



## IDENTIFYING DATA

### AI in big data environments

Subject	AI in big data environments		
Code	O06M193V01303		
Study programme	Máster universitario en Inteligencia artificial		
Descriptors	ECTS Credits	Choose	Year
	6	Optional	2nd
Teaching language	English		
Department			
Coordinator	Ribadas Pena, Francisco José		
Lecturers	A0075-Ax2tc-1 A0075-Ax2tc-1, A0075-Ax2tc-1 A0075-Ax2tc-2 A0075-Ax2tc-2, A0075-Ax2tc-2 Ribadas Pena, Francisco José		
E-mail	ribadas@uvigo.es		
Web	<a href="http://guiadocente.udc.es/guia_docent/index.php?centre=614&amp;ensenyament=614544&amp;consulta=assignatures&amp;ny_academic=2023_24">http://guiadocente.udc.es/guia_docent/index.php?centre=614&amp;ensenyament=614544&amp;consulta=assignatures&amp;ny_academic=2023_24</a>		
General description	The every time greater quantity of accessible information through Internet does that the efficient processing of big quantities of data was every time of greater interest. This has carried to the development of new technicians of storage and processing of *ingentes quantities of information, technicians that adapt of natural form to the systems distributed.		
	The main aim of this matter is to provide to the students the knowledges and necessary skills to comprise, develop and apply technicians of artificial intelligence (*IA) in surroundings of *Big Dates.		

## 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.
C10	Ability to build, validate and apply a stochastic model of a real system from observed data and the critical analysis of the results obtained
C11	Understanding and mastery of the main data analysis techniques and tools, both from a statistical and machine learning point of view, including those dedicated to the processing of large volumes of data, and the ability to select the most appropriate ones for problem solving.
C12	Ability to plan, formulate and resolve all stages of a data project, including understanding and mastery of basic fundamentals and techniques for searching and filtering information in large data collections.
C15	Knowledge of computer tools in the field of machine learning, and ability to select the most appropriate for solving a problem.
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
New	A2 B2 C10 C11 C12 C15 D3
New	A1 A2 A3 B3 B4 B5 C10 C11 C12 C15 D3 D7 D8 D9
New	A1 A2 A3 B2 B3 B4 B5 C10 C11 C12 C15 D3 D7 D8 D9
New	A1 A2 A3 B2 C12 C15 D3 D7 D9
New	A1 A2 B3 B5 C11 C15 D3 D7 D9

New	A1 A3 B2 B3 B5 C11 C12 C15 D3 D7 D8 D9
New	A2 A3 B2 B3 C10 C11 C15 D3 D9

## Contents

Topic	
(*)Introducción ao Big Data	(*)Que é Big Data Aplicacións Big Data Analítica Big Data Problemática da análise de datos en contornas Big Data
(*)Preparación e visualización de datos	(*)Técnicas de preprocesado de datos Técnicas de visualización
(*)Infraestructuras para o almacenamento e procesamento de Big Data: Apache Hadoop e Apache Spark	(*)Procesamento distribuído e infraestructuras Aprendizaxe por lotes en plataformas paralelas e distribuídas Aprendizaxe distribuída en vertical e horizontal
(*)Tratamento de datos en continuo	(*)Aprendizaxe incremental Aprendizaxe en tempo real Problemas de cambio de concepto

## Planning

	Class hours	Hours outside the classroom	Total hours
Laboratory practical	14	44	58
Mentored work	9	20	29
Lecturing	20	21	41
Objective questions exam	2	20	22

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

## Methodologies

	Description
Laboratory practical	(*)Clases prácticas en el aula de informática, que permiten al alumno familiarizarse desde un punto de vista práctico con las cuestiones expuestas en las clases teóricas.
Mentored work	(*)Aprendizaje basado en problemas, seminarios, estudio de casos o proyectos, que permiten que el alumnado adquiera determinadas competencias en base a la resolución de ejercicios, estudio de casos y realización de proyectos.
Lecturing	(*)Clases de teoría, en las que se expone el contenido de cada tema. El alumno dispondrá de copias de las transparencias con anterioridad y el profesor promoverá una actitud activa, realizando preguntas que permitan aclarar aspectos concretos y dejando cuestiones abiertas para la reflexión del alumno.

## Personalized assistance

Methodologies	Description
Laboratory practical	
Mentored work	

## Assessment

Description	Qualification	Training and Learning Results
Laboratory practical	50	A1 B2 C10 D3 A2 B3 C11 D7 A3 B4 C12 D8 B5 C15 D9
Objective questions exam	50	A1 B2 C10 D8 A2 C11 D9 A3 C12 C15

### Other comments on the Evaluation

#### Sources of information

##### Basic Bibliography

##### Apuntes y material proporcionado por el profesorado.

Tom White, **Hadoop: The Definitive Guide, 4th Edition**, 9781491901632, 4, O'Reilly Media, Inc., 2015

Bill Chambers, Matei Zaharia, **Spark: The Definitive Guide**, 9781491912218, 1, O'Reilly Media, Inc., 2018

##### Complementary Bibliography

Rezaul Karim, Sridhar Alla, **Scala and Spark for Big Data Analytics**, 978-1785280849, 1, Packt Publishing, 2017

Nick Pentreath, **Machine Learning with Spark**, 978-1783288519, 1, Packt Publishing, 2015

Michael Bowles, **Machine Learning with Spark and Python: Essential Techniques for Predictive Analytics**, 978-1-119-56193-4, 2, Wiley, 2019

#### Recommendations

##### Subjects that it is recommended to have taken before

Machine learning I/O06M193V01105

Machine learning II/O06M193V01207

Deep learning/O06M193V01206

Data engineering/O06M193V01102