



Guía docente				
Datos Identificativos				2022/23
Asignatura (*)	Aprendizaje Automático a Gran Escala		Código	614G02032
Titulación	Grao en Ciencia e Enxeñaría de Datos			
Descriptorios				
Ciclo	Periodo	Curso	Tipo	Créditos
Grado	1º cuatrimestre	Cuarto	Optativa	6
Idioma	Castellano			
Modalidad docente	Presencial			
Prerrequisitos				
Departamento	Ciencias da Computación e Tecnoloxías da Información			
Coordinador/a	Hernandez Pereira, Elena Maria	Correo electrónico	elena.hernandez@udc.es	
Profesorado	Alvarez Estevez, Diego Blanco Mallo, Eva Cancela Barizo, Brais Fontenla Romero, Oscar Hernandez Pereira, Elena Maria Morán Fernández, Laura	Correo electrónico	diego.alvareze@udc.es eva.blanco@udc.es brais.cancela@udc.es oscar.fontenla@udc.es elena.hernandez@udc.es laura.moranf@udc.es	
Web	campusvirtual.udc.gal			
Descripción general	<p>En esta materia se aborda el área del aprendizaje automático en entornos donde existe una gran cantidad de datos a analizar. En este contexto surgen ciertas problemáticas que hacen que muchos de los sistemas de aprendizaje clásicos no sean directamente aplicables por motivos de complejidad computacional. El aprendizaje automático a escala trata dos ámbitos de escalabilidad diferentes. La primera es el entrenamiento de un modelo con grandes conjuntos de datos, que necesita las funcionalidades de la escalabilidad sobre un clúster de ordenadores para realizar el entrenamiento. El segundo se centra en la puesta en operación del modelo entrenado de manera que se pueda escalar para cumplir las necesidades de las aplicaciones que lo consumen.</p> <p>Con el aprendizaje automático a gran escala, el enfoque se desplaza hacia los datos y la tarea. El tiempo dedicado a la tarea y los datos es significativo y, a menudo, mucho más grande de lo previsto. El objetivo de esta materia es proporcionar una visión clara sobre los conceptos del 'Big Data' en el ámbito del aprendizaje automático.</p>			

Competencias del título	
Código	Competencias del título
A24	CE24 - Comprensión y dominio de las principales técnicas básicas y avanzadas de aprendizaje automático, incluyendo las dedicadas al tratamiento de grandes volúmenes de datos.
A25	CE25 - Capacidad para identificar la adecuación de cada una de las técnicas de aprendizaje automático a la resolución de un problema, incluyendo los aspectos relacionados con su complejidad computacional o su capacidad explicativa, de acuerdo a los requisitos establecidos.
A26	CE26 - Conocimiento de las herramientas informáticas actuales en el campo del aprendizaje automático, y capacidad para seleccionar la más adecuada para la resolución de un problema.
B2	CB2 - Que los estudiantes sepan aplicar sus conocimientos a su trabajo o vocación de una forma profesional y posean las competencias que suelen demostrarse por medio de la elaboración y defensa de argumentos y la resolución de problemas dentro de su área de estudio
B3	CB3 - Que los estudiantes tengan la capacidad de reunir e interpretar datos relevantes (normalmente dentro de su área de estudio) para emitir juicios que incluyan una reflexión sobre temas relevantes de índole social, científica o ética
B4	CB4 - Que los estudiantes puedan transmitir información, ideas, problemas y soluciones a un público tanto especializado como no especializado
B7	CG2 - Elaborar adecuadamente y con cierta originalidad composiciones escritas o argumentos motivados, redactar planes, proyectos de trabajo, artículos científicos y formular hipótesis razonables.
B8	CG3 - Ser capaz de mantener y extender planteamientos teóricos fundados para permitir la introducción y explotación de tecnologías nuevas y avanzadas en el campo.



B9	CG4 - Capacidad para abordar con éxito todas las etapas de un proyecto de análisis de datos: exploración previa de los datos, preprocesado, análisis, visualización y comunicación de resultados.
B10	CG5 - Ser capaz de trabajar en equipo, especialmente de carácter multidisciplinar, y ser hábiles en la gestión del tiempo, personas y toma de decisiones.
C1	CT1 - 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.
C4	CT4 - Valorar la importancia que tiene la investigación, la innovación y el desarrollo tecnológico en el avance socioeconómico y cultural de la sociedad.

Resultados de aprendizaje			
Resultados de aprendizaje	Competencias del título		
Identificar y saber afrontar los problemas más frecuentes relacionados con la explosión de datos, conocidos como las "Vs del Big Data" y cómo algunos de ellos afectan a la aplicación de técnicas de Aprendizaje Automático.	A25	B2 B9 B10	C1 C4
Conocer los métodos y técnicas más representativos y actuales de preprocesado de datos para tratar grandes volúmenes de datos.	A24 A25 A26	B9	C1
Conocer los métodos y técnicas más representativas y actuales de Aprendizaje Automático en entornos afectados por problemas como el volumen, la velocidad o la privacidad de los datos.	A24 A25 A26	B3 B10	C1 C4
Saber manejar las herramientas y entornos de trabajo más actuales en el ámbito del aprendizaje automático para tratar grandes volúmenes de datos.	A24	B2 B4 B7	C1
Conocer técnicas para la representación visual de datos complejos y saber utilizar herramientas de visualización de datos para poder comunicar eficazmente los resultados de los análisis realizados.	A24 A26	B2 B3 B4 B7 B8 B9 B10	C1
Conocer técnicas analíticas y escalables basadas en grafos.	A24 A26	B2 B8 B9 B10	C1

Contenidos	
Tema	Subtema
Problemáticas del análisis de datos en entornos "Big Data";	
Tratamiento y visualización de grandes volúmenes de datos	Técnicas de preprocesado de datos Técnicas de visualización
Aprendizaje distribuido	Aprendizaje por lotes en plataformas paralelas y distribuidas Aprendizaje distribuido en vertical y horizontal
Aprendizaje federado	Aprendizaje en el borde Preservación de la privacidad
Tratamiento de datos en continuo	Aprendizaje incremental Aprendizaje en tiempo real Problemas de cambio de concepto
Analítica de datos basada en grafos	



Planificación				
Metodoloxías / probas	Competencias	Horas presenciales	Horas no presenciales / traballo autónomo	Horas totales
Solución de problemas	A24 A25 A26 B2 B4 B7 B10 C1 C4	0	40	40
Prueba objetiva	A24 A25 A26 B2 B3 B8 B9	3	0	3
Prácticas de laboratorio	A24 A25 A26 B2 B3 B7 B9 B10 C1	21	21	42
Sesión magistral	A25 A26 B2 B3 B8 B9	21	42	63
Atención personalizada		2	0	2

(\*) Los datos que aparecen en la tabla de planificación són de carácter orientativo, considerando la heterogeneidad de los alumnos

Metodoloxías	
Metodoloxías	Descrición
Solución de problemas	Se desenvolverán exemplos e exercicios en los que los alumnos tendrán que aplicar los conocimientos teóricos de la asignatura a casos concretos. Se garantizará la interactividad, resolviendo dudas por parte de los alumnos y animándolos a contrastar sus soluciones y plantear cuestiones relevantes. Parte de los problemas realizados serán evaluados.
Prueba objetiva	Se llevará a cabo una evaluación de la materia mediante una prueba que incluirá tanto preguntas sobre los contenidos teóricos, como supuestos prácticos y exercicios de aplicación relacionados con los distintos temas vistos en la asignatura.
Prácticas de laboratorio	Se desenvolverán exemplos e exercicios en los que los alumnos tendrán que aplicar los conocimientos teóricos de la asignatura a casos concretos. Se garantizará la interactividad, resolviendo dudas por parte de los alumnos y animándolos a contrastar sus soluciones y plantear cuestiones relevantes.
Sesión magistral	Clases magistrales donde se expondrán los conceptos teóricos de la asignatura, sin perder nunca de vista exemplos de aplicación para motivar y contextualizar los contenidos de la materia. Se fomentará la interactividad en clase mediante la formulación de preguntas y se utilizarán distintos recursos como transparencias o demostraciones.

Atención personalizada	
Metodoloxías	Descrición
Solución de problemas Prácticas de laboratorio	<p>El desarrollo, tanto de las clases magistrales como de las de resolución de problemas y los laboratorios de prácticas, se realizará atendiendo al progreso de los alumnos en las capacidades de comprensión y asimilación de los contenidos impartidos. El avance general de la clase se compaginará con una atención específica a aquellos alumnos que presenten mayores dificultades en la tarea del aprendizaje y con un apoyo adicional a aquellos que presenten mayor desentortura y deseen ampliar conocimientos.</p> <p>La atención personalizada se realiza a través de los siguientes canales:</p> <ul style="list-style-type: none"> <li>- Correo-e: De uso para hacer consultas de respuesta corta o solicitar encuentros para resolver dudas.</li> <li>- Teams: De uso en el horario oficial de tutorías en los que se garantiza una respuesta inmediata. Sin embargo, dado el elevado número de estudiantes, se recomienda solicitud previa de tutoría a través del correo-e.</li> <li>- Campus virtual: Se pondrán a disposición de los estudiantes "foros temáticos" que resuelvan dudas generales ligadas a las actividades específicas como las prácticas.</li> </ul>

Evaluación			
Metodoloxías	Competencias	Descrición	Calificación
Solución de problemas	A24 A25 A26 B2 B4 B7 B10 C1 C4	Se valorarán los resultados, forma y condiciones de realización de diversos trabajos puntuables que se detallarán durante el curso.	40



Prueba objetiva	A24 A25 A26 B2 B3 B8 B9	Realización obligatoria. Se evaluará el dominio de los conocimientos teóricos y operativos de la asignatura.	60
Prácticas de laboratorio	A24 A25 A26 B2 B3 B7 B9 B10 C1	Son obligatorias para poder aprobar los trabajos de la solución de problemas e influyen en la calificación final de estos, pero no se puntúan al margen de la nota otorgada a la solución de problemas.	0

### Observaciones evaluación

Para poder aprobar la materia el estudiante deberá cumplir los siguientes requisitos (puntuación entre 0 y 10 en todas las actividades):- Lograr una nota superior o igual a 3,5 en la prueba mixta final realizada al final del cuatrimestre.- Lograr una nota superior o igual a 5 al realizar la suma de todas las pruebas de evaluación. Notas sobre los trabajos prácticos y solución de problemas:- Todas las actividades tendrán una única oportunidad para su entrega durante lo curso académico, salvo la prueba objetiva final que tendrá dos oportunidades oficiales de examen.

Sobre la responsabilidad compartida de los trabajos en grupo.- En las actividades que se llevan a cabo en grupos, todos los miembros del grupo serán responsables solidarios del trabajo realizado y entregado, así como de las consecuencias que se deriven del incumplimiento de las normas de autoría del mismo.

Matrícula a tiempo parcial

- Los alumnos matriculados a tiempo parcial tendrán que entregar las actividades evaluables en las condiciones y plazos específicos que se establecerán. Será obligación del estudiante comunicar su situación al profesorado.

No presentado

-

Quien no concurra a la prueba objetiva en el período oficial de evaluación tendrá la condición de "No presentado" (NP). En la primera oportunidad, esto será extensible a quien no entregue las prácticas en tiempo y forma.

### Fuentes de información

**Básica**

- Jules S. Damji, Brooke Wenig, Tathagata das, and Denny Lee (2020). Learning Spark : Lightning-fast big data analytics. Sebastopol, CA : O'Reilly
- Bill Chambers and Matei Zaharia (2018). Spark : the definitive guide :big data processing made simple . Sebastopol, CA : O'Reilly Media, Inc.

**Complementaria**

### Recomendaciones

#### Asignaturas que se recomienda haber cursado previamente

Procesamiento Paralelo/614G02023  
Aprendizaje Automático III/614G02026  
Aprendizaje Automático I/614G02019  
Aprendizaje Automático II/614G02021

#### 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