



IDENTIFYING DATA

Data engineering

Subject	Data engineering			
Code	O06M193V01102			
Study programme	Máster universitario en Inteligencia artificial			
Descriptors	ECTS Credits	Choose	Year	Quadmester
	3	Mandatory	1st	1st
Teaching language	English			
Department				
Coordinator	García Lourenco, Analía María			
Lecturers	García Lourenco, Analía María			
E-mail	analía@uvigo.es			
Web	http://http://moovi.uvigo.gal			
General description	The aim of this course is to introduce the basics of data engineering, notably in the scope of Big Data. The acquired skills will allow the analysis and the efficient management of heterogeneous information, both structured and non structured, within the development of AI applications, whenever traditional methods show insufficiency.			

Training and Learning Results

Code	
A1	CB6 - Possess and understand knowledge that provides a basis or opportunity to be original in the development and/or application of ideas, often in a research context
A2	CB7 - Students should be able to apply their acquired knowledge and problem-solving skills in new or unfamiliar environments within broader (or multidisciplinary) contexts related to their area of study.
A3	CB8 - the complexity of making judgments based on information that, while incomplete or limited, includes reflections on the social and ethical responsibilities linked to the application of their knowledge and judgments.
B2	Successfully address all stages of an Artificial Intelligence project.
B3	Search and select useful information needed to solve complex problems, handling with fluency the bibliographic sources of the field.
B4	Elaborate adequately and with certain originality written compositions or motivated arguments, write plans, work projects, scientific articles and formulate reasonable hypotheses in the field.
B5	Work in teams, especially multidisciplinary teams, and be skilled in time management, people management and decision making.
C16	Knowledge of the process and tools for data processing and preparation from data acquisition or extraction, cleaning, transformation, loading, organization and access.
D3	Utilizar las herramientas básicas de las tecnologías de la información y las comunicaciones (TIC) necesarias para el ejercicio de su profesión y para el aprendizaje a lo largo de su vida.
D7	Develop the ability to work in interdisciplinary or transdisciplinary teams to offer proposals that contribute to sustainable environmental, economic, political and social development.
D8	Value the importance of research, innovation and technological development in the socioeconomic and cultural progress of society.
D9	Have the ability to manage time and resources: develop plans, prioritize activities, identify critical ones, set deadlines and meet them.

Expected results from this subject

Expected results from this subject	Training and Learning Results
RA1: Develop the capacity to analyse and model data for processing in intelligent systems.	A1 A2 C16 D3 D9

RA2: Know and understand the process of extraction, cleaning, transformation, load and preprocessing of data.	A3 B2 B3 C16 D3 D7 D9
RA3: Know and learn how to use multidimensional and NoSQL databases.	A2 B3 B4 D8
RA4: Know the foundations of data lakes and data warehouses.	A2 A3 B2 B5 D3 D7 D8

Contents

Topic	
Concepts and foundations of Data Engineering	Concepts and basic definitions, problems of efficient data load in Big Data scenarios, massive data storage and access.
Techniques of data cleaning and preparation	Common techniques. Definition of processing flows. Quality metrics.
Efficient advanced structures and data warehouses for Big Data	Data warehouses and multidimensional databases, data lakes, NoSQL databases.

Planning

	Class hours	Hours outside the classroom	Total hours
Lecturing	12	0	12
Laboratory practical	8	0	8
Project	0	50	50
Problem and/or exercise solving	5	0	5

*The information in the planning table is for guidance only and does not take into account the heterogeneity of the students.

Methodologies

	Description
Lecturing	The teacher will introduce given subjects to the students with the aim to acquire information valuable within a specific scope. CONTINUOUS EVALUATION Mandatory character Facultative attendance GLOBAL EVALUATION Mandatory character
Laboratory practical	Problem or problems of practical character whose resolution requires the understanding and application of the theoretical and practical contents covered by the course. The students can work the solution to the proposed problems individually or in groups. CONTINUOUS EVALUATION Mandatory character Mandatory attendance (min. 75% of lab practices) GLOBAL EVALUATION Mandatory character

Personalized assistance

Tests	Description
Project	Doubts related to the planning and development of the project will be addressed.
Problem and/or exercise solving	Doubts related to the cases to be analyzed will be answered.

Assessment

Description		Qualification	Training and Learning Results			
Laboratory practical	Several tests aimed to evaluate the understanding of the knowledge exposed in theory and/or practical classes. Learning results evaluated: RA3, RA4	30	A2	B2 B5	C16	D3 D7
Project	Development of a project to evaluate the ability of the students to work autonomously, including their critical spirit and the ability to apply the acquired knowledge to real-world problems. At the end, the students must submit a detailed project report. Learning results evaluated: RA2, RA4	40	A2	B2 B4	C16	D8 D9
Problem and/or exercise solving	Defense of the solution proposed by the student before the teacher and the rest of the students. Learning results evaluated: RA1, RA2	30	A1 A2 A3	B2 B3	C16	D7 D8

Other comments on the Evaluation

CONTINUOUS EVALUATION SYSTEM

TEST 1: Laboratory practical

Description: Several tests aimed to evaluate the understanding of the knowledge exposed in theory and/or practical classes.

Methodology to be applied: Laboratory practical.

% Qualification: 30%.

Minimum %: To pass this part of the course the student has to obtain a grade equal or greater than 5 points (out of 10).

Training: A2,B2,B5,C16,D3,D7.

Learning results: RA3, RA4

TEST 2: Problem and/or exercise solving

Description: Defense of the solution proposed by the student before the teacher and the rest of the students.

Methodology to be applied: Problem and/or exercise solving.

% Qualification: 30%.

Minimum %: To pass this part of the course the student has to obtain a grade equal or greater than 5 points (out of 10).

Training: A1,A2,A3,B2,B3,C16,D7,D8

Learning results: RA1, RA2

TEST 3: Project

Description: Development of a project to evaluate the ability of the students to work autonomously, including their critical spirit and the ability to apply the acquired knowledge to real-world problems. At the end, the students must submit a detailed project report.

Methodology to be applied: Project.

% Qualification: 40%.

Minimum %: To pass this part of the course the student has to obtain a grade equal or greater than 5 points (out of 10).

Training: A2,B2,B4,C16,D8,D9

Learning results: RA2, RA4

GLOBAL EVALUATION SYSTEM

Procedure for choosing the global evaluation modality: students are considered to have chosen the global evaluation

system if they do not take Test 1 of the continuous evaluation system.

TEST 1: Laboratory practical

Description: Several tests aimed to evaluate the understanding of the knowledge exposed in theory and/or practical classes.

Methodology to be applied: Laboratory practical.

% Qualification: 30%.

Minimum %: To pass this part of the course the student has to obtain a grade equal or greater than 5 points (out of 10).

Training: A2,B2,B5,C16,D3,D7.

Learning results: RA3, RA4

TEST 2: Problem and/or exercise solving

Description: Defense of the solution proposed by the student before the teacher and the rest of the students.

Methodology to be applied: Problem and/or exercise solving.

% Qualification: 30%.

Minimum %: To pass this part of the course the student has to obtain a grade equal or greater than 5 points (out of 10).

Training: A1,A2,A3,B2,B3,C16,D7,D8

Learning results: RA1, RA2

TEST 3: Project

Description: Development of a project to evaluate the ability of the students to work autonomously, including their critical spirit and the ability to apply the acquired knowledge to real-world problems. At the end, the students must submit a detailed project report.

Methodology to be applied: Project.

% Qualification: 40%.

Minimum %: To pass this part of the course the student has to obtain a grade equal or greater than 5 points (out of 10).

Training: A2,B2,B4,C16,D8,D9

Learning results: RA2,RA4

CRITERIA OF EVALUACION FOR EXTRAORDINARY AND END OF CAREER CALLS

The continuous and global evaluation systems described above will be used.

MINUTES QUALIFICATION PROCESS

Regardless of the evaluation system and the call, in case of failing any part of the evaluation, but the overall score is higher than 4 (out of 10), the grade in the minutes will be 4).

EVALUATION DATES

The dates of the tests corresponding to the continuous evaluation system will be published in the calendar of activities, available on the website of the ESEI <https://esei.uvigo.es/docencia/horarios/>. The official exam dates for the different calls, officially approved by the ESEI Xunta de Centro, are published on the ESEI website <https://esei.uvigo.es/docencia/exames/>.

USE OF MOBILE DEVICES

All students are reminded of the prohibition of the use of mobile devices in exercises and practices, in compliance with Article 13.2.d) of the University Student Statute, regarding the duties of university students, which establishes the duty to "Refrain from the use or cooperation in fraudulent procedures in evaluation tests, in the work performed or in official university documents."

CONSULTATION/REQUEST FOR TUTORIALS Tutorials can be consulted through the faculty member's personal page, accessible through <https://esei.uvigo.es/docencia/profesorado/>.

OTHER CONSIDERATIONS

If plagiarism is detected in any of the works (essays or project), the final grade will be "Suspenso" (0) and the situation will be notified to the School's Board to take the appropriate disciplinary actions. If translation errors cause any contradictions between the various versions of this syllabus, the English will be the prevailing version.

Sources of information

Basic Bibliography

Sadalage, Fowler, **NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence**, 978-0321826626, 1, Addison-Wesley, 2012

Avi Silberschatz, Henry F. Korth, S. Sudarshan, **Database System Concepts**, 0-07-352332-1, 6, McGraw-Hill, 2010

Ihab F. Ilyas and Xu Chu, **Data Cleaning. Association for Computing Machinery**, <https://doi.org/10.1145/3310205>, Association for Computing Machinery, 2019

Alex Gorelik, **The Enterprise Big Data Lake: Delivering the Promise of Big Data and Data Science**, 9781491931554, O'Reilly Media, Inc., 2019

Matt Casters, Roland Bouman, Jos van Dongen,, **Pentaho Kettle Solutions: Building Open Source ETL Solutions with Pentaho Data Integration**, 978-0470635179, Wiley, 2013

Complementary Bibliography

Recommendations

Other comments

Follow the proposed methodology, class assistance and participation, devoting the necessary time to the study and the development of the proposed project and specific problems/case studies with the help of the teacher.

The virtual campus will be used to improve the communication between the students and the teachers, to store the necessary materials and to support in the processes of evaluation.
