

**FITXA IDENTIFICATIVA****DADES DE L'ASSIGNATURA**

**Codi:** 46573  
**Nom:** Anàlisi exploratòria de dades  
**Cicle:** Màster Universitari Oficial  
**Crèdits ECTS:** 4,5  
**Curs acadèmic:** 2025-26

**TITULACIONS**

Titulació	Centre	Curs	Període
2262 - Màster Universitari en Ciència de Dades	Escola Tècnica Superior d'Enginyeria	1	Primer quadrimestre

**MATÈRIES**

Titulació	Matèria	Caràcter
2262 - Màster Universitari en Ciència de Dades	Anàlisi exploratòria de dades	OBLIGATÒRIA

**COORDINACIÓ**

MARTINEZ SOBER MARCELINO

GOMEZ SANCHIS JUAN

**RESUM**

En aquesta assignatura es descriuen les primeres etapes d'un problema d'anàlisi de dades.

El o la científica de dades s'enfronta a conjunts de dades de procedència, format, organització, codificació, etc., molt diversos. L'adquisició correcta, l'organització, l'eliminació de possibles dades errònies (valors atípics o *outliers*), la imputació de valors absents (*missing data imputation*), la transformació de les dades, la reducció de la dimensionalitat (mitjançant la selecció de les característiques més rellevants *feature selection* i l'extracció de característiques (*feature extraction*), així com l'eliminació de dades redundants, constitueixen una de les etapes més costoses d'un problema d'anàlisi de dades. Aquesta etapa és clau per al tractament correcte del problema i per a garantir la fiabilitat i solidesa dels resultats obtinguts en etapes posteriors d'anàlisi (selecció de models, classificació, agrupament, estimació, contrastos d'hipòtesis, etc.). En aquest mòdul ens centrarem en les etapes de preparació de les dades

**CONEIXEMENTS PREVIS****RELACIÓ AMB ALTRES ASSIGNATURES DE LA MATEIXA TITULACIÓ**



No s'ha especificat restriccions de matrícula amb altres assignatures del pla d'estudis.

## ALTRES TIPUS DE REQUISITS

Introducció a la Ciència de Dades

## COMPETÈNCIES / RESULTATS D' APRENTATGE

### 2262 - Màster Universitari en Ciència de Dades

Capacitat d'accés i gestió de la informació en diferents formats per al seu posterior anàlisi a fi d'obtenir coneixement a partir de dades.

Capacitat d'anàlisi i síntesi, en l'elaboració d'informes, en l'exposició, comunicació i defensa d'idees.

Capacitat d'organització i planificació d'activitats d'investigació, desenrotllament i consultoria en l'àrea de ciència de dades.

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.

Extraure coneixement de conjunts de dades en diferents formats.

Que els estudiants posseïsquen les habilitats d'aprenentatge que els permeten continuar estudiant d'una forma que haurà de ser en gran manera autòdrida o autònoma.

Ser capaços d'accedir a ferramentes d'informació (bibliogràfiques i d'ocupació) i utilitzar-les apropiadament.

Ser capaços d'assumir la responsabilitat del seu propi desenvolupament professional i de la seua especialització en un o més camps d'estudi, aplicant els coneixements adquirits en la identificació d'eixides professionals i jaciments d'ocupació.

Ser capaços de valorar la necessitat de completar la seua formació tècnica, científica, en llengües, en informàtica, en literatura, en ètica, social i humana en general, i d'organitzar el seu propi autoaprenentatge amb un alt grau d'autonomia

## DESCRIPCIÓ DE CONTINGUTS

### 1. Introducció a l'anàlisi exploratori de dades

En aquest bloc d'introducció es presentaran els principals aspectes a tenir en compte per a realitzar una correcta visualització de les dades.



## 2. Adquisició i neteja de dades

En aquest bloc es presentaran els diversos tipus de dades, (continus, discrets) , importació dels formats més habituals, conversió de dades, detecció de dades anòmales.

## 3. Anàlisi estadístic descriptiu

En aquest bloc es mostren com caracteritzar diferents tipus de dades, mitjançant els seus estadístics bàsics i diversos tipus de representacions gràfiques com a part fonamental en la comprensió de les dades disponibles i la detecció d'errors en importació o en els valors originals dels mateixos (anàlisi univariant , bivariant , multivariant , correlació , covariància , etc. )

## 4. Anàlisi Exploratori de dades II. Anomalies.

En aquest bloc es presenten diferents tipus d'anomalies en variables numèriques. Outliers. Caracterització d'outliers. Mètodes de detecció. dades perdudes i absents, mètodes d'imputació

## 5. Transformacions en les dades

En aquest bloc es presenten mètodes de transformació de dades. En aquesta etapa les dades són transformats perquè el procés d'anàlisi sigui més eficient i es faciliti la comprensió de la informació que contenen.

## 6. Introducció a la reducció de la dimensionalitat

En aquest bloc es presenten una introducció a tècniques de reducció de la dimensionalitat. En particular, es presentaran els paradigmes de selecció de característiques i extracció de característiques. En particular es presentaran mètodes de selecció de característiques filter senzills i mètodes més complexos wrapper. De la mateixa manera, es presentarà el paradigma d'extracció de característiques, en particular l'anàlisi de components principals (PCA).

### VOLUM DE TREBALL (HORES)

#### ACTIVITATS PRESENCIALS

Activitat	Hores
Teoria-Pràctiques	4,00
Teoria	25,00
Laboratori	16,00
<b>Total hores</b>	<b>45,00</b>

#### ACTIVITATS NO PRESENCIALS



Activitat	Hores
Assistència a altres activitats	0,00
Elaboració de treballs individuals o en grup	10,00
Estudi i treball autònom	7,50
Preparació de classes	16,50
Preparació d'activitats d'avaluació	6,00
Resolució de casos pràctics	5,00
<b>Total hores</b>	<b>45,00</b>

## METODOLOGIA DOCENT

Les classes combinaran el contingut teòric i el pràctic, sense distinció entre sessions de teoria i de laboratori. Totes les sessions s'impartiran en aules d'informàtica.

**Activitats teòriques.** Desenvolupament expositiu de la matèria amb la participació de l'estudiant en la resolució de qüestions puntuals. Possible realització de qüestionaris individuals d'avaluació.

**Activitats pràctiques.** Aprenentatge mitjançant resolució de problemes, exercicis i casos d'estudi que permeten adquirir competències sobre els diferents aspectes de la matèria.

**Treballs en laboratori i/o aula ordinador.** Aprenentatge mitjançant la realització d'activitats desenvolupades de forma individual o en grups reduïts i dutes a terme en aules d'ordinador

## AVALUACIÓ

L'avaluació de l'aprenentatge dels coneixements i competències obtinguts pels estudiants es realitzarà de forma continuada al llarg del curs, i constarà dels següents blocs d'avaluació:

1. Exercicis i treballs entregats durant el curs i/o exàmens parcials: 60% de la nota final
2. Examen final: 40% de la nota final.

Les qualificacions obtingudes en l'apartat 1 sols es conservaran en les dues convocatòries del curs acadèmic en què s'hagen realitzat, donat que la seua avaluació sols és possible en el període de docència

## BIBLIOGRAFIA

- 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.tmw.org/dimensionality.html>
- Alice Zheng, Amanda Casari (2018)- Feature Engineering for Machine Learning\_ Principles and Techniques for Data Scientists-O'Reilly Media