Teaching Guide				
Identifying Data				2023/24
Subject (*)	AI in Big Data Environments		Code	614544016
Study programme	Máster Universitario en Intelixencia Artificial			
Descriptors				
Cycle	Period	Year	Type	Credits
Official Master's Degree	1st four-month period	Second	Optional	6
Language	English			
Teaching method	Face-to-face			
Prerequisites				
Department	Ciencias da Computación e Tecnoloxías da Información			
Coordinador	Bolón Canedo, Verónica	E-mail	veronica.bolon@udc.es	
Lecturers	Alonso Ríos, David Bolón Canedo, Verónica	E-mail	david.alonso@udc.es veronica.bolon@udc.es	
Web				
General description	A cada vez maior cantidade de información accesible a través de Internet fai que o procesamiento eficiente de grandes cantidades de datos sexa cada vez de maior interese. Isto levou ao desenvolvemento de novas técnicas de almacenamento e procesamiento de inxentes cantidades de información, técnicas que se adaptan de forma natural aos sistemas distribuídos. O obxectivo principal desta materia é proporcionar aos estudantes os coñecementos e habilidades necesarios para comprender, desenvolver e aplicar técnicas de intelixencia artificial (IA) en contornas de Big Data.			

Study programme competences / results	
Code	Study programme competences / results
A11	CE10 - Ability to implement, validate and apply a stochastic model starting from the observed data on a real system, and to perform a critical analysis of the obtained results, selecting those ones most suitable for problem solving
A12	CE11 - Understanding and command of the main techniques and tools for data analysis, both from the statistical and the machine learning viewpoints, including those devised for large volumes of data, and ability to select those ones most suitable for problem solving
A13	CE12 - Ability to outline, formulate and solve all the stages of a data project, including the understanding and command of basic concepts and techniques for information search and filtering in big collections of data
A16	CE15 - Knowledge of computer tools in the field of machine learning and ability to select those ones most suitable for problem solving
B2	CG02 - Successfully addressing each and every stage of an AI project
B3	CG03 - Searching and selecting that useful information required to solve complex problems, with a confident handling of bibliographical sources in the field
B4	CG04 - Suitably elaborating written essays or motivated arguments, including some point of originality, writing plans, work projects, scientific papers and formulating reasonable hypotheses in the field
B5	CG05 - Working in teams, especially of multidisciplinary nature, and being skilled in the management of time, people and decision making
B6	CB01 - Acquiring and understanding knowledge that provides a basis or opportunity to be original in the development and/or application of ideas, frequently in a research context
B7	CB02 - The students will be able to apply the acquired knowledge and to use their capacity of solving problems in new or poorly explored environments inside wider (or multidisciplinary) contexts related to their field of study
B8	CB03 - The students will be able to integrate different pieces of knowledge, to face the complexity of formulating opinions (from information that may be incomplete or limited) and to include considerations about social and ethical responsibilities linked to the application of their knowledge and opinions
B9	CB04 - The students will be able to communicate their conclusions, their premises and their ultimate justifications, both to specialised and non-specialised audiences, using a clear style language, free from ambiguities
C3	CT03 - Use of the basic tools of Information and Communications Technology (ICT) required for the student's professional practice and learning along her life

C4	CT04 - Acquiring a personal development for practicing a citizenship under observation of the democratic culture, the human rights and the gender perspective
C7	CT07 - Developing the ability to work in interdisciplinary or cross-disciplinary teams to provide proposal that contribute to a sustainable environmental, economic, political and social development
C8	CT08 - Appreciating the importance of research, innovation and technological development in the socioeconomic and cultural progress of society
C9	CT09 - Being able to manage time and resources: outlining plans, prioritising activities, identifying criticisms, fixing deadlines and sticking to them

Learning outcomes			
Learning outcomes	Study programme competences / results		
Know the techniques that allow the design of scalable AI techniques at software and hardware resources level.	AC10 AC11 AC12 AC15	BC2 BC7	CC3 CC4
Acquire the skills to integrate large volume and variety of data in AI Big Data projects.	AC10 AC11 AC12 AC15	BC3 BC4 BC5 BC6 BC7 BC8 BC9	CC3 CC4 CC7 CC8 CC9
To know the scalability paradigms in machine learning algorithms.	AC10 AC11 AC12 AC15	BC2 BC3 BC4 BC5 BC6 BC7 BC8 BC9	CC3 CC4 CC7 CC8 CC9
Understand, analyze and design the necessary infrastructures for Big Data AI projects: local/cloud environment and physical/virtual equipment with low latency storage systems and distributed file systems	AC12 AC15	BC2 BC6 BC7 BC8	CC3 CC4 CC7 CC9
To know the languages, frameworks and components that allow us to increase performance in hardware infrastructures with CPU and GPU.	AC11 AC15	BC3 BC7 BC8	CC3 CC4 CC7 CC9
To know the techniques that allow, with low latency, the visualization of data in environments with large volume of information.	AC11 AC12 AC15	BC2 BC3 BC5 BC6 BC7 BC8 BC9	CC3 CC4 CC7 CC8 CC9

Use and be able to apply the correct KPIs in each environment.	AC10 AC11 AC15	BC2 BC3 BC7 BC8	CC3 CC9

Contents	
Topic	**Sub-topic**
Introduction to Big Data	What is Big Data Big Data applications Big Data analytics Data analysis problems in big data environments
Data preparation and visualization	Data preprocessing techniques Visualization techniques
Federated learning	Edge learning Privacy preservation
Infrastructures for Big Data storage and processing: Apache Hadoop and Apache Spark	Distributed processing and infrastructures Batch learning in parallel and distributed platforms Vertical and horizontal distributed learning
Streaming learning	Incremental learning Real-time learning Concept-drift problems

Planning				
Methodologies / tests	**Competencies / Results**	**Teaching hours (in-person & virtual)**	**Student?s personal work hours**	**Total hours**
ICT practicals	A11 A12 A13 A16 B2 B3 B4 B5 B6 B7 B8 B9 C3 C7 C8 C9	14	44	58
Supervised projects	A11 A12 B3 B4 B5 B6 B9 C4 C7 C8	7	20	27
Objective test	A11 A12 A13 B2 B6 B7 B8 B9 C4 C8 C9	2	20	22
Guest lecture / keynote speech	A11 A12 A13 A16 B2 B3 B4 B6 B8 B9 C4 C8 C9	21	20	41
Personalized attention		2	0	2
(*)The information in the planning table is for guidance only and does not take into account the heterogeneity of the students.				

Methodologies	
Methodologies	**Description**
ICT practicals	Practical classes in the computer classroom, which allow the student to familiarize himself/herself from a practical point of view with the issues exposed in the theoretical classes.
Supervised projects	Learning based on problems, seminars, case studies or projects, which allow students to acquire certain competences based on the resolution of exercises competencies based on the resolution of exercises, case studies and projects.
Objective test	Test in which the student must demonstrate the acquired knowledge from the course
Guest lecture / keynote speech	Theory classes, in which the content of each topic is exposed. The student will have copies of the transparencies beforehand and the professor will promote an active attitude, asking questions to clarify specific aspects and leaving open questions for the student's reflection

Personalized attention	
Methodologies	**Description**
ICT practicals Supervised projects	Realization of the practical work with the advice of the teacher. Writing documents summarizing the results in the form of reports or articles, as well as the presentation of the results with the teacher or in public sessions within the class.

Assessment			
Methodologies	**Competencies / Results**	**Description**	**Qualification**
ICT practicals	A11 A12 A13 A16 B2 B3 B4 B5 B6 B7 B8 B9 C3 C7 C8 C9	Assessment of practical work: 50% marks the solutions proposed by the students to the exposed practices will be evaluated. The internship evaluation can take place through a correction by the teacher, a defense of the solution provided by the student before the teacher or an oral presentation of the developed solution. All work must be delivered before the dates to be specified and must meet minimum quality requirements to be considered. The degree of compliance with the specifications, the methodology and rigor and the presentation of results will be assessed	50
Objective test	A11 A12 A13 B2 B6 B7 B8 B9 C4 C8 C9	Questions about the contents of the subject (which can be of the test type or problems to solve), based on the different advanced machine learning techniques and their applications.	50

Assessment comments
To pass the subject, a total score of 5 or higher must be achieved. It is essential to pass all the practices indicated as mandatory. Condition for qualification of Not Presented: do not present any practice and do not attend the final exam. Students who are not newly enrolled do not retain grades from previous courses. Recovery opportunity (July) and extraordinary: The assessment will be the same as in the ordinary opportunity. Students who have not submitted the proposed assignments throughout the semester must submit them before the established date. Condition for qualification of Not Presented: do not present any practice and do not attend the final exam. The submitted work must be original by the student. In accordance with article 14, section 4, of the regulations, the delivery of non-original works or with duplicate parts (either by copies between colleagues or by obtaining from other sources...) will carry a global grade of SUSPENSION IN THE ANNUAL CALL, both for the /a student who presents copied material as if to whom it was provided.

Sources of information	
Basic	- Apuntes proporcionados por el profesor- T. White, Hadoop: The Definitive Guide, 4th Edition, O'Reilly, 2015- B. Chambers, M. Zaharia, Spark: The Definitive Guide, O'Reilly, 2018- Apuntes proporcionados por el profesor- T. White, Hadoop: The Definitive Guide, 4th Edition, O'Reilly, 2015- B. Chambers, M. Zaharia, Spark: The Definitive Guide, O'Reilly, 2018
Complementary	- Karim, Md. Rezaul, Sridhar Alla. Scala and Spark for Big Data Analytics: Tame Big Data with Scala and Apache Spark! 1st edition. Birmingham: Packt, 2017.- Pentreath, Nick. Machine Learning with Spark Create Scalable Machine Learning Applications to Power a Modern Data-Driven Business Using Spark Packt Publishing Ltd., 2015.- Bowles, Michael. Machine Learning with Spark and Python: Essential Techniques for Predictive Analytics 2nd ed. Wiley, 2019- Karim, Md. Rezaul, Sridhar Alla. Scala and Spark for Big Data Analytics: Tame Big Data with Scala and Apache Spark! 1st edition. Birmingham: Packt, 2017.- Pentreath, Nick. Machine Learning with Spark Create Scalable Machine Learning Applications to Power a Modern Data-Driven Business Using Spark Packt Publishing Ltd., 2015.- Bowles, Michael. Machine Learning with Spark and Python: Essential Techniques for Predictive Analytics 2nd ed. Wiley, 2019

Recommendations

Subjects that it is recommended to have taken before
AI Fundamentals/614544001 Machine Learning I /614544012 Machine Learning II /614544014
Subjects that are recommended to be taken simultaneously
Subjects that continue the syllabus
Other comments

(*)The teaching guide is the document in which the URV publishes the information about all its courses. It is a public document and cannot be modified. Only in exceptional cases can it be revised by the competent agent or duly revised so that it is in line with current legislation.