2001. | Bibliography | |
| Jiawei Han, Micheline Kamber, Data Mining : Concepts and Techniques, 2nd edition, Morgan Kaufmann, ISBN 1558609016, 2006. | Bibliography | |
| Data Mining Techniques: Marketing, Sales and Customer Support, Michael J. A. Berry, Gordon Linoff, John Wiley & Sons, 1997. | Bibliography | |
| Pang-Ning Tan, Michael Steinbach, Vipin Kumar, Introduction to Data Mining, Pearson Addison Wesley (May, 2005). Hardcover: 769 pages. ISBN: 0321321367 | Bibliography | MOST RECOMMENDED BOOK |

| | | |
|--|--------------|----------------------|
| Ian Witten, Eibe Frank, Mark Hall, Data Mining: Practical Machine Learning Tools and Techniques, 3rd Edition, Morgan Kaufmann, ISBN 978-0-12-374856-0, 2011. | Bibliography | |
| Página web de la asignatura en moodle | Web resource | |
| IBM SPSS MODELER | Others | THE TOOL WE WILL USE |
| Sala de trabajo en grupo con ordenadores | Equipment | |
| aula | Equipment | |

9. Other information

9.1. Other information about the subject

We will use Teams for communication and collaboration