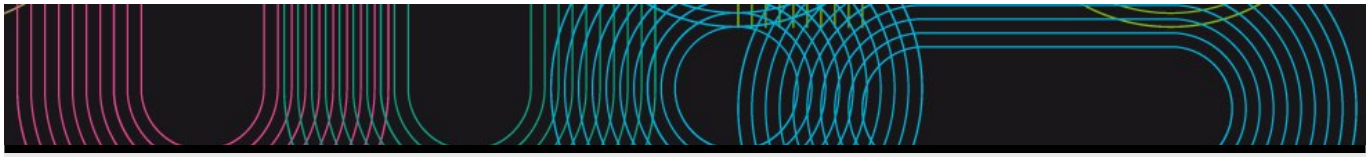

DATOS IDENTIFICATIVOS				
IA en entornos big data				
Asignatura	IA en entornos big data			
Código	O06M193V01303			
Titulacion	Máster universitario en Inteligencia artificial			
Descriptores	Creditos ECTS	Seleccione	Curso	Cuatrimestre
	6	OP	2	1c
Lengua Impartición	Inglés			
Departamento				
Coordinador/a	Ribadas Pena, Francisco José			
Profesorado	A0075-Ax2tc-1 A0075-Ax2tc-1, A0075-Ax2tc-1 A0075-Ax2tc-2 A0075-Ax2tc-2, A0075-Ax2tc-2 Ribadas Pena, Francisco José			
Correo-e	ribadas@uvigo.es			
Web	http://guiadocente.udc.es/guia_docent/index.php?centre=614&ensenyament=614544&consulta=assignatures&any_academic=2023_24			
Descripción general	La cada vez mayor cantidad de información accesible a través de Internet hace que el procesamiento eficiente de grandes cantidades de datos sea cada vez de mayor interés. Esto ha llevado al desarrollo de nuevas técnicas de almacenamiento y procesamiento de ingentes cantidades de información, técnicas que se adaptan de forma natural a los sistemas distribuidos. El objetivo principal de esta materia es proporcionar a los estudiantes los conocimientos y habilidades necesarios para comprender, desarrollar y aplicar técnicas de inteligencia artificial (IA) en entornos de Big Data.			

Resultados de Formación y Aprendizaje	
Código	
A1	CB6 - 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
A2	CB7 - 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
A3	CB8 - 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
B2	Abordar con éxito todas las etapas de un proyecto de Inteligencia Artificial
B3	Buscar y seleccionar la información útil necesaria para resolver problemas complejos, manejando con soltura las fuentes bibliográficas del campo.
B4	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	Trabajar en equipo, especialmente de carácter multidisciplinar, y ser hábiles en la gestión del tiempo, personas y toma de decisiones.
C10	Capacidad para la construcción, validación y aplicación de un modelo estocástico de un sistema real a partir de los datos observados y el análisis crítico de los resultados obtenidos.
C11	Comprensión y dominio de las principales técnicas y herramientas de análisis de datos, tanto desde el punto de vista estadístico como del aprendizaje automático, incluyendo las dedicadas al tratamiento de grandes volúmenes de datos, y capacidad para seleccionar las más adecuadas para la resolución de problemas.
C12	Capacidad para plantear, formular y resolver todas las etapas de un proyecto de datos, incluyendo la compresión y dominio de fundamentos y técnicas básicas para la búsqueda y el filtrado de información en grandes colecciones de datos.
C15	Conocimiento de las herramientas informáticas en el campo del aprendizaje automático, y capacidad para seleccionar la más adecuada para la resolución de un problema

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	Desarrollar la capacidad de trabajar en equipos interdisciplinares o transdisciplinares, para ofrecer propuestas que contribuyan a un desarrollo sostenible ambiental, económico, político y social.
D8	Valorar la importancia que tiene la investigación, la innovación y el desarrollo tecnológico en el avance socioeconómico y cultural de la sociedad
D9	Tener la capacidad de gestionar tiempos y recursos: desarrollar planes, priorizar actividades, identificar las críticas, establecer plazos y cumplirlos.

Resultados previstos en la materia

Resultados previstos en la materia	Resultados de Formación y Aprendizaje
Conocer las técnicas que permiten el diseño de técnicas de IA escalables a nivel software y de recursos hardware	A2 B2 C10 C11 C12 C15 D3
Adquirir las competencias que permitan integran gran volumen y variedad de datos en proyectos de Big Data en IA	A1 A2 A3 B3 B4 B5 C10 C11 C12 C15 D3 D7 D8 D9
Conocer los paradigmas de escalabilidad en algoritmos de aprendizaje automático	A1 A2 A3 B2 B3 B4 B5 C10 C11 C12 C15 D3 D7 D8 D9
Comprender, analizar y diseñar las infraestructuras necesarias para proyectos de IA en BigData: entorno local/nube y equipamiento físico/virtual con sistemas de almacenamiento de baja latencia y sistemas de ficheros distribuidos.	A1 A2 A3 B2 C12 C15 D3 D7 D9
Conocer los lenguajes, frameworks y componentes que nos permiten incrementar el rendimiento en las infraestructuras hardware con CPU y GPU.	A1 A2 B3 B5 C11 C15 D3 D7 D9

Conocer las técnicas que permiten, con baja latencia, la visualización de datos en entornos con gran volumen de información.	A1 A3 B2 B3 B5 C11 C12 C15 D3 D7 D8 D9
Usar y poder aplicar los KPI correctos en cada entorno.	A2 A3 B2 B3 C10 C11 C15 D3 D9

Contenidos	
Tema	
Introducción al Big Data	Quá es Big Data Aplicaciones Big Data Analítica Big Data Problemática del análisis de datos en entornos Bi gData
Preparación y visualización de datos	Técnicas de preprocesado de datos Técnicas de visualización
Infraestructuras para el almacenamiento y procesamiento de Big Data: Apache Hadoop y Apache Spark	Procesamento distribuido e infraestructuras Aprendizaje por lotes en plataformas paralelas y distribuidas Aprendizaje distribuida en vertical y horizontal
Tratamiento de datos en continuo	Aprendizaje incremental Aprendizaje en tiempo real Problemas de cambio de concepto

Planificación			
	Horas en clase	Horas fuera de clase	Horas totales
Prácticas de laboratorio	14	44	58
Trabajo tutelado	9	20	29
Lección magistral	20	21	41
Examen de preguntas objetivas	2	20	22

*Los datos que aparecen en la tabla de planificación son de carácter orientativo, considerando la heterogeneidad de alumnado

Metodologías	
	Descripción
Prácticas de laboratorio	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.
Trabajo tutelado	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.
Lección magistral	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.

Atención personalizada	
Metodologías	**Descripción**
Prácticas de laboratorio	Realización del trabajo práctico con el asesoramiento del profesor. Redacción de documentos que resumen los resultados en forma de informes o artículos, así como la presentación de los resultados con el profesor o en sesiones públicas dentro de la clase.

Trabajo tutelado	Realización del trabajo práctico con el asesoramiento del profesor. Redacción de documentos que resumen los resultados en forma de informes o artículos, así como la presentación de los resultados con el profesor o en sesiones públicas dentro de la clase.

Evaluación

	Descripción	Calificación	Resultados de Formación y Aprendizaje			
Prácticas de laboratorio	Se evaluarán las soluciones propuestas por el alumnado a las prácticas planteadas. La evaluación de prácticas puede llevarse acabo mediante una corrección por parte del profesor, una defensa de la solución aportada por parte del alumno ante el profesor o una presentación oral de la solución desarrollada. Todos los trabajos deberán ser entregados antes de las fechas que se especificarán y deberán cumplir unos requisitos mínimos de calidad para ser tenidos en consideración. Se valorará el grado de cumplimiento de las especificaciones, la metodología y rigurosidad y la presentación de resultados. Es imprescindible para aprobar haber entregado todas las prácticas indicadas cómo obligatorias. El seguimento continuado y objetivo de una participación activa supone un 10% de la nota final de la materia.	50	A1 A2 A3	B2 B3 B4 B5	C10 C11 C12 C15	D3 D7 D8 D9
Examen de preguntas objetivas	Evaluación mediante un examen final de los contenidos revisados en las clases teóricas y/o prácticas de la materia.	50	A1 A2 A3	B2	C10 C11 C12 C15	D8 D9

Otros comentarios sobre la Evaluación

Para superar la materia, debe conseguirse una *puntuación total de 5 o superior. Es imprescindible para aprobar entregar todas las prácticas indicadas cómo obligatorias.

Condición para cualificación de No Presentado: no presentar ninguna práctica y no asistir al examen final.

Los alumnos que no sean de nueva matrícula cula no conservan notas de cursos anteriores.

Oportunidad de recuperación (julio) y extraordinaria:

La valoración será igual que en la oportunidad común. Los alumnos que no entregaron los trabajos propuestos al largo del cuatrimestre los deberán entregar antes de la fecha establecida.

Condición para cualificación de No Presentado: no presentar ninguna práctica y no asistir al examen final.

El trabajo entregado deberá ser original del alumno. La entrega de trabajos no originales o con partes duplicadas (sea por copias entre compañeros o por obtención de otras fuentes...) llevará una nota global de SUSPENSO EN La CONVOCATORIA ANUAL, tanto para el/la estudiante que presente material copiado cómo la quien lo facilitara.

Fuentes de información

Bibliografía Básica

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

Bibliografía Complementaria

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

Recomendaciones

Asignaturas que se recomienda haber cursado previamente

Aprendizaje automático I/O06M193V01105
Aprendizaje Automático II/O06M193V01207
Aprendizaje profundo/O06M193V01206
Ingeniería de datos/O06M193V01102

Otros comentarios

Materia coordinada por la Universidad de A Coruña y la Universidad de Santiago.