



Universidad  
Rey Juan Carlos

ESCUELA TÉCNICA SUPERIOR  
DE INGENIERÍA INFORMÁTICA

## Guía de la asignatura

APRENDIZAJE AUTOMÁTICO I  
Grado en Ciencia e Ingeniería de Datos

2023-2024

**Autor: Carmen Lancho, Isaac Martín de Diego**



Copyright (c) 2023 Carmen Lancho, Isaac Martín de Diego. Esta obra está bajo la licencia CC BY-SA 4.0, [Creative Commons Atribución-Compartir Igual 4.0 Internacional](https://creativecommons.org/licenses/by-sa/4.0/).

## 1. Introducción de la asignatura

En el campo de la tecnología, una revolución está transformando la manera en que las máquinas interpretan y toman decisiones. A lo largo de este curso llevaremos a cabo un apasionante recorrido por el universo del aprendizaje automático (**ML**, de “Machine Learning” en inglés). El **aprendizaje automático** es una subdisciplina de la IA que se centra en el desarrollo de algoritmos y modelos que pueden aprender de datos y realizar tareas sin estar explícitamente programados. Esto es fundamental en multitud de aplicaciones, muchas de las cuales trataremos a lo largo de este curso. Este campo ha revolucionado la inteligencia artificial (**IA**, o AI, de “Artificial Intelligence” en inglés), dando forma al futuro de la tecnología y la ciencia de datos. Desde sus cimientos más elementales hasta algunas de las técnicas más avanzadas, exploraremos cómo las máquinas pueden aprender de los datos y mejorar su rendimiento en tareas que abarcan desde el reconocimiento de patrones hasta la toma de decisiones complejas. A lo largo de varios temas, desvelaremos los secretos que se esconden en los algoritmos que impulsan la IA moderna, permitiéndote aprender cómo aplicar estas técnicas para resolver desafíos del mundo real.

En esta asignatura emplazaremos el aprendizaje automático en el contexto de la ciencia e ingeniería de datos, remarcando la importancia que tiene en cualquier proyecto de ciencia de datos. Comenzaremos en los datos, destacando la importancia de su correcto tratamiento y comprensión (Análisis Exploratorio de Datos) para la posterior aplicación de técnicas de aprendizaje automático. Una vez establecida esta base, ahondaremos en métodos de aprendizaje automático como la reducción de la dimensión, el aprendizaje no supervisado y el aprendizaje supervisado.

**Objetivos de la asignatura:** Aprender a tratar correctamente distintos conjuntos de datos, a limpiarlos, procesarlos y entenderlos. Con ello, saber qué preguntas realizarse sobre los datos y saber qué técnicas o qué combinación de técnicas de aprendizaje automático aplicar para responderlas.

---

## 2. Temario de la asignatura

Tema 1: Introducción al Aprendizaje Automático

Tema 2: Datos

Tema 3: Análisis Exploratorio de Datos

Tema 4: Técnicas de reducción de la dimensionalidad

Tema 5: Aprendizaje no supervisado

Tema 6: Medidas de rendimiento

Tema 7: Aprendizaje supervisado

Tema 8: Reglas de asociación

Tema 9: Nuevas tendencias

## 3. Desarrollo de la asignatura en 15 semanas

Con 15 semanas de clase y 2 clases por semana (1 dedicada a teoría y 1 a práctica), se establece la siguiente configuración de la asignatura. Nótese que en los apuntes de esta asignatura se incluyen códigos que el alumno puede ir implementando tema por tema para realizar las prácticas correspondientes.

### Semana 1

La primera clase presencial siempre se dedica a presentar la asignatura y explicar la guía docente, el sistema de evaluación y la dinámica general de la asignatura. El tiempo restante de la primera clase y la segunda clase se da el tema 1 que es introductorio para poner la asignatura en contexto y se presenta el software con el que se va a trabajar. En este caso se ha optado por R pero podría escogerse Python sin ningún tipo de problemas.

Para el seguimiento autónomo de esta primera semana se recomienda leer la guía de la asignatura, el tema introductorio y recopilar la bibliografía recomendada.

## **Semana 2**

Esta semana se dará el tema 2 dedicado a los datos. Además, los alumnos deberán buscar conjuntos de datos (en los repositorios que se facilitarán durante dicho tema) para comenzar a trabajar y jugar con ellos. Esto se realizará en la sesión práctica.

## **Semanas 3 y 4**

Estas semanas se dedicarán al tema 3 de Análisis Exploratorio de Datos. Se irán explicando los contenidos de forma teórica y se irán implementando en las clases prácticas con los conjuntos de datos buscados la semana 2.

## **Semanas 5**

Esta semana se cubre el temario del tema 4: reducción de la dimensión. De nuevo, se realizará tanto de forma teórica como práctica en la sesión de prácticas.

## **Semanas 6-8**

Estas semanas se dedican al tema 5 de aprendizaje no supervisado que se divide en algoritmos no jerárquico, algoritmos jerárquicos y mapas auto-organizativos. En las sesiones prácticas se aplicarán dichas técnicas.

## **Semana 9**

Antes de comenzar el temario de aprendizaje supervisado, se presenta el tema 6 de medidas de rendimiento para evaluar el rendimiento y comportamiento de los modelos supervisados. También se comenzará la explicación del tema 7 de aprendizaje supervisado con los modelos lineales (nótese que los estudiantes de este grado tienen una asignatura de Modelos de regresión que ya cubre el modelo de regresión lineal).

## **Semanas 10-12**

Estas semanas se utilizarán para dar tanto la teoría como la práctica del tema 7 de aprendizaje supervisado

## **Semana 13**

La semana 13 se dedica para dar el tema 8 de reglas de asociación y hacer su práctica.

## **Semana 14-15**

La semana 14 se usa para el temario de nuevas tendencias (tema 9) y para aclarar dudas de las prácticas y de cara al examen.

---

## 4. Bibliografía

A continuación se recomiendan una serie de libros de referencia:

- Hastie, R. Tibshirani, J. Friedman; The Elements of Statistical Learning Theory: Data Mining, Inference and Prediction; Springer, 2001.
- G. James, D. Witten, T. Hastie, R. Tibshirani; An Introduction to Statistical Learning with Applications in R; Springer, 2013.
- Kelleher, J. D., & Tierney, B. (2018). Data science. MIT Press.
- Kelleher, J. D., Mac Namee, B., & D'arcy, A. (2020). Fundamentals of machine learning for predictive data analytics: algorithms, worked examples, and case studies. MIT press.
- Murphy, K. P. (2012). Machine learning: a probabilistic perspective. MIT press.
- "Fundamentos de ciencia de datos con R" coordinado por Gema Fernández-Avilés y José-María Montero. <https://cdr-book.github.io/>