ANX-PR/CL/001-01
LEARNING GUIDE

SUBJECT
93000944 - Large-scale Media Analytics

DEGREE PROGRAMME
09AT - Master Universitario En Teoria De La Señal Y Comunicaciones

ACADEMIC YEAR & SEMESTER
2023/24 - 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..................................................................................................................................................................6
6. Activities and assessment criteria............................................................................................................................8
7. Teaching resources.................................................................................................................................................10
8. Other information....................................................................................................................................................11
1. Description

1.1. Subject details

<table>
<thead>
<tr>
<th>Name of the subject</th>
<th>93000944 - Large-Scale Media Analytics</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>09AT - Master Universitario en Teoria de la Señal y Comunicaciones</td>
</tr>
<tr>
<td>Centre</td>
<td>09 - Escuela Tecnica Superior De Ingenieros De Telecomunicacion</td>
</tr>
<tr>
<td>Academic year</td>
<td>2023-24</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>Alberto Belmonte Hernandez</td>
<td>D-112.2</td>
<td><a href="mailto:alberto.belmonte@upm.es">alberto.belmonte@upm.es</a></td>
<td>Sin horario. Appointment arranged by email</td>
</tr>
<tr>
<td>(Subject coordinator)</td>
<td></td>
<td></td>
<td>Sin horario. Appointment arranged by email</td>
</tr>
<tr>
<td>Federico Alvarez Garcia</td>
<td>D-103</td>
<td><a href="mailto:federico.alvarez@upm.es">federico.alvarez@upm.es</a></td>
<td>Sin horario. Appointment arranged by email</td>
</tr>
</tbody>
</table>
Jose Manuel Menendez
Garcia

C-300

jm.menendez@upm.es

Sin horario.
Appointment
arranged by email

* 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

CB06 - Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en el desarrollo y/o aplicación de ideas, a menudo en un contexto de investigación

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

CB09 - Que los estudiantes sepan comunicar sus conclusiones y los conocimientos y razones últimas que las sustentan a públicos especializados y no especializados de un modo claro y sin ambigüedades

CB10 - Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo

CE01 - Analizar y aplicar técnicas para el diseño y desarrollo avanzado de equipos y sistemas, basándose en la teoría de la señal y las comunicaciones, en un entorno internacional
CE02 - Evaluar y sintetizar los resultados de un trabajo en equipo en proyectos relacionados con la teoría de la señal y las comunicaciones, en un entorno internacional.

CT05 - Capacidad para gestionar la información, identificando las fuentes necesarias, los principales tipos de documentos técnicos y científicos, de una manera adecuada y eficiente

3.2. Learning outcomes

RA34 - Capability to develop and evaluate machine-learning techniques and to design big data learning systems

RA43 - Ability to develop basic applications in relevant current use cases in the media industry (media search, content recommendation, etc.)

RA42 - Knowledge on Big Data technologies and their application to multimedia content

RA41 - Ability to select and apply adequate machine learning techniques to large-scale multimedia datasets and evaluate their performance

* 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

Current data analysis applications require the management of extremely large collections of heterogeneous multimedia data. The extraction of knowledge from these huge datasets is a difficult problem with a broad scope.

This subject aims at presenting the most relevant techniques and methodologies for large scale multimedia analysis.

In particular, we will discuss the application of widely used feature extraction and machine learning techniques (dimensionality reduction, classification, clustering) to text, image, audio, spatio-temporal and tabular data. Heterogeneous information networks and suitable data mining techniques will also be described.

Big data technologies will be introduced, including efficient acquisition, storage and processing of huge amounts of heterogeneous data. Some of the described techniques will be applied to relevant use cases, such as content search, summarization, content recommendation...
Practical sessions will be proposed in which students will apply these tools to real datasets and become familiar with powerful analysis frameworks. Finally a final project will be developed by students applying the techniques reviewed in the lessons over a selected dataset.

4.2. Syllabus

1. Exploratory Data Analysis
   1.1. Analytic graphics and techniques for exploratory data analysis
   1.2. Data transformation and dimensionality reduction, clustering and outliers detection
   1.3. Algorithms analysis, hyperparameters, evaluation metrics
   1.4. Explainable AI
   1.5. Lab session: EDA and analysis applied to a dataset

2. Image and Audio Feature Extraction
   2.1. Computer vision and image descriptors: Histograms, Texture, Spatio-Temporal
   2.2. Image/Video Movement
   2.3. Deep Learning for image feature extraction
   2.4. Audio signal and audio descriptors
   2.5. Deep Learning for audio analysis
   2.6. Lab session: Image/audio search and classification

3. Recommender Systems
   3.1. Simple recommenders
   3.2. Content-based recommenders
   3.3. Collaborative filtering
   3.4. Hybrid recommenders
   3.5. Lab session: Simple, content-based and collaborative filtering recommenders

4. Text analysis
   4.1. Text mining and cleaning
   4.2. Vector space models
   4.3. Keywords extraction and summarization
   4.4. Text clustering, classification and recommendation
4.5. Lab session: Text analytics

5. Machine/Deep Learning Applications with multimedia content
   5.1. Advanced techniques for multimedia content
   5.2. Recent deep learning architectures for multimedia content

6. Project development in pairs or individually
   6.1. Project guidance
   6.2. Project results and presentation
5. Schedule

5.1. Subject schedule*

<table>
<thead>
<tr>
<th>Week</th>
<th>Classroom activities</th>
<th>Laboratory activities</th>
<th>Distant / On-line</th>
<th>Assessment activities</th>
</tr>
</thead>
</table>
| 1    | Chapter 1 - EDA and data understanding  
Duration: 03:00  
Lecture |                      |                   |                       |
| 2    |                      | Lab 1 - EDA and data understanding  
Duration: 03:00  
Laboratory assignments |                   |                       |
| 3    | Chapter 2 - Image feature extraction  
Duration: 03:00  
Lecture |                      |                   |                       |
| 4    | Chapter 2 - Image/Video and audio feature extraction  
Duration: 03:00  
Lecture |                      |                   |                       |
| 5    |                      | Lab 2 - Image/Video/Audio feature extraction  
Duration: 03:00  
Laboratory assignments |                   |                       |
| 6    | Chapter 3 - Recommender systems  
Duration: 03:00  
Lecture |                      |                   |                       |
| 7    |                      | Lab 3 - Recommender systems  
Duration: 03:00  
Laboratory assignments |                   |                       |
| 8    | Chapter 4 - Text feature extraction  
Duration: 03:00  
Laboratory assignments |                      |                   |                       |
| 9    |                      | Lab 4 - Text feature extraction  
Duration: 03:00  
Laboratory assignments |                   |                       |
| 10   | Chapter 5 - Advanced techniques  
Duration: 03:00  
Lecture |                      |                   |                       |
| 11   |                      | Project session: Project proposal and dataset used  
Duration: 03:00  
Laboratory assignments |                   | Lab sessions report  
Problem-solving test  
Continuous assessment  
Not Present  
Duration: 00:00 |
| 12   |                      | Project session: Feature extraction techniques  
Duration: 03:00  
Laboratory assignments |                   |                       |
<table>
<thead>
<tr>
<th></th>
<th>Project session: Preliminary results</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>Duration: 03:00</td>
<td>Laboratory assignments</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Project session: Final projects ideas</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Duration: 00:00</td>
<td>Laboratory assignments</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td>Project development and presentation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Group work</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Continuous assessment</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Presential</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Duration: 00:30</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td>Exam: Test / Short Questions</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Written test</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Continuous assessment</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Presential</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Duration: 01:30</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td></td>
<td>Final exam</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Written test</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Final examination</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Presential</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Duration: 01:30</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lab sessions report</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Problem-solving test</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Final examination</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Not Presential</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Duration: 00:00</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.

* 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

<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>11</td>
<td>Lab sessions report</td>
<td>Problem-solving</td>
<td>No Presental</td>
<td>00:00</td>
<td>25%</td>
<td>4 / 10</td>
<td>CB07 CT05</td>
</tr>
<tr>
<td>15</td>
<td>Project development and presentation</td>
<td>Group work</td>
<td>Face-to-face</td>
<td>00:30</td>
<td>35%</td>
<td>5 / 10</td>
<td>CB06 CB07 CB09 CB10 CT05 CE01 CE02</td>
</tr>
<tr>
<td>17</td>
<td>Exam: Test / Short Questions</td>
<td>Written test</td>
<td>Face-to-face</td>
<td>01:30</td>
<td>40%</td>
<td>4 / 10</td>
<td>CB07 CB09 CB10 CT05 CE01 CE02</td>
</tr>
</tbody>
</table>

6.1.2. 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 exam</td>
<td>Written test</td>
<td>Face-to-face</td>
<td>01:30</td>
<td>75%</td>
<td>5 / 10</td>
<td>CB06 CB07 CB09 CB10 CT05 CE01 CE02</td>
</tr>
<tr>
<td>17</td>
<td>Lab sessions report</td>
<td>Problem-solving test</td>
<td>No Presental</td>
<td>00:00</td>
<td>25%</td>
<td>5 / 10</td>
<td>CB07 CT05</td>
</tr>
</tbody>
</table>

6.1.3. 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>
</table>
6.2. Assessment criteria

Students will be qualified through progressive evaluation. Evaluation will assess if students have acquired all the competences of the subject. Thus, evaluation through final assessment will be carried out considering all the evaluation techniques used in continuous evaluation (EX, ET, TG, etc.), and will be celebrated in the exam period approved by Junta de Escuela for the current academic semester and year. Evaluation activities that assess learning outcomes that cannot be evaluated through a single exam can be carried out along the semester.

The progressive evaluation will be based on the following elements:

- Attend and follow the theory and practical sessions. 4 laboratory sessions will be held. Each lab will incorporate some questions that need to be solved during the session. After the laboratory sessions a homework over a different dataset need to be done (mandatory practices).
- Develop a final project in pairs or individually and present the results to the classroom. This activity includes the reading and extracting of the main ideas from relevant papers in the field. Project will include a report and the performance of a presentation to the group (15 minutes plus 5 minutes discussion) by each pair/individual of students. During the project, a presentation of the actual results will be carry out to the rest of the class (project follow up) every week.
- Individual exam: questions on the theoretical content of the course, selected papers and code.

All parts are required to pass the subject.

The weight of such activities, all mandatory are:

- Lab sessions (individual) 25% - minimal threshold 4/10
- Project (pairs/individual) 35% - minimal threshold 5/10
- Exam [Test / Short questions] (individual) 40% - minimal threshold of 4/10
In the event that a student does not pass Lab sessions or/and Project, the student will be evaluated
a final exam with the requirement of delivering the laboratory practices tasks (homework). In this case the
percentages for the evaluations are the following:

- Lab sessions (individual) 25% - minimal threshold 5/10
- Exam [Test / Short questions] (individual) 75% - minimal threshold of 5/10

Extraordinary examination will be carried out by a final exam and the requirement of completing the laboratory
practices tasks (homework). The minimum percentages to pass the subject in this case are:

- Lab sessions (individual) 25% - minimal threshold 5/10
- Exam [Test / Short questions] (individual) 75% - minimal threshold of 5/10

7. Teaching resources

7.1. Teaching resources for the subject

<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Python Data Science Handbook - Jake VanderPlas</td>
<td>Bibliography</td>
<td>Data analysis</td>
</tr>
<tr>
<td>Python for Data Analysis, Data Wrangling with Pandas, NumPy, and IPython - Wes Mckinney</td>
<td>Bibliography</td>
<td>Data analysis</td>
</tr>
</tbody>
</table>
8. Other information

8.1. Other information about the subject

In this subject we align with the Sustainable Development Goals (SDG) 4, 5 and 9.

Especially this subject will support in the activities to be carried out:

4.4 By 2030, substantially increase the number of youth and adults who have relevant skills, including technical and vocational skills, for employment, decent jobs and entrepreneurship.

5. B Enhance the use of enabling technology, in particular information and communications technology, to promote the empowerment of women.

9.5 Enhance scientific research, upgrade the technological capabilities of industrial sectors in all countries, in particular developing countries, including, by 2030, encouraging innovation and substantially increasing the number of research and development workers per 1 million people and public and private research and development spending.