



DATOS IDENTIFICATIVOS

Ingeniería de datos

Asignatura	Ingeniería de datos			
Código	O06M193V01102			
Titulación	Máster universitario en Inteligencia artificial			
Descriptores	Creditos ECTS	Seleccione	Curso	Cuatrimestre
	3	OB	1	1c
Lengua Impartición	Inglés			
Departamento				
Coordinador/a	García Lourenco, Analia María			
Profesorado	García Lourenco, Analia María			
Correo-e	analia@uvigo.es			
Web	http://http://moovi.uvigo.gal			
Descripción general	El objetivo de la asignatura 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.			

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.
C16	Conocimiento del proceso y las herramientas para el procesamiento y preparación de datos desde su adquisición o extracción, limpieza, transformación, carga, organización y acceso.
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
------------------------------------	---------------------------------------

RA1: Desarrollar la capacidad de analizar y modelar datos para su procesado en sistemas inteligentes.	A1 A2 C16 D3 D9
RA2: Conocer y comprender el proceso de extracción, limpieza, transformación, carga y preprocesado de datos.	A3 B2 B3 C16 D3 D7 D9
RA3: Conocer y saber utilizar bases de datos multidimensionales y de tipo NoSQL.	A2 B3 B4 D8
RA4: Conocer los fundamentos de data lakes y data warehouses.	A2 A3 B2 B5 D3 D7 D8

Contenidos

Tema	
Conceptos y fundamentos de Ingeniería de datos	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 de datos.	Técnicas más comunes. Definición de flujos de procesamiento. Métricas de calidad.
Estructuras avanzadas y almacenes de datos eficientes para Big Data	Data warehouses y BD multidimensionales, Data lakes, Bases de Datos NoSQL.

Planificación

	Horas en clase	Horas fuera de clase	Horas totales
Lección magistral	12	0	12
Prácticas de laboratorio	8	0	8
Proyecto	0	50	50
Resolución de problemas y/o ejercicios	5	0	5

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

Metodologías

	Descripción
Lección magistral	El profesorado presenta un tema al alumnado con el objetivo de facilitar un conjunto de información con alcance concreto. EVALUACIÓN CONTINUA Carácter obligatorio Asistencia no obligatoria EVALUACIÓN GLOBAL Carácter obligatorio
Prácticas de laboratorio	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. EVALUACIÓN CONTINUA Carácter obligatorio Asistencia obligatoria (min. 75% de las horas de prácticas) EVALUACIÓN GLOBAL Carácter obligatorio

Atención personalizada

Pruebas	Descripción
---------	-------------

Proyecto	Se atenderán las dudas relacionadas con la planificación y el desarrollo del proyecto.
Resolución de problemas y/o ejercicios	Se atenderán las dudas relacionadas con los casos a analizar.

Evaluación							
	Descripción	Calificación		Resultados de Formación y Aprendizaje			
Prácticas de laboratorio	Entrega de prácticas de laboratorio planteadas a lo largo del curso en las fechas estipuladas previamente. Resultados de aprendizaje evaluados: RA3, RA4	30	A2	B2 B5	C16	D3 D7	
Proyecto	Desarrollo de un proyecto en que se valora la capacidad de trabajo autónomo y el espíritu crítico del alumno y su capacidad para aplicar los conocimientos adquiridos a entornos reales. Al final, el alumnado deberá entregar una memoria detallada del proyecto. Resultados de aprendizaje evaluados: RA2, RA4	40	A2	B2 B4	C16	D8 D9	
Resolución de problemas y/o ejercicios	Se llevará a cabo la defensa de la solución aportada por parte del alumno ante el profesor y el resto del alumnado. Resultados de aprendizaje evaluados: RA1, RA2	30	A1 A2 A3	B2 B3	C16	D7 D8	

Otros comentarios sobre la Evaluación

SISTEMA DE EVALUACIÓN CONTINUA

PRUEBA 1: Prácticas de laboratorio

Descripción: Entrega de prácticas de laboratorio planteadas a lo largo del curso en las fechas estipuladas previamente.

Metodología(s) aplicada(s): Prácticas de laboratorio.

% Calificación: 30%.

% Mínimo: Para la liberación de esta parte de la asignatura el estudiante deberá obtener una calificación igual o superior a 5 puntos (sobre 10).

Competencias evaluadas: A2,B2,B5,C16,D3,D7.

Resultados de aprendizaje evaluados: RA3, RA4

PRUEBA 2: Resolución de problemas y/o ejercicios

Descripción: Se llevará a cabo la defensa de la solución aportada por parte del alumno ante el profesor y el resto del alumnado.

Metodología(s) aplicada(s): Resolución de problemas y/o ejercicios.

% Calificación: 30%.

% Mínimo: Para la liberación de esta parte de la asignatura el estudiante deberá obtener una calificación igual o superior a 5 puntos (sobre 10).

Competencias evaluadas: A1,A2,A3,B2,B3,C16,D7,D8

Resultados de aprendizaje evaluados: RA1, RA2

PRUEBA 3: Proyecto

Descripción: 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.

Metodología(s) aplicada(s): Proyecto.

% Calificación: 40%

% Mínimo: Para la liberación de esta parte de la asignatura el estudiante deberá obtener una calificación igual o superior a 5 puntos (sobre 10).

Competencias evaluadas: A2,B2,B4,C16,D8,D9.

Resultados de aprendizaje evaluados: RA2, RA4

SISTEMA DE EVALUACIÓN GLOBAL

Procedimiento para la elección de la modalidad de evaluación global: se considera que el estudiantado opta por el sistema de evaluación global si no se presenta a la Prueba 1 del sistema de evaluación continua.

PRUEBA 1: Prácticas de laboratorio

Descripción: Entrega de prácticas de laboratorio planteadas a lo largo del curso en las fechas estipuladas previamente.

Metodología(s) aplicada(s): Prácticas de laboratorio.

% Calificación: 30%.

% Mínimo: Para la liberación de esta parte de la asignatura el estudiante deberá obtener una calificación igual o superior a 5 puntos (sobre 10).

Competencias evaluadas: A2,B2,B5,C16,D3,D7.

Resultados de aprendizaje evaluados: RA3, RA4

PRUEBA 2: Resolución de problemas y/o ejercicios

Descripción: Se llevará a cabo la defensa de la solución aportada por parte del alumno ante el profesor y el resto del alumnado.

Metodología(s) aplicada(s): Resolución de problemas y/o ejercicios.

% Calificación: 30%.

% Mínimo: Para la liberación de esta parte de la asignatura el estudiante deberá obtener una calificación igual o superior a 5 puntos (sobre 10).

Competencias evaluadas: A1,A2,A3,B2,B3,C16,D7,D8

Resultados de aprendizaje evaluados: RA1, RA2

PRUEBA 3: Proyecto

Descripción: 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.

Metodología(s) aplicada(s): Proyecto.

% Calificación: 40%

% Mínimo: Para la liberación de esta parte de la asignatura el estudiante deberá obtener una calificación igual o superior a 5 puntos (sobre 10)).

Competencias evaluadas: A2,B2,B4,C16,D8,D9.

Resultados de aprendizaje evaluados: RA2, RA4

CRITERIOS DE EVALUACIÓN PARA CONVOCATORIA EXTRAORDINARIA Y FIN DE CARRERA

Se emplearán los sistemas de evaluación continua y global expuestos anteriormente.

PROCESO DE CALIFICACIÓN DE ACTAS

Independientemente del sistema de evaluación y la convocatoria, en caso de no superar alguna parte de la evaluación, pero la puntuación global fuese superior a 4 (sobre 10), la calificación en actas será 4.

FECHAS DE EVALUACIÓN

Las fechas de las pruebas correspondientes al sistema de evaluación continua se publicará en el calendario de actividades, disponible en la página web de la ESEI <https://esei.uvigo.es/docencia/horarios/>.

Las fechas oficiales de examen de las diferentes convocatorias, aprobadas oficialmente por la Xunta de Centro de la ESEI, se encuentran publicadas en la página web de la ESEI <https://esei.uvigo.es/docencia/exames/>.

EMPLEO DE DISPOSITIVOS MÓVILES

Se recuerda a todo el alumnado la prohibición del uso de dispositivos móviles en ejercicios y prácticas, en cumplimiento del artículo 13.2.d) del Estatuto del Estudiante Universitario, relativo a los deberes del estudiantado universitario, que establece el deber de "Abstenerse de la utilización o cooperación en procedimientos fraudulentos en las pruebas de evaluación, en los trabajos que se realicen o en documentos oficiales de la universidad."

CONSULTA/SOLICITUD DE TUTORÍAS

Las tutorías pueden consultarse a través de la página personal del profesorado, accesible a través de <https://esei.uvigo.es/docencia/profesorado/>

OTRAS CONSIDERACIONES

En el caso de detección de plagio en alguno de los trabajos (teóricos o práctico), la calificación final de la materia será de Suspenso (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

Bibliografía Básica

Sadalage, Fowler, **NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence**, 978-0321826626, 1, Addison-Wesley, 2012

Avi Silberschatz, Henry F. Korth, S. Sudarshan, **Database System Concepts**, 0-07-352332-1, 6, McGraw-Hill, 2010

Ihab F. Ilyas and Xu Chu, **Data Cleaning. Association for Computing Machinery**, <https://doi.org/10.1145/3310205>, Association for Computing Machinery, 2019

Alex Gorelik, **The Enterprise Big Data Lake: Delivering the Promise of Big Data and Data Science**, 9781491931554, O'Reilly Media, Inc., 2019

Matt Casters, Roland Bouman, Jos van Dongen,, **Pentaho Kettle Solutions: Building Open Source ETL Solutions with Pentaho Data Integration**, 978-0470635179, Wiley, 2013

Bibliografía Complementaria

Recomendaciones

Otros comentarios

Seguir la metodología propuesta, asistiendo a las clases, dedicando el tiempo necesario al estudio y a la realización de trabajos y resolviendo problemas específicos con la ayuda del profesorado en las sesiones de tutorías.

Se hará uso del campus virtual, para mejorar la comunicación entre el alumnado y el profesorado, para alojar el material necesario y para apoyar en los procesos de evaluación.