

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 2019-2020

### 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 (1042)
Otros profesores	STEVEN JOHAN MARIA VAN VAERENBERGH JOAQUIN BEDIA JIMENEZ RODRIGO GARCIA MANZANAS MAIALEN ITURBIDE MARTINEZ DE ALBENIZ

### 2. CONOCIMIENTOS PREVIOS

Los requeridos en la admisión al Máster.

### 3. COMPETENCIAS GENÉRICAS Y ESPECÍFICAS DEL PLAN DE ESTUDIOS TRABAJADAS

<b>Competencias Genéricas</b>
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.
<b>Competencias Específicas</b>
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.
<b>Competencias Básicas</b>
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.
<b>Competencias Transversales</b>
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
<b>ACTIVIDADES PRESENCIALES</b>	
<b>HORAS DE CLASE (A)</b>	
- Teoría (TE)	13
- Prácticas en Aula (PA)	8
- Prácticas de Laboratorio (PL)	24
- Horas Clínicas (CL)	
Subtotal horas de clase	45
<b>ACTIVIDADES DE SEGUIMIENTO (B)</b>	
- Tutorías (TU)	12
- Evaluación (EV)	3
Subtotal actividades de seguimiento	15
<b>Total actividades presenciales (A+B)</b>	<b>60</b>
<b>ACTIVIDADES NO PRESENCIALES</b>	
Trabajo en grupo (TG)	30
Trabajo autónomo (TA)	60
Tutorías No Presenciales (TU-NP)	
Evaluación No Presencial (EV-NP)	
<b>Total actividades no presenciales</b>	<b>90</b>
<b>HORAS TOTALES</b>	<b>150</b>

## 6. ORGANIZACIÓN DOCENTE

CONTENIDOS		TE	PA	PL	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	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	6,00		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	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	4,00	0,00	2,00	3,00	6,00	10,00	0,00	0,00	6
<b>TOTAL DE HORAS</b>		<b>13,00</b>	<b>8,00</b>	<b>24,00</b>	<b>0,00</b>	<b>12,00</b>	<b>3,00</b>	<b>30,00</b>	<b>60,00</b>	<b>0,00</b>	<b>0,00</b>	

Esta organización tiene carácter orientativo.

TE	Horas de teoría
PA	Horas de prácticas en aula
PL	Horas de prácticas de laboratorio
CL	Horas 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
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
Calif. mínima	3,00			
Duración	Un máximo de dos horas			
Fecha realización	Durante el periodo de impartición de la asignatura.			
Condiciones recuperación				
Observaciones				
<b>TOTAL</b>				<b>100,00</b>
<b>Observaciones</b>				
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.				
<b>Observaciones para alumnos a tiempo parcial</b>				
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 <a href="http://www-bcf.usc.edu/~gareth/ISL/">http://www-bcf.usc.edu/~gareth/ISL/</a> Dimension Reduction: A Guided Tour, Christopher J. C. Burges, Foundations and Trends® in Machine Learning: Vol. 2: No. 4, pp 275-365. <a href="http://dx.doi.org/10.1561/2200000002">http://dx.doi.org/10.1561/2200000002</a> 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**