

**FITXA IDENTIFICATIVA****DADES DE L'ASSIGNATURA****Codi:** 36423**Nom:** Tractament de les dades**Cicle:** Grau**Crèdits ECTS:** 6**Curs acadèmic:** 2025-26**TITULACIONS**

Titulació	Centre	Curs	Període
1406 - Grau en Ciència de Dades	Escola Tècnica Superior d'Enginyeria	1	Segon quadrimestre

MATÈRIES

Titulació	Matèria	Caràcter
1406 - Grau en Ciència de Dades	Gestió de la informació	OBLIGATÒRIA

COORDINACIÓ

MARTINEZ SOBER MARCELINO

GOMEZ CHOVA LUIS

RESUM

El científic o la científica de dades s'enfronta a un conjunt de dades de molt diversa procedència, format, organització, codificació, etc. La correcta adquisició, organització, eliminació de possibles dades errònies (outliers), imputació de dades mancants (missing data imputation), transformació de les dades, selecció de les característiques més rellevants d'un conjunt d'alta dimensionalitat (feature selection), eliminació de dades redundants, etc. és una de les etapes més costoses d'un problema d'anàlisi de dades. Aquesta etapa és crucial per al correcte tractament del problema i la fiabilitat i solidesa dels resultats obtinguts en etapes posteriors d'anàlisi (selecció de models, classificadors, agrupament, estimació, contrastos d'hipòtesis, etc.). Totes aquestes tasques seran abordades en l'assignatura obligatòria, 36423 Tractament de Dades, que s'imparteix en el segon quadrimestre de primer curs.

Les classes de teoria s'impartiran en castellà i les classes pràctiques i de laboratori segons consta en la fitxa de l'assignatura disponible en la web del grau.

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

Es recomana haver superat l'assignatura Dades, Societat i Universitat que s'imparteix en el primer quadrimestre del primer curs del grau.

COMPETÈNCIES / RESULTATS D' APRENTATGE

1406 - Grau en Ciència de Dades

(CB3) Que els estudiants tinguen la capacitat d'arreglar i interpretar dades rellevants (normalment dins de la seua àrea d'estudi) per emetre judicis que incloguen una reflexió sobre temes rellevants d'índole social, científica o ètica.

(CE02) Conèixer i aplicar de forma metodològica les tècniques de programació i l'algorísmia necessàries per al processament eficient d'informació i la resolució informàtica de problemes que utilitzen grans volums de dades.

(CE06) Capacitat per representar i visualitzar conjunts de dades per a l'extracció de coneixement.

(CE11) Capacitat per dissenyar i implementar la presa de dades, la seua integració, transformació, selecció, comprovació de la seua qualitat i veracitat a partir de diferents fonts, tenint en compte el seu caràcter, heterogeneïtat i variabilitat.

(CE13) Saber dissenyar, aplicar i avaluar algorismes de Ciència de Dades per a la resolució de problemes complexos.

(CG06) Capacitat d'accés i gestió de la informació en diferents formats per a la seua posterior anàlisi amb la finalitat d'obtenir coneixement a partir de dades.

(CT03) Habilitat per defensar el seu treball amb rigor i arguments, exposant-ho de forma adequada i precisa, recolzant-se en els mitjans necessaris.

(CT04) Ser responsables del seu propi desenvolupament professional i de la seua especialització, aplicant els coneixements adquirits en la identificació de sortides professionals i jaciments d'ocupació.

DESCRIPCIÓ DE CONTINGUTS

1. Introducció al tractament de dades

1.1. Per què analitzar dades.

1.2. Visió global d'un problema de tractament de dades.



2. Obtenció de dades

- 2.1. Introducció.
- 2.2. Introducció de dades.
- 2.3. Repositoris de dades.
- 2.4. Formats de fitxers de dades.
- 2.5. Fusió de dades procedents de diferents fonts.
- 2.6. Accés a bases de dades.

3. Visualització de dades.

- 3.1. Gràfics explicatius i exploratoris.
- 3.2. Sistemes gràfics en R: base, grid, lattice, ggplot2.
- 3.3. Llibreria ggplot2: representacions bàsiques.

4. Preparació de les dades

- 4.1. Estructura d'un conjunt de dades per a la seva anàlisi: operacions bàsiques.
- 4.2. Manipulació de dades. Llibreria tidy.
- 4.3. Maneig de dades. Llibreria dplyr.

5. Anàlisi exploratòria de dades I. Definicions

- 5.1. Explorant un nou dataset.
- 5.2. Caracterització de variables.
- 5.3. Visualització de relacions entre variables.

6. Anàlisi Exploratori de dades II. Anomalies.

- 6.1. Anomalies en variables numèriques: Outliers. Caracterització d'outliers. Mètodes de detecció.
- 6.2. Anomalies en variables numèriques: dades perdudes i absents

7. Treballant amb dades de tipus text

- 7.1. Introducció
- 7.2. Bases de l'anàlisi de dades de text.
- 7.3. Funcions bàsiques per al maneig de caràcters en R.
- 7.4. Expressions regulars.
- 7.5 Codificació de caràcters: ascii vs Unicode.



8. Pràctiques de Tractament de les dades

En aquest bloc es presentaran una sèrie de supòsits pràctics a manera de pràctiques de laboratori.

Pràctica 1. Importació de dades.

Pràctica 2. Visualització de dades.

Pràctica 3. Preparació de dades amb tidyR.

Pràctica 4. Maneig de dades amb dplyr.

Pràctica 5. Dades anòmals.

Pràctica 6. Anàlisi exploratòria de dades.

Pràctica 7. Anàlisi completa d'un conjunt de dades.

VOLUM DE TREBALL (HORES)

ACTIVITATS PRESENCIALS

Activitat	Hores
Teoria	34,00
Pràctiques a l'aula	6,00
Laboratori	20,00
Total hores	60,00

ACTIVITATS NO PRESENCIALS

Activitat	Hores
Assistència a altres activitats	0,00
Elaboració de treballs individuals o en grup	15,00
Estudi i treball autònom	20,00
Preparació de classes	30,00
Preparació d'activitats d'avaluació	25,00
Resolució de casos pràctics	0,00
Total hores	90,00

METODOLOGIA DOCENT

Les classes combinaran el contingut teòric i pràctic.

MD1 - Activitats teòriques. Desenvolupament expositiu de la matèria amb la participació de l'estudiantat en la resolució de qüestions puntuals. Realització de qüestionaris individuals d'avaluació.

En les activitats teòriques de caràcter presencial es desenvoluparan els temes de l'assignatura proporcionant una visió global i integradora, analitzant amb major detall els aspectes clau i de major complexitat, fomentant, en tot moment, la participació de l'alumnat (CB03, CT03).

MD2 - Activitats pràctiques. Aprenentatge mitjançant resolució de problemes, exercicis i casos d'estudi a través dels quals s'adquireixen competències sobre els diferents aspectes de la matèria. (CB03, CG06,



CE02, CE06, CE11, CE13)

Les activitats teòriques es complementen amb activitats pràctiques amb l'objectiu d'aplicar els conceptes bàsics i ampliar-los amb el coneixement i l'experiència que es vagin adquirint durant la realització dels treballs proposats.

MD4 -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 laboratoris i / o aules d'ordinador. (CB03, CG06, CE02, CE06, CE11, CE13)

A més de les activitats presencials, els estudiants hauran de realitzar tasques personals (fora de l'aula) sobre: qüestions i problemes, així com la preparació de classes i exàmens (estudi). Aquestes tasques es realitzaran principalment de manera individual, per tal de potenciar el treball autònom, però addicionalment s'inclouran treballs, especialment la preparació i resolució de pràctiques laboratori, que requereixin la participació de l'estudiants en petits grups (2-3) per fomentar la capacitat d'integració en grups de treball.

S'utilitzarà les plataformes d'e-learning (Aula Virtual) de la Universitat de València, Microsoft Teams i Blackboard Collaborate, com a suport de comunicació amb l'alumnat. A través d'elles es tindrà accés al material didàctic utilitzat en classe, així com els problemes i exercicis a resoldre.

AVALUACIÓ

L'avaluació de l'aprenentatge dels coneixements i competències obtinguts per l'estudiantat es farà de manera contínua al llarg del curs, i comptarà dels següents blocs d'avaluació:

Primera i segona convocatòries:

SE1 - Prova objectiva, consistent en un o diversos exàmens que consten tant de qüestions teòriques-pràctiques com de problemes (avaluació de competències CB03, CG06, CT03, CE02, CE06, CE11, CE13) (53%) (Nota: Tots els percentatges estan referits a la nota final)

SE1-1 (35%) Examen de teoria

SE1-2 (18%) Examen de laboratori

SE2 - Avaluació de les pràctiques a partir de l'elaboració de treballs / memòries i / o exposicions orals (avaluació de competències CB03, CG06, CT03, CT04, CE02, CE06, CE11, CE13) (27%).

SE2-1 (15%) Realització d'un miniproyecto consistent en l'anàlisi completa d'un conjunt de dades.



SE2-2 (12%) Sessions Laboratori (Activitat NO RECUPERABLE)

SE3 - Avaluació contínua de l'estudiantat, basada en la participació i grau d'implicació de l'alumnat en el procés d'ensenyament-aprenentatge, tenint en compte la regular assistència a les activitats presencials previstes i la resolució de qüestions i problemes proposats periòdicament. (20%)

SE3-1 (5%) Assistència regular a les activitats presencials previstes (avaluació de competències CB04, CG01). (Activitat NO RECUPERABLE)

SE3-2 (15%) Resolució de qüestions i problemes proposats (avaluació de competències CB02, CB04, CG01, CT03). (Activitat NO RECUPERABLE)

La nota final de l'assignatura es calcularà com la mitjana ponderada de cada un dels apartats anteriors, d'acord amb el següent criteri: SE-1 (53%), SE-2 (27%), SE-3 (20%).

Consideracions particulars sobre l'avaluació:

- És necessari obtenir una qualificació mínima de 5 (sobre 10) en els apartats d'avaluació SE1-1 (examen Teoria), SE1-2 (examen de laboratori) i SE2-1 (mini projecte), per fer la mitjana.

- Les activitats SE2-2, SE3-1 i SE3-2 no són recuperables.

- L'activitat SE1-2 es realitzarà a l'acabar l'examen de teoria el dia de la convocatòria oficial.

En qualsevol cas, el sistema d'avaluació es regirà pel que estableix el Reglament d'Avaluació i Qualificació de la Universitat de València per a Grans i Mestres:

<https://webges.uv.es/uvTaeWeb/MuestraInformacionEdictoPublicoFrontAction.do?accion=inicio&idEdictoSeleccionado=5639>

La còpia o plagi manifest de qualsevol activitat que forma part de l'avaluació suposarà la impossibilitat de superar l'assignatura, sotmetent-se seguidament als procediments disciplinaris oportuns indicats en el **PROTOCOL D'ACTUACIÓ DAVANT PRÀCTIQUES FRAUDULENTES A LA UNIVERSITAT DE VALÈNCIA (ACGUV 123/2020)**.

BIBLIOGRAFIA

- R.K.Pearson (2018) Exploratory Data Analysis Using R. CRC.
- H. Wickham, Mine Çetinkaya-Rundel, G. Grolemund. (2023) R for data Science 2nd Edition. O'Reilly Media Inc. <http://r4ds.had.nz/>
- B. S. Baumer, D. T. Kaplan, N. J. Horton (2021) Modern Data Science with R. 2nd Edition. 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)
- L. Han, M. Kamber, and J. Pei. (2012) Data Mining Concepts and Techniques (third Edition). (disponible e-libro)
- N. Zumel and J. Mount (2014). Practical Data Science with R. Manning Publications Co.
- A.Cirillo (2017) R Data Mining. Pack Publishing (disponible e-libro)
- C.Aggarwal (2015) Data mining: the textbook. Springer (disponible e-libro)