



DATOS IDENTIFICATIVOS

IA en contornas big data

Materia	IA en contornas big data			
Código	O06M193V01303			
Titulación	Máster universitario en Intelixencia artificial			
Descritores	Creditos ECTS	Sinale	Curso	Cuadrimestre
	6	OP	2	1c
Lingua de 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&ny_academic=2023_24			
Descrición xeral	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.			

Resultados de Formación e Aprendizaxe

Código	
A1	CB6 - Posuír e comprender coñecementos que acheguen unha base ou oportunidade de ser orixinais no desenvolvemento e/ou aplicación de ideas, a miúdo nun contexto de investigación
A2	CB7 - Que os estudantes saiban aplicar os coñecementos adquiridos e a súa capacidade de resolución de problemas en contornas novas ou pouco coñecidos dentro de contextos máis amplos (ou multidisciplinares) relacionados coa súa área de estudo
A3	CB8 - Que os estudantes sexan capaces de integrar coñecementos e enfrontarse á complexidade de formular xuízos a partir dunha información que, sendo incompleta ou limitada, inclúa reflexións sobre as responsabilidades sociais e éticas vinculadas á aplicación dos seus coñecementos e xuízos
B2	Abordar con éxito todas as etapas dun proxecto de Intelixencia Artificial
B3	Buscar e seleccionar a información útil necesaria para resolver problemas complexos, manexando con soltura as fontes bibliográficas do campo.
B4	Elaborar adecuadamente e con certa orixinalidade composicións escritas ou argumentos motivados, redactar plans, proxectos de traballo, artigos científicos e formular hipóteses razoables no campo.
B5	Traballar en equipo, especialmente de carácter multidisciplinar, e ser hábiles na xestión do tempo, persoas e toma de decisións.
C10	Capacidade para a construción, validación e aplicación dun modelo estocástico dun sistema real a partir dos datos observados e a análise crítica dos resultados obtidos.
C11	Comprensión e dominio das principais técnicas e ferramentas de análises de datos, tanto desde o punto de vista estatístico como da aprendizaxe automática, incluíndo as dedicadas ao tratamento de grandes volumes de datos, e capacidade para seleccionar as máis adecuadas para a resolución de problemas.
C12	Capacidade para expor, formular e resolver todas as etapas dun proxecto de datos, incluíndo a compresión e dominio de fundamentos e técnicas básicas para a procura e o filtrado de información en grandes coleccións de datos.
C15	Coñecemento das ferramentas informáticas no campo da aprendizaxe automática, e capacidade para seleccionar a máis adecuada para a resolución dun problema
D3	Utilizar as ferramentas básicas das tecnoloxías da información e as comunicacións (TIC) necesarias para o exercicio da súa profesión e para a aprendizaxe ao longo da súa vida.

D7	Desenvolver a capacidade de traballar en equipos interdisciplinares ou transdisciplinares, para ofrecer propostas que contribúan a un desenvolvemento sustentable ambiental, económico, político e social.
D8	Valorar a importancia que ten a investigación, a innovación e o desenvolvemento tecnolóxico no avance socioeconómico e cultural da sociedade
D9	Ter a capacidade de xestionar tempos e recursos: desenvolver plans, priorizar actividades, identificar as críticas, establecer prazos e cumprilos.

Resultados previstos na materia

Resultados previstos na materia	Resultados de Formación e Aprendizaxe
Coñecer as técnicas que permiten o deseño de técnicas de IA escalables a nivel software e de recursos hardware	A2 B2 C10 C11 C12 C15 D3
Adquirir as competencias que permitan integran gran volume e variedade de datos en proxectos de Big Data en IA	A1 A2 A3 B3 B4 B5 C10 C11 C12 C15 D3 D7 D8 D9
Coñecer as paradigmas de escalabilidade en algoritmos de aprendizaxe automática	A1 A2 A3 B2 B3 B4 B5 C10 C11 C12 C15 D3 D7 D8 D9
Comprender, analizar e deseñar as infraestruturas necesarias para proxectos de IA en BigData: contorna local/nube e equipamento físico/virtual con sistemas de almacenamento de baixa latencia e sistemas de ficheiros distribuídos.	A1 A2 A3 B2 C12 C15 D3 D7 D9
Coñecer as linguaxes, frameworks e compoñentes que nos permiten incrementar o rendemento nas infraestruturas hardware con CPU e GPU	A1 A2 B3 B5 C11 C15 D3 D7 D9

Cofecer as técnicas que permiten, con baixa latencia, a visualización de datos en contornas con gran volume de información.	A1 A3 B2 B3 B5 C11 C12 C15 D3 D7 D8 D9
---	---

Usar e poder aplicar os KPI correctos en cada contorna.	A2 A3 B2 B3 C10 C11 C15 D3 D9
---	---

Contidos

Tema	
Introdución 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

Planificación

	Horas na aula	Horas fóra da aula	Horas totais
Prácticas de laboratorio	14	44	58
Traballo tutelado	9	20	29
Lección maxistral	20	21	41
Exame de preguntas obxectivas	2	20	22

*Os datos que aparecen na táboa de planificación son de carácter orientador, considerando a heteroxeneidade do alumnado.

Metodoloxía docente

	Descrición
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.
Traballo 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 maxistral	(*)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

Metodoloxías	Descrición
Prácticas de laboratorio	
Traballo tutelado	

Avaliación

Descrición	Cualificación	Resultados de Formación e Aprendizaxe
Prácticas de laboratorio	50	A1 B2 C10 D3 A2 B3 C11 D7 A3 B4 C12 D8 B5 C15 D9
Exame de preguntas obxectivas	50	A1 B2 C10 D8 A2 C11 D9 A3 C12 C15

Outros comentarios sobre a Avaliación

Para superar a materia, debe conseguirse unha puntuación total de 5 ou superior. É imprescindible para aprobar entregar todas as prácticas indicadas como obrigatorias.

Condición para cualificación de Non Presentado: non presentar ningunha práctica e non asistir ao exame final.

Os alumnos que non sexan de nova matrícula non conservan notas de cursos anteriores.

Oportunidade de recuperación (xullo) e extraordinaria:

A valoración será igual que na oportunidade ordinaria. Os alumnos que non entregaron os traballos propostos ao longo do cuadrimestre deberanos entregar antes da data establecida.

Condición para cualificación de Non Presentado: non presentar ningunha práctica e non asistir ao exame final.

O traballo entregado deberá ser orixinal do alumno. A entrega de traballos non orixinais ou con partes duplicadas (sexa por copias entre compañeiros ou por obtención doutras fontes...) levará unha nota global de SUSPENSO NA CONVOCATORIA ANUAL, tanto para o/a estudante que presente material copiado como a quen o facilítase.

Bibliografía. Fontes 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

Recomendacións

Materias que se recomenda ter cursado previamente

Aprendizaxe automático I/O06M193V01105

Aprendizaxe Automático II/O06M193V01207

Aprendizaxe profunda/O06M193V01206

Enxeñaría de datos/O06M193V01102

Outros comentarios

Materia cordenada polas universidades de Santiago e A Coruña