UNIVERSIDADE DA CORUÑA

Guia docente					
Datos Identificativos					2023/24
Asignatura (*)	Ingeniería de Datos		Código		614544002
Titulación	Máster Universitario en Intelixencia Artificial				
Descriptores					
Ciclo	Periodo	Curso	Tipo		Créditos
Máster Oficial	1º cuatrimestre	Primero	Obligatoria		3
Idioma	Inglés				
Modalidad docente	Presencial				
Prerrequisitos					
Departamento	Ciencias da Computación e Tecnoloxías da Información				
Coordinador/a	Bernardo Roca, Guillermo de		Correo electrónico	guillermo.debernardo@udc.es	
Profesorado	Bernardo Roca, Guillermo de		Correo electrónico	guillermo.debernardo@udc.es	
Web					
Descripción general	El objetivo de la materia es la introducción de los aspectos básicos de la ingeniería de datos, fundamentalmente en el ámbito del Big Data. Las competencias adquiridas permitirán el análisis y la gestión eficiente de información heterogénea, tanto estructurada como no estructurada, dentro del desarrollo de aplicaciones IA, allí donde los métodos tradicionales muestren su insuficiencia.				

Competencias / Resultados del título	
Código	Competencias / Resultados del título
A17	CE16 - Conocimiento del proceso y las herramientas para el procesamiento y preparación de datos desde su adquisición o extracción, limpieza, transformación, carga, organización y acceso
B2	CG02 - Abordar con éxito todas las etapas de un proyecto de Inteligencia Artificial
B3	CG03 - Buscar y seleccionar la información útil necesaria para resolver problemas complejos, manejando con soltura las fuentes bibliográficas del campo
B4	CG04 - Elaborar adecuadamente y con cierta originalidad composiciones escritas o argumentos motivados, redactar planes, proyectos de trabajo, artículos científicos y formular hipótesis razonables en el campo
B5	CG05 - Trabajar en equipo, especialmente de carácter multidisciplinar, y ser hábiles en la gestión del tiempo, personas y toma de decisiones
B6	CB01 - 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
B7	CB02 - 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
B8	CB03 - Que los estudiantes sean capaces de integrar conocimientos y enfrentarse a la complejidad de formular juicios a partir de una información que, siendo incompleta o limitada, incluya reflexiones sobre las responsabilidades sociales y éticas vinculadas a la aplicación de sus conocimientos y juicios
C3	CT03 - 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
C7	CT07 - Desarrollar la capacidad de trabajar en equipos interdisciplinares o transdisciplinares, para ofrecer propuestas que contribuyan a un desarrollo sostenible ambiental, económico, político y social
C8	CT08 - Valorar la importancia que tiene la investigación, la innovación y el desarrollo tecnológico en el avance socioeconómico y cultural de la sociedad
C9	CT09 - Tener la capacidad de gestionar tiempos y recursos: desarrollar planes, priorizar actividades, identificar las críticas, establecer plazos y cumplirlos

Resultados de aprendizaje	
Resultados de aprendizaje	Competencias / Resultados del título

Desarrollar la capacidad de analizar y modelar datos para su procesado en sistemas inteligentes	AM16	BM6 BM7	CM3 CM9
Conocer y comprender el proceso de extracción, limpieza, transformación, carga y preprocesado de datos.	AM16	BM2 BM3 BM8	CM3 CM9
Conocer y saber utilizar bases de datos multidimensionales y de tipo NoSQL		BM3 BM4 BM7	CM8
Conocer los fundamentos de data lakes y data warehouses		BM2 BM5 BM7 BM8	CM3 CM7 CM8

Contenidos	
Tema	**Subtema**
Conceptos y fundamentos	Conceptos y definiciones básicas, problemas de carga eficiente en escenarios Big Data, almacenamiento de datos masivos y acceso a los mismos.
Técnicas de limpieza y preparación	Técnicas más comunes. Definición de flujos de procesamiento. Métricas de calidad.
EStruturas avanzadas	Data warehouses y BD multidimensionales, Data lakes, Bases de Datos NoSQL.

Planificación				
Metodologías / pruebas	**Competencias / Resultados**	**Horas lectivas (presenciales y virtuales)**	**Horas trabajo autónomo**	**Horas totales**
Sesión magistral	B4 B5 C3 C9	12	0	12
Prueba práctica	A17 B2 B5 B7 C3	8	0	8
Solución de problemas	A17 B2 B4 B7 C7 C9	0	50	50
Trabajos tutelados	A17 B2 B3 B6 B7 B8 C7 C8	5	0	5
Atención personalizada		0		0
(*)Los datos que aparecen en la tabla de planificación són de carácter orientativo, considerando la heterogeneidad de los alumnos				

Metodologías	
Metodologías	**Descripción**
Sesión magistral	El profesorado presenta un tema al alumnado con el objetivo de facilitar un conjunto de información con alcance concreto.
Prueba práctica	El profesorado de la materia plantea al alumnado un problema o problemas de carácter práctico cuya resolución requiere la comprensión y aplicación de los contenidos teórico-prácticos incluidos en los contenidos de la materia. El alumnado puede trabajar la solución a los problemas planteados de forma individual o en grupos.
Solución de problemas	El profesorado plantea al alumnado un trabajo cuyo alcance y objetivos requieren que sea trabajado por los alumnos de forma autónoma, aunque con la tutela del profesorado de la asignatura. En general, se aplica a trabajos con un alcance temporal y de esfuerzo superior al de las prácticas de laboratorio.

Trabajos tutelados	Se plantea al alumnado proyectos prácticos cuyo alcance requiere que se le dedique un parte importante de la dedicación total del alumno a la asignatura. Además, por el alcance de los trabajos a realizar, se requiere no sólo que el alumnado aplique competencias de gestión además de competencias de índole técnica.

Atención personalizada	
Metodologías	**Descripción**
Trabajos tutelados Solución de problemas	Trabajos tutelados (aprendizaje basado en proyectos): Se plantea al alumnado un escenario de trabajo, real o ficticio, que presenta una determinada problemática. El alumnado debe aplicar los conocimientos teórico-prácticos de la asignatura para buscar una solución a la cuestión o cuestiones planteadas. Como norma general, el estudio de casos se realizará en grupos. Los distintos grupos de trabajo expondrán y pondrán en común sus soluciones. Resolución de problemas de forma autónoma: El profesorado atenderá al alumnado en sesiones de tutorías individualizadas dedicadas al desarrollo del proyecto.

Evaluación			
Metodologías	**Competencias / Resultados**	**Descripción**	**Calificación**
Trabajos tutelados	A17 B2 B3 B6 B7 B8 C7 C8	Se llevará a cabo la defensa de la solución aportada por parte del alumno ante el profesor o una presentación oral de la solución desarrollada	30
Prueba práctica	A17 B2 B5 B7 C3	Se realizarán distintas pruebas de evaluación, orientadas especialmente a evaluar la comprensión de los conocimientos expuestos en las clases de teoría y/o prácticas. Estas pruebas no pueden repetirse en la segunda convocatoria.	30
Solución de problemas	A17 B2 B4 B7 C7 C9	La evaluación del trabajo autónomo tutelado se llevará a cabo mediante la entrega de un informe y una defensa en la que el alumnado explica su propuesta y conclusiones ante el profesorado, o mediante una presentación oral de la solución ante el aula	40

Observaciones evaluación
PRIMERA CONVOCATORIA y SEGUNDA CONVOCATORIA [Asistentes y No Asistentes]Calificación final = 0,30 * Prácticas de laboratorio + 0,40 * Resolución de problemas de forma autónoma + 0,30 * trabajos tutelados (aprendizaje basado en proyectos).Los estudiantes con dispensa académica realizarán las mismas entregas y pruebas que los alumnos asistentes.PROCESO DE CALIFICACIÓN DE ACTASPara superar la asignatura en cualquier convocatoria, la calificación final debe serigual o superior a 5, debiéndose obtener como mínimo un 5 (sobre 10) en cada una de las partes.En la segunda oportunidad no pueden recuperarse las pruebas prácticas de laboratorio, por lo que no se aplica la nota mínima en esta parte.FECHAS DE EVALUACIÓNLas fechas de realización de las pruebas periódicas y entregas de trabajos se publicarán en la primera semana de clases en la plataforma de teledocencia.OTRAS CONSIDERACIONESEn el caso de detección de plagio en alguno de los trabajos (teóricos o práctico), la calificación final de la materia será deSuspenso (0) y el hecho será comunicado a la Dirección del Centro para que tome las medidas oportunas. Ante cualquier contradicción que se pudiera dar entre las distintas versiones de la guía debido a algún error de traducción, la versión que prevalecerá es la versión en inglés.

Fuentes de información	
Básica	- Sadalage, Fowler (2012). NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence. Addison-Wesley - Avi Silberschatz, Henry F. Korth, S. Sudarshan (2010). Database System Concepts. McGraw-Hill - Ihab F. Ilyas, Xu Chu, (2019). Data Cleaning. Association for Computing Machinery. ACM - Alex Gorelik (). The Enterprise Big Data Lake: Delivering the Promise of Big Data and Data Science. O'Reilly
Complementária	- Matt Casters, Roland Bouman, Jos van Dongen (2013). Pentaho Kettle Solutions: Building Open Source ETL Solutions with Pentaho Data Integration. Wiley

Recomendaciones
Asignaturas que se recomienda haber cursado previamente
Asignaturas que se recomienda cursar simultáneamente
Asignaturas que continúan el temario
Otros comentarios

(*) La Guía Docente es el documento donde se visualiza la propuesta académica de la UDC. Este documento es público y no se puede modificar, salvo cosas excepcionales bajo la revisión del órgano competente de acuerdo a la normativa vigente que establece el proceso de elaboración de guías