

**FICHA IDENTIFICATIVA****Datos de la Asignatura**

<b>Código</b>	44276
<b>Nombre</b>	Análisis exploratorio de datos
<b>Ciclo</b>	Máster
<b>Créditos ECTS</b>	3.0
<b>Curso académico</b>	2023 - 2024

**Titulación(es)**

<b>Titulación</b>	<b>Centro</b>	<b>Curso</b>	<b>Periodo</b>
2199 - M.U. en Ingeniería Electrónica	Escuela Técnica Superior de Ingeniería	1	Segundo cuatrimestre
3131 - Ingeniería Electrónica	Escuela de Doctorado	0	Primer cuatrimestre

**Materias**

<b>Titulación</b>	<b>Materia</b>	<b>Caracter</b>
2199 - M.U. en Ingeniería Electrónica	1 - Tratamiento digital de señales	Obligatoria
3131 - Ingeniería Electrónica	1 - Complementos de Formación	Optativa

**Coordinación**

<b>Nombre</b>	<b>Departamento</b>
MARTIN GUERRERO, JOSE DAVID	242 - Ingeniería Electrónica
SERRANO LOPEZ, ANTONIO JOSE	242 - Ingeniería Electrónica

**RESUMEN**

Esta asignatura tiene como objetivo principal describir las técnicas más importantes de Análisis Exploratorio de Datos que permiten extraer conocimiento de un problema a través del estudio estadístico de los datos adquiridos.

La asignatura está formada por cinco temas. En el primero de ellos se presentan y repasan algunas de las herramientas matemáticas básicas para el análisis de datos, como la Probabilidad, la Estadística o el Álgebra Lineal. A continuación, en el segundo tema, se describen algunos conceptos básicos, que permiten definir las características del tipo de problema a resolver y la técnica que se utilizará para ello (número de patrones, datos atípicos, valores perdidos, tipo de aprendizaje a utilizar, etc.). Posteriormente, en el tercer tema, se presentan las técnicas para llevar a cabo las primeras exploraciones sobre los datos para tener una descripción somera del tipo de información que almacenan, destacando el estudio de algoritmos de agrupamiento (*clustering*) así como las técnicas de extracción y selección de variables



(características). Los dos últimos temas describen los modelos de clasificación y regresión (lineales y no lineales, respectivamente) más ampliamente utilizados en el Análisis Exploratorio de Datos.

Esta es una asignatura de carácter obligatorio, que se imparte en el primer cuatrimestre del Master en Ingeniería Electrónica. La carga lectiva total es de 3 ECTS. La carga de trabajo para el alumno es de 75 horas a lo largo del cuatrimestre, de las cuales 30 son presenciales y 45 son de trabajo individual.

Para las sesiones de laboratorio, se utilizará preferentemente Matlab, sin perjuicio de que los alumnos puedan utilizar otras soluciones, como Python o R.

## CONOCIMIENTOS PREVIOS

### Relación con otras asignaturas de la misma titulación

No se han especificado restricciones de matrícula con otras asignaturas del plan de estudios.

### Otros tipos de requisitos

Es recomendable que el alumno tenga una base matemática mínima de Álgebra, Cálculo, Estadística y Probabilidad. Si no la posee, se le facilitará una serie de tutoriales para que se adapte al curso sin problemas.

## COMPETENCIAS

### 2199 - M.U. en Ingeniería Electrónica

- Que los/las 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/las 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.
- Que los/las 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/las estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo
- Considerar el contexto económico y social en las soluciones de ingeniería siendo consciente de la diversidad y la multiculturalidad, y garantizando la sostenibilidad y el respeto a los derechos humanos y a la igualdad hombre-mujer.
- Diseñar un sistema, componente o proceso que cumpla unas especificaciones desde diferentes puntos de vista: electrónico, económico, social, ético y medioambiental.



- Demostrar una comprensión sistemática de un campo de estudio y el dominio de las habilidades.
- Realizar un análisis crítico, evaluación y síntesis de ideas nuevas y complejas.
- Ser capaz de fomentar, en contextos académicos y profesionales, el avance tecnológico, social o cultural dentro de una sociedad basada en el conocimiento.
- Capacidad para proyectar, calcular y diseñar productos, procesos e instalaciones en todos los ámbitos de la Ingeniería Electrónica y en particular los de tratamiento de la señal, sistemas digitales y de comunicaciones y electrónica industrial.
- Capacidad para el modelado matemático, cálculo y simulación en todos los ámbitos relacionados con la Ingeniería Electrónica y campos multidisciplinares afines. En especial los de tratamiento de la señal, sistemas digitales y de comunicaciones y electrónica industrial.
- Conocer las técnicas avanzadas de análisis de datos.
- Capacidad de analizar, especificar y diseñar sistemas de tratamiento digital de señales desde su concepción hasta su implementación en sistemas hardware de tiempo real..

## RESULTADOS DE APRENDIZAJE

Una vez se haya desarrollado esta asignatura el alumno deberá saber afrontar con garantías un problema de análisis de datos desde un principio, siendo capaz de filtrar la información útil de la que no lo es (ruido), y teniendo capacidad para encontrar los perfiles típicos de comportamiento que se observen en el conjunto de datos.

Asimismo, tendrá los suficientes elementos de juicio y la capacidad de establecer las ventajas e inconvenientes de diferentes alternativas en el procesado de datos. Podrá utilizar aproximaciones lineales o no lineales dependiendo del problema a resolver.

## DESCRIPCIÓN DE CONTENIDOS

### 1. Fundamentos básicos para el análisis exploratorio de datos

TEMA 1.

- 1.1 Probabilidad (axiomas; prob. condicional y conjunta)
- 1.2 Teorema de Bayes. Aplicaciones
- 1.3 Variable Aleatoria. Distribuciones típicas; momentos
- 1.4 Estadística. Contraste de hipótesis
- 1.5 Repaso de Álgebra lineal
- 1.6 Optimización de funciones

PRÁCTICA: Aplicación de estadística y contrastes de hipótesis para el análisis de datos.



## 2. Introducción al análisis de datos

TEMA 2.

- 2.1 Variables: tipos. Patrones
- 2.3 Tipos de problemas a resolver
- 2.4 Aprendizaje. Tipos

## 3. Análisis exploratorio descriptivo

TEMA 3.

- 3.1 Adquisición y limpieza de los datos
- 3.2 Caracterización de las variables (estadísticos y gráficas)
- 3.3 Transformaciones de los datos
- 3.4 Selección y extracción de características
- 3.5 Algoritmos de agrupamiento (HCM, FCM y SOM)

PRÁCTICA: Preprocesado de datos

Codificación, normalización, eliminación de datos atípicos, procesamiento de datos perdidos, obtención de perfiles típicos con métodos de agrupamiento.

## 4. Modelos lineales

TEMA 4.

- 4.1 Descripción de modelos lineales
- 4.2 Obtención de parámetros: ecuaciones normales. Regularización
- 4.3 Validación de modelos

PRÁCTICA: Aplicación de modelos lineales para el tratamiento de datos

Regresión lineal de mínimos cuadrados para el modelado de funciones. Regularización. Regresión logística aplicada a problemas de clasificación. Regresión con funciones de coste robustas.

## 5. Modelos no lineales

TEMA 5.

- 5.1 Árboles (regresión y clasificación)
- 5.2 Redes Neuronales Artificiales
- 5.3 Máquinas de Vectores Soporte
- 5.4 Extracción de reglas

PRÁCTICA: Aplicación de modelos no lineales al tratamiento de datos (I)

Aplicación de modelos no lineales vistos en teoría; se usarán conjuntos de datos similares a la práctica anterior para poder comparar su funcionamiento.

PRÁCTICA: Aplicación de modelos no lineales al tratamiento de datos (II)

Dado que se muestran varios modelos no lineales en teoría se plantea una segunda práctica para



analizar los que no se hayan visto en a práctica anterior.

## VOLUMEN DE TRABAJO

ACTIVIDAD	Horas	% Presencial
Clases de teoría	15,00	100
Prácticas en laboratorio	15,00	100
Estudio y trabajo autónomo	10,00	0
Lecturas de material complementario	5,00	0
Preparación de actividades de evaluación	5,00	0
Preparación de clases de teoría	5,00	0
Preparación de clases prácticas y de problemas	10,00	0
Resolución de casos prácticos	10,00	0
<b>TOTAL</b>	<b>75,00</b>	

## METODOLOGÍA DOCENTE

Las metodologías docentes a emplear en el desarrollo de la asignatura son las siguientes:

a) Actividades teóricas.

Desarrollo expositivo de la materia con la participación del estudiante en la resolución de cuestiones puntuales.

b) Actividades prácticas.

Resolución de casos prácticos

c) Trabajo personal del estudiante.

Descripción: Realización fuera del aula de cuestiones y problemas, así como la preparación de clases y exámenes (estudio). Esta tarea se realizará de manera individual e intenta potenciar el trabajo autónomo.

Se utilizarán las plataformas de e-learning (Aula Virtual) como soporte de comunicación con los estudiantes. A través de ella se tendrá acceso al material didáctico utilizado en clase, así como los problemas y ejercicios a resolver

## EVALUACIÓN

La evaluación de la asignatura se llevará a cabo mediante la realización de una o varias pruebas de conocimiento en formato de examen y/o trabajo individual y/o trabajo por grupos.



## REFERENCIAS

### Básicas

- Advances in knowledge discovery and data mining. Edited by Usama M. Fayyad [et al.]. MIT Press, 1996.
- Data mining for scientific and engineering applications. Edited by Robert L. Grossman [et al.]. Kluwer, 2001.
- Análisis de datos experimentales. Emilio Soria, José D. Martín, Antonio J. Serrano, Daniel Aguado. Universidad Politécnica de Valencia, 2007.
- Machine Learning. Ethem Alpaydin, MIT Press, 2009.

### Complementarias

- Neural Networks and Learning Machines. Simon Haykin. Pearson Education, 2009.
- Time Series Prediction: Forecasting the Future and Understanding the Past: Proceedings of the NATO by Andreas Weigend y Neil Gershenfeld (Editores). Addison-Wesley, 1993.
- Pattern Classification (2nd Edition). Richard O. Duda, Peter E. d G. Stork. Wiley-Interscience, 2n edition, November 2000.
- Kernel Methods for Pattern Analysis. John Shawe-Taylor and Nello Cristianini. Cambridge University Press, New York, NY, USA, 2004.
- The elements of statistical learning: data mining, inference, and prediction. Trevor Hastie, Robert Tibshirani, Jerome Friedman. Springer, 2001.