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

2018/19 - 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....11

1. Description

1.1. Subject details

Name of the subject	93000944 - Large-scale media analytics
No of credits	4 ECTS
Type	Optional
Academic year ot the programme	First year
Semester of tuition	Semester 2
Tuition period	February-June
Tuition languages	English
Degree programme	09AT - Master universitario en teoria de la señal y comunicaciones
Centre	09 - Escuela Tecnica Superior de Ingenieros de Telecomunicacion
Academic year	2018-19

2. Faculty

2.1. Faculty members with subject teaching role

Name and surname	Office/Room	Email	Tutoring hours *
Federico Alvarez Garcia (Subject coordinator)	D-103	federico.alvarez@upm.es	Sin horario. Appointment arranged by email
Jose Manuel Menendez Garcia	C-300	jm.menendez@upm.es	Sin horario. Appointment arranged by email

Guillermo Cisneros Perez	C-300	guillermo.cisneros@upm.es	Sin horario.

* 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

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

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

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.)

* 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 machine learning techniques (dimensionality reduction, classification, clustering) to textual, image and spatio-temporal 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 web content search and recommendation.

Practical sessions will be proposed in which students will apply these tools to real datasets and become familiar with powerful analysis frameworks.

4.2. Syllabus

1. Multimedia analytics
   1.1. Introduction to multimedia analytics: Multimedia content analysis and applications
   1.2. Content descriptors extraction: video, image, audio and text
   1.3. Content Handling, Search and Retrieval at big scale
   1.4. Use of multimedia analytics in the industry
   1.5. Lab session: descriptors and multimedia search
2. Machine learning tools for multimedia analysis
   2.1. Classification and regression for multimedia
   2.2. Clustering applied to multimedia applications
3. Use cases: a practical approach including lab sessions
   3.1. Use case 1: Search and ranking
      3.1.1. Mining homogeneous graphs
      3.1.2. Web Search
      3.1.3. Search and ranking (PageRank) with Python
   3.2. Use case 2: Text processing
      3.2.1. Text analysis and clustering
      3.2.2. Graph analysis (keywords, summarization)
      3.2.3. Lab session: Text processing with Python
   3.3. Use case 3: Content recommendation
      3.3.1. Collaborative filtering (CF)
      3.3.2. Content-based filtering (link to video/image descriptors)
      3.3.3. Lab session: CF recommendation with Spark
4. Project development in pairs (application of ML tools to a real dataset)
   4.1. Lab session: Exploratory data analysis

4.2. Lab session: Inference on the data

4.3. Lab session: Visualization of results

5. Evaluation

5.1. Test and project presentations

5.2. Selected papers presentations

5. Schedule

5.1. Subject schedule*

Week	Face-to-face classroom activities	Face-to-face laboratory activities	Other face-to-face activities	Assessment activities
1	**1.1 Introduction to multimedia analytics: Multimedia content analysis and applications** Duration: 02:00 Lecture **1.2 Content descriptors extraction: video, image, audio and text. Part 1** Duration: 01:00 Lecture			
2	**1.2 Content descriptors extraction: video, image, audio and text. part 2** Duration: 01:00 Lecture **1.3. Content Handling, Search and Retrieval at big scale** Duration: 02:00 Lecture			
3	**1.4. Use of multimedia analytics in the industry** Duration: 01:00 Lecture	**1.5. Lab session: descriptors and multimedia search** Duration: 02:00 Laboratory assignments		
4	**2.1. Classification and regression for multimedia** Duration: 01:30 Lecture **2.2. Clustering applied to multimedia applications** Duration: 01:30 Lecture			
5	**Use Case 1: Search and Ranking. Mining homogeneous graphs and web search description** Duration: 01:00 Lecture	**Search and ranking with Python** Duration: 02:00 Laboratory assignments		
6	**Use case 2: Text processing. Text analysis and clustering. Graph analysis** Duration: 01:00 Lecture	**Lab session: Text processing with Python** Duration: 02:00 Laboratory assignments		
7	**Use case 3: Content recommendation. Collaborative filtering (CF). Content-based filtering (link to video/image descriptors)** Duration: 02:00 Lecture	**Lab session: CF recommendation with Spark. part 1** Duration: 01:00 Laboratory assignments		

8	**Lab session: CF recommendation with Spark. part 1** Duration: 02:00 Laboratory assignments			
9	**Project development in pairs (application of ML tools to a real dataset)** Duration: 01:00 Lecture	**Lab session: Exploratory data analysis** Duration: 02:00 Laboratory assignments		
10	**Lab session: Inference on the data** Duration: 02:00 Laboratory assignments			
11	**Lab session: Visualization of results** Duration: 02:00 Laboratory assignments			
12	**Lab session: Integration of results and validation** Duration: 03:00 Laboratory assignments			
13				**Lab sessions (pairs) report** Problem-solving test Continuous assessment Duration: 00:20 **Project presentation** Group work Continuous assessment Duration: 02:30
14				**Exam: Test / Short Questions** Written test Continuous assessment Duration: 01:00
15				
16				
17				**Global exam** Written test Final examination Duration: 01:30

The independent study hours are training activities during which students should spend time on individual study or individual assignments.

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 subject schedule is based on a previous theorical planning of the subject plan and might go to through experience some unexpected changes along throughout the academic year.

6. Activities and assessment criteria

6.1. Assessment activities

6.1.1. Continuous assessment

Week	Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
13	Lab sessions (pairs) report	Problem-solving test	Face-to-face	00:20	20%	3 / 10	CB07 CT05
13	Project presentation	Group work	Face-to-face	02:30	35%	3 / 10	CB09 CB07 CB06 CE02 CE01 CT05 CB10
14	Exam: Test / Short Questions	Written test	Face-to-face	01:00	45%	4 / 10	CB09 CB07 CE01 CT05 CB10

6.1.2. Final examination

Week	Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
17	Global exam	Written test	Face-to-face	01:30	100%	5 / 10	CB09 CB07 CB06 CE02 CE01 CT05 CB10

6.1.3. Referred (re-sit) examination

Description	Modality	Type	Duration	Weight	Minimum grade	Evaluated skills
Extraordinary assessment	Written test	Face-to-face	01:30	100%	5 / 10	CB09 CB07 CB06 CE02 CE01 CT05 CB10

6.2. Assessment criteria

Students will be qualified through continuous evaluation by default. According to the Normativa de Evaluación del Aprendizaje de la Universidad Politécnica de Madrid, students willing to renounce to continuous evaluation must complete the Moodle task entitled "Renounce to continuous evaluation" before the end of the 4th week after the subject start (deadline will be announced in Moodle).

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.

Extraordinary examination will be carried out exclusively by the final assessment method.

The continuous evaluation will be based on the following elements;

- Attend and follow the theory and practical sessions and hand in a report for each lab session
- Develop a final project in pairs 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 of students.
- 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 (pairs) 20% - minimal threshold 3/10
- Project (pairs) 35% - minimal threshold 3/10
- Exam [Test / Short questions] (individually) 45% - with a minimal threshold of 4/10 in the total mark of the exam

Considering the nature of the subject all students are encouraged to follow the evaluation procedure described above.

In case of students not taking the option of "continuous assessment" they should carry out only a final exam (weight 100% of the final mark); nevertheless, the student should bring the results from the lab sessions and project to the final (global) exam.

Extraordinary examination will be carried out exclusively by the final assessment method, with the same method and conditions as indicated for students not following the "continous assessment" (the exam will include questions on the lab sessions reports, questions on the project results and test/short questions on the theoretical aspects of the subject).

7. Teaching resources

7.1. Teaching resources for the subject

Name	Type	Notes
Pentreath, N. (2015). Machine Learning with Spark. Packt Publishing Ltd.	Bibliography	Machine Learning with Spark
McKinney, W. (2012). Python for data analysis: Data wrangling with Pandas, NumPy, and IPython. " O'Reilly Media, Inc.".	Bibliography	McKinney, W. (2012). Python for data analysis: Data wrangling with Pandas, NumPy, and IPython
Segaran, T. (2007). Programming collective intelligence: building smart web 2.0 applications. " O'Reilly Media, Inc.".	Bibliography	Programming collective intelligence: building smart web 2.0 applications
Aggarwal, C. C., & Zhai, C. (2012). Mining text data. Springer Science & Business Media.	Bibliography	Mining text data. Springer Science & Business Media.