

Facultad de Ciencias

GUÍA DOCENTE DE LA ASIGNATURA

M1966 - Minería de Datos

Máster Universitario en Ciencia de Datos / Master in Data Science
Obligatoria. Curso 1

Curso Académico 2022-2023

1. DATOS IDENTIFICATIVOS

Título/s	Máster Universitario en Ciencia de Datos / Master in Data Science	Tipología y Curso	Obligatoria. Curso 1
Centro	Facultad de Ciencias		
Módulo / materia	FUNDAMENTOS MÉTODOS EN CIENCIA DE DATOS		
Código y denominación	M1966 - Minería de Datos		
Créditos ECTS	6	Cuatrimestre	Cuatrimestral (1)
Web			
Idioma de impartición	Español	English friendly	Sí
		Forma de impartición	Presencial

Departamento	DPTO. MATEMATICA APLICADA Y CIENCIAS DE LA COMPUTACION
Profesor responsable	SIXTO HERRERA GARCIA
E-mail	sixto.herrera@unican.es
Número despacho	E.T.S. de Ingenieros de Caminos, Canales y Puertos. Planta: + 1. DESPACHO PROFESORES (1034)
Otros profesores	STEVEN JOHAN MARIA VAN VAERENBERGH JOAQUIN BEDIA JIMENEZ RODRIGO GARCIA MANZANAS ANA CASANUEVA VICENTE

2. CONOCIMIENTOS PREVIOS

Los requeridos en la admisión al Máster.

3. COMPETENCIAS GENÉRICAS Y ESPECÍFICAS DEL PLAN DE ESTUDIOS TRABAJADAS

Competencias Genéricas
Integrarse eficazmente en un grupo de trabajo y trabajar en equipo, compartir la información disponible e integrar su actividad en la actividad del grupo colaborando de forma activa en la consecución de objetivos comunes.
Capacidad de estudio, síntesis y autonomía suficientes para desarrollar de forma autónoma proyectos básicos de investigación.
Buscar, obtener, procesar, comunicar información y transformarla en conocimiento.
Conocer las herramientas metodológicas necesarias para desarrollar proyectos avanzados.
Capacidad de actualización de los conocimientos expuestos en el ámbito de la comunidad científica.
Competencias Específicas
Utilizar el análisis predictivo para analizar grandes volúmenes de datos y descubrir nuevas relaciones.
Utilizar técnicas estadísticas apropiadas sobre los datos disponibles para lograr una visión adecuada de los mismos.
Aplicar el ingenio propio para resolver problemas complejos y desarrollar ideas innovadoras.
Competencias Básicas
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.
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.
Que los estudiantes sepan comunicar sus conclusiones y los conocimientos y razones últimas que las sustentan a públicos especializados y no especializados de un modo claro y sin ambigüedades.
Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo.
Competencias Transversales
Analizar y combinar información utilizando diferentes fuentes.
Dominio de la gestión del tiempo.
Capacidad de trabajo autónomo y toma de decisiones.
Capacidades asociadas al trabajo en equipo: cooperación, liderazgo, saber escuchar.

3.1 RESULTADOS DE APRENDIZAJE

- Entender las técnicas para la descripción estadística de un conjunto de datos y saber aplicar críticamente las que mejor se adapten a un caso práctico.
- Conocer y aplicar el concepto de regularización en problemas de minería de datos.
- Saber identificar los distintos problemas de minería de datos (asociación, segmentación, clasificación y predicción) y saber qué técnicas se aplican en cada caso.
- Entender las principales metodologías para aprendizaje automático (basado en casos, supervisado o no supervisado) .
- Entender el funcionamiento y saber aplicar las principales técnicas de minería de datos.
- Conocer el concepto de generalización y entender el problema del sobreajuste.
- Saber validar de forma crítica los modelos aprendidos y saber identificar el modelo más adecuado para un caso práctico concreto.
- Saber utilizar paquetes de software para resolver casos prácticos de minería de datos y analizar críticamente las soluciones obtenidas.

4. OBJETIVOS

Esta materia tiene como objetivo proporcionar a los estudiantes conocimientos básicos sobre los métodos estadísticos para la descripción, presentación y análisis de datos, así como los principios generales para la aplicación práctica de las principales técnicas automáticas de modelización y predicción de minería de datos para el tratamiento de grandes masas de datos.

5. MODALIDADES ORGANIZATIVAS Y MÉTODOS DOCENTES

ACTIVIDADES	HORAS DE LA ASIGNATURA
ACTIVIDADES PRESENCIALES	
HORAS DE CLASE (A)	
- Teoría (TE)	13
- Prácticas en Aula (PA)	8
- Prácticas de Laboratorio Experimental(PLE)	
- Prácticas de Laboratorio en Ordenador (PLO)	24
- Prácticas Clínicas (CL)	
Subtotal horas de clase	45
ACTIVIDADES DE SEGUIMIENTO (B)	
- Tutorías (TU)	12
- Evaluación (EV)	3
Subtotal actividades de seguimiento	15
Total actividades presenciales (A+B)	60
ACTIVIDADES NO PRESENCIALES	
Trabajo en grupo (TG)	30
Trabajo autónomo (TA)	60
Tutorías No Presenciales (TU-NP)	
Evaluación No Presencial (EV-NP)	
Total actividades no presenciales	90
HORAS TOTALES	150

6. ORGANIZACIÓN DOCENTE

CONTENIDOS		TE	PA	PLE	PLO	CL	TU	EV	TG	TA	TU-NP	EV-NP	Semana
1	Introducción y perspectiva histórica; Problemas Canónicos, Paradigmas de Aprendizaje y Validación; Aprendizaje no supervisado con Reglas de Asociación; Aprendizaje supervisado con técnicas de vecinos k-nn; Cross-Validación y Aprendizaje	6,00	2,00	0,00	8,00	0,00	4,00	0,00	10,00	20,00	0,00	0,00	1-3
2	Regularización y selección de variables Reducción de la dimensionalidad con Técnicas lineales y No lineales	4,00	4,00	0,00	6,00	0,00	3,00	0,00	7,00	15,00	0,00	0,00	3-4
3	Técnicas de embedding Técnicas basadas en árboles Ensemble Methods: Bagging and Random Forests	2,00	2,00	0,00	6,00	0,00	3,00	0,00	7,00	15,00	0,00	0,00	4-5
4	Técnicas de Segmentación	1,00	0,00	0,00	4,00	0,00	2,00	3,00	6,00	10,00	0,00	0,00	6
TOTAL DE HORAS		13,00	8,00	0,00	24,00	0,00	12,00	3,00	30,00	60,00	0,00	0,00	

Esta organización tiene carácter orientativo.

TE	Horas de teoría
PA	Horas de prácticas en aula
PLE	Horas de prácticas de laboratorio experimental
PLO	Horas de prácticas de laboratorio en ordenador
CL	Horas de prácticas clínicas
TU	Horas de tutoría
EV	Horas de evaluación
TG	Horas de trabajo en grupo
TA	Horas de trabajo autónomo
TU-NP	Tutorías No Presenciales
EV-NP	Evaluación No Presencial

7. MÉTODOS DE LA EVALUACIÓN

Descripción	Tipología	Eval. Final	Recuper.	%										
Valoración de informes y trabajos escritos	Actividad de evaluación con soporte virtual	Sí	Sí	60,00										
<table border="1"> <tr> <td>Calif. mínima</td> <td>3,00</td> </tr> <tr> <td>Duración</td> <td></td> </tr> <tr> <td>Fecha realización</td> <td>Durante el periodo de impartición de la asignatura.</td> </tr> <tr> <td>Condiciones recuperación</td> <td></td> </tr> <tr> <td>Observaciones</td> <td>Evaluación de los trabajos de grupo e individuales entregados por el alumno.</td> </tr> </table>					Calif. mínima	3,00	Duración		Fecha realización	Durante el periodo de impartición de la asignatura.	Condiciones recuperación		Observaciones	Evaluación de los trabajos de grupo e individuales entregados por el alumno.
Calif. mínima	3,00													
Duración														
Fecha realización	Durante el periodo de impartición de la asignatura.													
Condiciones recuperación														
Observaciones	Evaluación de los trabajos de grupo e individuales entregados por el alumno.													
Examen (escrito, oral y/o práctico en el aula de computación)	Actividad de evaluación con soporte virtual	Sí	Sí	40,00										
<table border="1"> <tr> <td>Calif. mínima</td> <td>3,00</td> </tr> <tr> <td>Duración</td> <td>Un máximo de dos horas</td> </tr> <tr> <td>Fecha realización</td> <td>Durante el periodo de impartición de la asignatura.</td> </tr> <tr> <td>Condiciones recuperación</td> <td></td> </tr> <tr> <td>Observaciones</td> <td></td> </tr> </table>					Calif. mínima	3,00	Duración	Un máximo de dos horas	Fecha realización	Durante el periodo de impartición de la asignatura.	Condiciones recuperación		Observaciones	
Calif. mínima	3,00													
Duración	Un máximo de dos horas													
Fecha realización	Durante el periodo de impartición de la asignatura.													
Condiciones recuperación														
Observaciones														
TOTAL				100,00										
Observaciones														
<p>Las calificaciones mínimas se aplican a los dos bloques de evaluación, la evaluación continua y el examen final. En base al artículo 35 del Reglamento, en el caso de que un estudiante no obtuviese la calificación mínima requerida en alguna de las pruebas de evaluación, la calificación global de la asignatura será el menor valor entre 4,9 y la media ponderada de las pruebas de evaluación. Con carácter obligatorio todas las pruebas se realizarán o entregarán usando la plataforma virtual de la asignatura. Por tanto es responsabilidad del alumno, asegurarse de que pueden acceder a la plataforma virtual de la asignatura, antes del comienzo de las sesiones en las que se realicen las pruebas.</p> <p>Si la nota final del alumno fuese menor que 5 sobre 10, entonces la recuperación consistirá en la realización de cada una de las tareas en las que hubiera obtenido una calificación menor que 5 sobre 10. El procedimiento de evaluación de una actividad recuperable será equivalente al de la actividad original.</p>														
Criterios de evaluación para estudiantes a tiempo parcial														
Se aplicará la misma evaluación que a los estudiantes a tiempo completo, teniendo en cuenta su disponibilidad de tiempo en cuanto a los plazos de entrega de los trabajos.														

8. BIBLIOGRAFÍA Y MATERIALES DIDÁCTICOS
BÁSICA

An Introduction to Statistical Learning with Applications in R, Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani, Editorial Springer. Available at <http://www-bcf.usc.edu/~gareth/ISL/>
 Dimension Reduction: A Guided Tour, Christopher J. C. Burges, Foundations and Trends® in Machine Learning: Vol. 2: No. 4, pp 275-365. <http://dx.doi.org/10.1561/22000000002>
 Machine learning: a probabilistic perspective, Kevin P. Murphy. ISBN 978-0-262-01802-9, MIT Press Cambridge, Massachusetts, London, England

Complementaria

The Elements of Statistical Learning Data Mining, Inference and Prediction, Jerome Friedman, Trevor Hastie and Robert Tibshirani, Ed. Friedman, Hastie & Tibshirani

9. SOFTWARE

PROGRAMA / APLICACIÓN	CENTRO	PLANTA	SALA	HORARIO
R Software				

10. COMPETENCIAS LINGÜÍSTICAS

- Comprensión escrita
- Comprensión oral
- Expresión escrita
- Expresión oral
- Asignatura íntegramente desarrollada en inglés

Observaciones