

# Minería de datos

**MÁSTER UNIVERSITARIO EN CIENCIA DE DATOS / MASTER IN  
DATA SCIENCE**

***UNIVERSIDAD INTERNACIONAL MENÉNDEZ PELAYO***

Este documento puede utilizarse como documentación de referencia de esta asignatura para la solicitud de reconocimiento de créditos en otros estudios. Para su plena validez debe estar sellado por la Secretaría de Estudiantes UIMP.



# DATOS GENERALES

## Breve descripción

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.

## Título asignatura

Minería de datos

## Código asignatura

102265

## Curso académico

2023-24

## Planes donde se imparte

[MÁSTER UNIVERSITARIO EN CIENCIA DE DATOS / MASTER IN DATA SCIENCE](#)

## Créditos ECTS

6

## Carácter de la asignatura

OBLIGATORIA

## Duración

Cuatrimestral

## Idioma

Castellano e Inglés

# CONTENIDOS

## Contenidos

1. Introducción y perspectiva histórica.
2. Problemas canónicos, paradigmas de aprendizaje y validación.
3. Aprendizaje no supervisado con Reglas de Asociación.
4. Aprendizaje supervisado con técnicas de vecinos k-nn.
5. Cross-Validación y Aprendizaje.
6. Regularización y selección de variables.
7. Reducción de la dimensionalidad con Técnicas lineales y No lineales.
8. Técnicas de embedding.
9. Técnicas basadas en árboles.
10. Ensemble Methods: Bagging and Random Forests.
11. Técnicas de Segmentación.

# COMPETENCIAS

## Generales

CG1 - 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

CG2 - Capacidad de estudio, síntesis y autonomía suficientes para desarrollar de forma autónoma proyectos básicos de investigación

CG6 - Buscar, obtener, procesar, comunicar información y transformarla en conocimiento

CG7 - Conocer las herramientas metodológicas necesarias para desarrollar proyectos avanzados

CG8 - Capacidad de actualización de los conocimientos expuestos en el ámbito de la comunidad científica

## Transversales

CT1 - Analizar y combinar información utilizando diferentes fuentes

CT3 - Dominio de la gestión del tiempo

CT5 - Capacidad de trabajo autónomo y toma de decisiones

CT6 - Capacidades asociadas al trabajo en equipo: cooperación, liderazgo, saber escuchar

## Específicas

DSDA01 - Utilizar el análisis predictivo para analizar grandes volúmenes de datos y descubrir nuevas relaciones

DSDA02 - Utilizar técnicas estadísticas apropiadas sobre los datos disponibles para lograr una visión adecuada de los mismos

DSRM06 - Aplicar el ingenio propio para resolver problemas complejos y desarrollar ideas innovadoras

# PLAN DE APRENDIZAJE

## Actividades formativas

AF1 - Participación y asistencia a lecciones magistrales y seminarios (13 horas)

AF2 - Realización de prácticas de computación y análisis de datos (32 horas)

AF6 - Tutorías (presenciales o por medio de recursos telemáticos) (12 horas)

AF7 - Elaboración de informes de laboratorio y trabajos (30 horas)

AF8 - Estudio individual de contenidos de la asignatura (30 horas)

AF9 - Trabajo en grupo (30 horas)

A10 - Pruebas de evaluación (3 horas)

## 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.
- Entender el modelo de regresión lineal y saber aplicarlo en casos prácticos.
- Conocer el concepto de regularización.
- 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.

# SISTEMA DE EVALUACIÓN

## Descripción del sistema de evaluación

SE1 - Examen (escrito, oral y/o práctico en el aula de computación) (40%)

SE2 - Valoración de informes y trabajos escritos (60%)

## PROFESORADO

### Profesor responsable

**Herrera García, Sixto**

*Profesor Ayudante Doctor de Matemática Aplicada  
Universidad de Cantabria (UC)*

### Profesorado

**Van Vaerenbergh , Steven Johan Maria**

*Investigador Contratado Doctor  
Universidad de Cantabria (UC)*

**Iturbide Martínez de Albéniz, Maialen**

*Investigadora Doctora en Agrobiología Ambiental  
Universidad de Cantabria (UC)*

**García Manzananas, Rodrigo**

*Profesor Ayudante Doctor  
Universidad de Cantabria (UC)*

**Gutiérrez Llorente, José Manuel**

*Profesor de Investigación  
Instituto de Física de Cantabria (IFCA), CSIC-UC*

**Bedia Jiménez, Joaquín**

*Profesor Doctor en Ciencia, Tecnología y Computación  
Universidad de Cantabria (UC)*

# HORARIO

## Horario

31/10/2023

17:30 - 19:30

Presentación, introducción y perspectiva histórica.

Sixto Herrera García

Profesor Ayudante Doctor de Matemática Aplicada  
Universidad de Cantabria (UC)

02/11/2023

17:30 - 18:30

Paradigmas, problemas canónicos y datasets.

Rodrigo García Manzananas

Profesor Ayudante Doctor  
Universidad de Cantabria (UC)

06/11/2023

17:30 - 19:30

Reglas de Asociación

Sixto Herrera García

Profesor Ayudante Doctor de Matemática Aplicada  
Universidad de Cantabria (UC)



07/11/2023

17:30 - 19:30

Práctica: Reglas de Asociación

Joaquín Bedia Jiménez

Profesor Doctor en Ciencia, Tecnología y Computación  
Universidad de Cantabria (UC)

14/11/2023

17:30 - 19:30

Evaluación, sobreajuste y cross-validation

Sixto Herrera García

Profesor Ayudante Doctor de Matemática Aplicada  
Universidad de Cantabria (UC)

16/11/2023

17:30 - 19:30

Cross-validation

Sixto Herrera García

Profesor Ayudante Doctor de Matemática Aplicada  
Universidad de Cantabria (UC)

21/11/2023

17:30 - 19:30

Árboles de clasificación y decisión

Rodrigo García Manzananas

Profesor Ayudante Doctor  
Universidad de Cantabria (UC)

23/11/2023

17:30 - 19:30

Práctica: Árboles de clasificación

Rodrigo García Manzananas

Profesor Ayudante Doctor  
Universidad de Cantabria (UC)

28/11/2023

15:30 - 17:30

Técnicas de vecinos cercanos (k-NN)

Joaquín Bedia Jiménez

Profesor Doctor en Ciencia, Tecnología y Computación  
Universidad de Cantabria (UC)

30/11/2023

17:30 - 19:30

Práctica: Vecinos cercanos (k-NN)

Rodrigo García Manzananas

Profesor Ayudante Doctor  
Universidad de Cantabria (UC)

05/12/2023

17:30 - 19:30

Comparación de técnicas de clasificación

Rodrigo García Manzananas

Profesor Ayudante Doctor  
Universidad de Cantabria (UC)

12/12/2023

17:30 - 19:30

Arboles de regresión

Rodrigo García Manzananas

Profesor Ayudante Doctor  
Universidad de Cantabria (UC)

14/12/2023

17:30 - 19:30

Arboles de regresión

Joaquín Bedia Jiménez

Profesor Doctor en Ciencia, Tecnología y Computación  
Universidad de Cantabria (UC)

18/12/2023

17:30 - 19:30

El paquete CARET

Sixto Herrera García

Profesor Ayudante Doctor de Matemática Aplicada  
Universidad de Cantabria (UC)

20/12/2023

17:30 - 19:30

Ensemble methods

Sixto Herrera García

Profesor Ayudante Doctor de Matemática Aplicada  
Universidad de Cantabria (UC)

21/12/2023

17:30 - 18:30

Random Forest

Joaquín Bedia Jiménez

Profesor Doctor en Ciencia, Tecnología y Computación  
Universidad de Cantabria (UC)

09/01/2024

17:30 - 18:30

Gradient Boosting

Sixto Herrera García

Profesor Ayudante Doctor de Matemática Aplicada  
Universidad de Cantabria (UC)

18:30 - 19:30

El parque CARET-ensembles

Sixto Herrera García

Profesor Ayudante Doctor de Matemática Aplicada  
Universidad de Cantabria (UC)

11/01/2024

15:30 - 17:30

Reducción de la dimensión (no lineal)

Steven Johan Maria Van Vaerenbergh

Investigador Contratado Doctor  
Universidad de Cantabria (UC)

15/01/2024

17:30 - 19:30

Práctica: Reducción de la dimensión (no lineal)

Steven Johan Maria Van Vaerenbergh

Investigador Contratado Doctor  
Universidad de Cantabria (UC)

16/01/2024

17:30 - 19:30

Clustering methods

Sixto Herrera García

Profesor Ayudante Doctor de Matemática Aplicada  
Universidad de Cantabria (UC)

18/01/2024

17:30 - 19:30

Clustering methods

Joaquín Bedia Jiménez

Profesor Doctor en Ciencia, Tecnología y Computación  
Universidad de Cantabria (UC)

22/01/2024

17:30 - 19:30

Predicción condicionada a un clustering

Rodrigo García Manzananas

Profesor Ayudante Doctor  
Universidad de Cantabria (UC)

24/01/2024

15:30 - 17:30

Sesión de refuerzo/repaso

Joaquín Bedia Jiménez

Profesor Doctor en Ciencia, Tecnología y Computación  
Universidad de Cantabria (UC)

25/01/2024

17:30 - 19:30

Examen

Sixto Herrera García

Profesor Ayudante Doctor de Matemática Aplicada  
Universidad de Cantabria (UC)

# BIBLIOGRAFÍA Y ENLACES RELACIONADOS

## Bibliografía

### Bibliografía 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

### Bibliografía complementaria

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