



FICHA IDENTIFICATIVA

DATOS DE LA ASIGNATURA

Código: 46573
Nombre: Análisis exploratorio de datos
Ciclo: Máster Universitario Oficial
Créditos ECTS: 4,5
Curso académico: 2025-26

TITULACIONES

Titulación	Centro	Curso	Periodo
2262 - Máster Universitario en Ciencia de Datos	Escola Tècnica Superior d'Enginyeria	1	Primer cuatrimestre

MATERIAS

Titulación	Materia	Carácter
2262 - Máster Universitario en Ciencia de Datos	Análisis exploratorio de datos	OBLIGATORIA

COORDINACIÓN

MARTINEZ SOBER MARCELINO

GOMEZ SANCHIS JUAN

RESUMEN

En esta asignatura se describen las primeras etapas de un problema de análisis de datos.

El o la científica de datos se enfrenta a un conjunto de datos de muy diversa procedencia, formato, organización, codificación, etc. La correcta adquisición, organización, eliminación de posibles datos erróneos (*outliers*), imputación de datos faltantes (*missing data imputation*), transformación de los datos, reducción de la dimensionalidad (selección de las características más relevante de un conjunto de alta dimensionalidad (*feature selection*) y extracción de características (*feature extraction*)), eliminación de datos redundantes, etc. es una de las etapas más costosas de un problema de análisis de datos. Esta etapa es crucial para el correcto tratamiento del problema y la fiabilidad y solidez de los resultados obtenidos en etapas posteriores de análisis (selección de modelos, clasificadores, agrupamiento, estimación, contrastes de hipótesis, etc). En este módulo nos centraremos en las etapas de preparación de los datos

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

Introducción a la Ciencia de Datos

COMPETENCIAS / RESULTADOS DE APRENDIZAJE

-

Capacidad de acceso y gestión de la información en diferentes formatos para su posterior análisis con el fin de obtener conocimiento a partir de datos.

Capacidad de análisis y síntesis, en la elaboración de informes, en la exposición, comunicación y defensa de ideas.

Capacidad de organización y planificación de actividades de investigación, desarrollo y consultoría en el área de ciencia de datos.

Entender la utilidad de la ciencia de datos y sus elementos asociados, así como su aplicación en la resolución de problemas, eligiendo las técnicas más adecuadas a cada problema, aplicando de forma correcta las técnicas de evaluación y, finalmente, interpretando los modelos y resultados.

Extraer conocimiento de conjuntos de datos en diferentes formatos.

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

Ser capaces de acceder a herramientas de información (bibliográficas y de empleo) y utilizarlas apropiadamente.

Ser capaces de asumir la responsabilidad de su propio desarrollo profesional y de su especialización en uno o más campos de estudio, aplicando los conocimientos adquiridos en la identificación de salidas profesionales y yacimientos de empleo.

Ser capaces de valorar la necesidad de completar su formación técnica, científica, en lenguas, en informática, en literatura, en ética, social y humana en general, y de organizar su propio autoaprendizaje con un alto grado de autonomía

DESCRIPCIÓN DE CONTENIDOS

1. Introducción al análisis exploratorio de datos

En este bloque de introducción se presentarán los principales aspectos a tener en cuenta para realizar una correcta preparación de los datos y sus implicaciones en las siguientes etapas.



2. Adquisición y limpieza de datos

En este bloque se presentarán los diversos tipos de datos, (continuos, discreto) , importación de los formatos más habituales, conversión de datos, detección de datos anómalos.

3. Análisis estadístico descriptivo

En este bloque se muestran como caracterizar diferentes tipos de datos, mediante sus estadísticos básico y diversos tipos de representaciones visuales básicas, como parte fundamental en la comprensión de los datos disponibles y la detección de errores en importación o en los valores originales de los mismos (análisis univariante, bivariante, multivariante, correlación, covarianza, etc)

4. Análisis Exploratorio de datos II. Anomalías.

En este bloque se presentan diferentes tipos de anomalías presentes en los datos numéricos como son los outliers, métodos de detección, así como la problemática de los valores perdidos, sus tipos y métodos de imputación.

5. Transformaciones en los datos

En este bloque se presentan métodos de transformación de datos. En esta etapa los datos son transformados para que el proceso de análisis sea más eficiente y se facilite la comprensión de la información que contienen.

6. Introducción a la reducción de la dimensionalidad

En este bloque se presentan una introducción a técnicas de reducción de la dimensionalidad. En particular se presentarán los paradigmas de selección de características y extracción de características. En particular se presentarán métodos de selección de características filter sencillos y métodos más complejos wrapper. Del mismo modo se presentará el paradigma de extracción de características, en particular el análisis de componentes principales (PCA).

VOLUMEN DE TRABAJO (HORAS)

ACTIVIDADES PRESENCIALES

Actividad	Horas
Teoría-Prácticas	4,00
Teoría	25,00
Laboratorio	16,00
Total horas	45,00

**ACTIVIDADES NO PRESENCIALES**

Actividad	Horas
Asistencia a otras actividades	0,00
Elaboración de trabajos individuales o en grupo	10,00
Estudio y trabajo autónomo	7,50
Preparación de clases	16,50
Preparación de actividades de evaluación	6,00
Resolución de casos prácticos	5,00
Total horas	45,00

METODOLOGÍA DOCENTE

Las clases combinarán el contenido teórico y el práctico, sin distinción entre sesiones de teoría y de laboratorio. Todas las sesiones se impartirán en aulas de informática.

Actividades teóricas. Desarrollo expositivo de la materia con la participación del estudiante en la resolución de cuestiones puntuales. Posible realización de cuestionarios individuales de evaluación.

Actividades prácticas. Aprendizaje mediante resolución de problemas, ejercicios y casos de estudio que permiten adquirir competencias sobre los diferentes aspectos de la materia.

Trabajos en laboratorio y/o aula ordenador. Aprendizaje mediante la realización de actividades desarrolladas de forma individual o en grupos reducidos y llevadas a cabo en aulas de ordenador.

EVALUACIÓN

La evaluación del aprendizaje de los conocimientos y competencias conseguidas por los estudiantes se hará de forma continuada a lo largo del curso, y constará de los siguientes bloques de evaluación:

1. Evaluación de las actividades prácticas a partir de la elaboración de trabajos/memorias y/o exposiciones orales: 60% de la nota final.
2. Prueba objetiva, consistente en uno o varios exámenes que constarán tanto de cuestiones teóricoprácticas como de problemas: 40 % de la nota final.

Las calificaciones obtenidas en el apartado 1 sólo se conservarán en las dos convocatorias del curso académico en que hayan sido realizadas, dado que su evaluación sólo es posible en el periodo de docencia.

BIBLIOGRAFÍA



- K.Pearson (2018) Exploratory Data Analysis Using R. CRC.
- H. Wickham, G. Grolemund. (2016) R for data Science. O'Reilly Media Inc. <http://r4ds.had.co.nz/>
- Max Kuhn, Kjell Johnson (2021) Feature Engineering and Selection A Practical Approach for Predictive Models-CRC Press
- GB. S. Baumer, D. T. Kaplan, N. J. Horton (2017) Modern Data Science with R. Boca Raton : Taylor & Francis CRC Press. (disponible e-libro)
- R. Buttres y, L.R. Whitaker (2018). A data scientist's guide to acquiring, cleaning and managing data in R . Wiley. (disponible e-libro)
- W. Graham, (2017). The Essentials of Data Science: Knowledge Discovery Using R. Chapman and Hall/CRC. (disponible e-libro)
- R. D. Peng (2016) Exploratory Data Analysis with R. Lean Publishing (<https://leanpub.com/exdata>)
- Max Kuhn, Julia Silge (2022), Tidy Modeling with R, O'Reilly Media, Inc. <https://www.tmrw.org/dimensionality.html>
- Alice Zheng, Amanda Casari (2018)- Feature Engineering for Machine Learning_ Principles and Techniques for Data Scientists-O'Reilly Media