



centro adscrito a:



UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH

GUIA DOCENT DE Mineria de Dades 2026-27

DADES GENERALS

Número:	Mineria de Dades
Codi:	801363 (BUSINESSTECH)
Curs:	2026-27
Titulació:	Grau en Empresa, Innovació i Tecnologia
Nº de crèdits (ECTS):	6
Ubicació en el pla d'estudis:	3r. Curs, 2n. quadrimestre
Departament:	Mètodes Quantitatius
Responsable departament:	Dr. Iván Romero
Data de l'última revisió:	Juliol 2026
Professor Responsable:	Xavi Royo Melero

1. DESCRIPCIÓ GENERAL

La mineria de dades constitueix una de les disciplines centrals de l'analítica de negoci, en integrar tècniques estadístiques, computacionals i d'aprenentatge automàtic per extreure coneixement accionable a partir de grans volums de dades. En un entorn empresarial on la generació de dades creix de forma exponencial, la capacitat de descobrir patrons, predir comportaments i segmentar col·lectius resulta indispensable per a la presa de decisions estratègiques.

L'assignatura s'estructura en sis blocs temàtics que recorren el cicle complet d'un projecte de mineria de dades: des de la comprensió del problema i els fonaments del machine learning, passant pel preprocessament i l'exploració de les dades (EDA), fins a la construcció i avaluació de models predictius i descriptius, així com el descobriment de patrons mitjançant tècniques de segmentació, l'associació i recomanació. Al llarg del curs s'emprarà el llenguatge Python com a eina principal, amb llibreries com pandas, scikit-learn, matplotlib i seaborn.

El primer bloc proporciona el marc conceptual de la mineria de dades i el machine learning, diferenciant els tipus d'aprenentatge supervisat i no supervisat, i situant la disciplina dins l'ecosistema més ampli de la ciència de dades.

El segon bloc aborda el preprocessament i l'exploració de dades, fase crítica que determina la qualitat i rellevància de qualsevol model posterior. S'estudien tècniques de neteja, transformació, tractament de valors absents i visualització exploratòria.

Els blocs tercer i quart se centren en els models supervisats: classificació i regressió, respectivament. Es presenten els algorismes més rellevants (regressió logística i lineal, arbres de decisió i random forest), mètriques d'avaluació i criteris de selecció segons el problema de negoci.

El cinquè bloc introdueix les tècniques d'aprenentatge no supervisat, amb especial èmfasi en el clustering (K-Means i clustering jeràrquic) i la segmentació de clients, eina clau per a l'anàlisi de negoci.

El sisè bloc aborda tècniques de descobriment de patrons i personalització basades en regles d'associació i sistemes de recomanació. S'estudia l'algoritme Apriori per a l'anàlisi de relacions entre productes i els enfocaments de collaborative filtering i content-based filtering per a la generació de recomanacions, examinant la seva utilitat en diferents contextos de negoci.

2. OBJECTIUS

En finalitzar el curs l'estudiant serà capaç de:

- Comprendre els fonaments conceptuals del machine learning i la mineria de dades, diferenciant tipus d'aprenentatge i paradigmes algorítmics.
- Aplicar tècniques de preprocessament i exploració de dades per garantir la qualitat i idoneïtat dels datasets abans del modelatge.
- Construir, entrenar i avaluar models de classificació i regressió utilitzant Python i la llibreria scikit-learn.
- Aplicar algorismes de clustering i segmentació per identificar grups naturals en les dades amb orientació a l'anàlisi de negoci.
- Aplicar regles d'associació i sistemes de recomanació per descobrir patrons de comportament i donar suport a decisions de personalització, recomanació i cross-selling.
- Seleccionar l'algoritme i les mètriques d'avaluació adequats en funció del problema de negoci i les característiques de les dades.
- Comunicar de forma efectiva els resultats d'una anàlisi de mineria de dades, tant de forma oral com escrita, a audiències tècniques i no tècniques.
- Desenvolupar una visió crítica sobre les limitacions, biaixos i consideracions ètiques associades a l'aplicació de models predictius en entorns empresarials.

3. RESULTATS DE L'APRENTATGE

- Identificar i diferenciar els principals tipus de problemes de mineria de dades (classificació, regressió, clustering i associació) i els paradigmes de machine learning associats.
- Aplicar un procés estructurat de preprocessament de dades (neteja, codificació, escalat i partició de dades) previ a la construcció de models.
- Construir pipelines de modelatge supervisat, avaluar el seu rendiment mitjançant mètriques apropiades i aplicar tècniques de validació creuada i ajust d'hiperparàmetres.
- Implementar tècniques de segmentació i clustering, interpretar els grups resultants i traslladar les troballes a recomanacions de negoci.
- Aplicar regles d'associació i sistemes de recomanació bàsics, interpretant els seus resultats en contextos de negoci.
- Reflexionar críticament sobre els resultats dels models, identificant possibles biaixos, limitacions i requisits ètics en la seva aplicació empresarial.

4. CONTINGUTS

TEMA 1. INTRODUCCIÓ A LA MINERIA DE DADES I MACHINE LEARNING

Resultats específics de l'aprenentatge

L'estudiant després d'estudiar el tema i realitzar els exercicis, serà capaç de:

- Definir els conceptes clau de mineria de dades, machine learning i intel·ligència artificial, situar-los en l'ecosistema de la ciència de dades.
- Distingir entre aprenentatge supervisat i no supervisat, identificant casos d'ús empresarial per a cada paradigma.
- Descriure les fases d'un projecte de mineria de dades (CRISP-DM) i relacionar-les amb el cicle de vida d'un model en producció.
- Identificar les principals llibreries de Python per a mineria de dades (pandes, scikit-learn, matplotlib, seaborn) i el seu rol en el flux de treball.

Contingut

- 1.1. Conceptes fonamentals: dades, informació, coneixement, mineria de dades i IA.
- 1.2. Tipus d'aprenentatge: supervisat i no supervisat.
- 1.3. Taxonomia de problemes: classificació, regressió, clustering i associació.
- 1.4. Metodologia CRISP-DM i cicle de vida d'un projecte de dades.
- 1.5. Desplegament de models com a fase final del cicle de vida: concepte i visió general.

- 1.6. Ecosistema Python per a mineria de dades: pandas, scikit-learn, matplotlib i seaborn.

TEMA 2. PREPROCESSAMENT I EXPLORACIÓ DE DADES (EDA)

Resultats específics de l'aprenentatge

L'estudiant després d'estudiar el tema i realitzar els exercicis, serà capaç de:

- Realitzar una anàlisi exploratòria de dades (EDA) completa, identificant la distribució, relacions i anomalies presents en un dataset.
- Aplicar tècniques de neteja de dades: detecció i tractament de valors absents, duplicats i outliers.
- Transformar variables mitjançant codificació (one-hot, label encoding) i escalat (normalització, estandardització).
- Utilitzar visualitzacions estadístiques (histogrames, boxplots, heatmaps de correlació) per comunicar troballes de l'EDA.

Contingut

- 2.1. Anàlisi exploratòria de dades (EDA): estadístiques descriptives i visualització.
- 2.2. Detecció i tractament de valors absents i duplicats.
- 2.3. Identificació i gestió d'outliers.
- 2.4. Codificació de variables categòriques: one-hot encoding i label encoding.
- 2.5. Escalat i normalització de variables: MinMaxScaler i StandardScaler.
- 2.6. Divisió de dades: train/test split i prevenció de data leakage
- 2.7. Pipelins de preprocessament amb scikit-learn: disseny i aplicació

TEMA 3. MODELS DE CLASSIFICACIÓ

Resultats específics de l'aprenentatge

L'estudiant després d'estudiar el tema i realitzar els exercicis, serà capaç de:

- Comprendre el funcionament dels principals algoritmes de classificació: regressió logística, arbres de decisió i random forest.
- Entrenar i avaluar models de classificació utilitzant scikit-learn, aplicant validació creuada.
- Interpretar mètriques d'avaluació de classificació: accuracy, precision, recall, F1-score, AUC-ROC i matriu de confusió
- Seleccionar l'algoritme de classificació més adequat en funció de les característiques del problema de negoci i del dataset.

Contingut

- 3.1. Fonaments de la classificació supervisada i aplicacions empresarials: detecció de frau, concessió de crèdits i predicció d'abandonament de clients
- 3.2. Regressió logística: fonaments i interpretació.
- 3.3. Arbres de decisió: criteris de divisió i control de l'overfitting.
- 3.4. Random forest como ensemble method.
- 3.5. Mètriques d'avaluació: accuracy, precision, recall, F1-score, AUC-ROC i matriu de confusió.
- 3.6. Validació creuada, selecció de models, ajust d'hiperparàmetres.

TEMA 4. MODELS DE REGRESSIÓ

Resultats específics de l'aprenentatge

L'estudiant després d'estudiar el tema i realitzar els exercicis, serà capaç de:

- Comprendre els fonaments de la regressió lineal simple i múltiple.
- Aplicar tècniques de regularització per gestionar el sobreajust: Ridge i Lasso
- Entrenar i avaluar models de regressió utilitzant scikit-learn: regressió lineal, arbres de regressió i mètodes ensemble.
- Interpretar mètriques d'avaluació de regressió: MAE, RMSE i R^2 .
- Seleccionar l'algoritme més adequat segons el problema de negoci i les característiques de les dades
- Comprendre i interpretar els resultats d'un model de regressió per a la presa de decisions

Contingut

- 4.1. Regressió lineal simple i múltiple: fonaments i interpretació.
- 4.2. Regressió regularitzada: Ridge i Lasso com a tècniques per evitar el sobreajust
- 4.3. Arbres de regressió i mètodes ensemble per a regressió.
- 4.4. Mètriques d'avaluació: MAE, RMSE, R^2 .
- 4.5. Criteris de selecció de l'algoritme segons el problema de negoci
- 4.6. Comprensió de models de regressió: importància de variables i comunicació de resultats

TEMA 5. TÈCNIQUES DE CLUSTERING I SEGMENTACIÓ

Resultats específics de l'aprenentatge

L'estudiant després d'estudiar el tema i realitzar els exercicis, serà capaç de:

- Comprendre els fonaments de l'aprenentatge no supervisat i distingir-lo de l'aprenentatge supervisat.
- Aplicar l'algoritme K-Means per segmentar datasets, determinant el nombre òptim de clusters mitjançant el mètode del colze i l'índex de silueta.

- Implementar clustering jeràrquic i interpretar dendrogrames per a la presa de decisions sobre el nombre de grups.
- Comprendre el concepte de reducció de dimensionalitat mitjançant PCA i la seva utilitat com a pas previ al clustering.
- Traslladar els resultats del clustering a perfils de segments accionables per al negoci.

Contingut

- 5.1. Fonaments de l'aprenentatge no supervisat i aplicacions de negoci.
- 5.2. K-Means: algoritme, selecció de K mitjançant el mètode del colze i índex de silueta
- 5.3. Clustering jeràrquic aglomeratiu i dendrogrames.
- 5.4. Reducció de dimensionalitat mitjançant PCA: concepte i aplicació en clustering.
- 5.5. Segmentació de clients: RFM (Recency, Frequency, Monetary) i aplicacions de màrqueting.
- 5.6. Interpretació i comunicació de perfils de segments per a la presa de decisions

TEMA 6. REGLES D'ASSOCIACIÓ I SISTEMES DE RECOMANACIÓ

Resultats específics de l'aprenentatge

L'estudiant després d'estudiar el tema i realitzar els exercicis, serà capaç de:

- Comprendre el concepte de regla d'associació i identificar aplicacions empresarials rellevants.
- Aplicar l'algoritme Apriori per descobrir patrons de comportament en dades de transaccions.
- Interpretar les mètriques clau de les regles d'associació: suport, confiança i lift.
- Distingir entre collaborative filtering i content-based filtering, identificant casos d'ús per a cada enfocament.
- Traslladar els resultats d'un sistema de recomanació a decisions de negoci: recomanació de productes, personalització i cross-selling.

Contingut

- 6.1. Aprenentatge no supervisat basat en associació: descobriment de patrons i aplicacions empresarials.
- 6.2. Regles d'associació i anàlisi de cistella de la compra: algoritme Apriori i mètriques clau.
- 6.3. Sistemes de recomanació: collaborative filtering i content-based filtering.
- 6.4. Aplicacions empresarials: recomanació de productes, personalització i cross-selling.
- 6.5. Interpretació i comunicació de resultats per a la presa de decisions de negoci.

5. METODOLOGIA D'ENSENYAMENT I APRENTATGE

L'assignatura Mineria de Dades, situada en el tercer curs de la menció d'Analítica de Negoci, integra tècniques estadístiques, computacionals i d'aprenentatge automàtic per extreure coneixement accionable a partir de grans volums de dades. Atès el seu caràcter eminentment aplicat, la modalitat semipresencial combina sessions presencials de treball pràctic intensiu amb sessions asíncrones que permeten consolidar l'aprenentatge autònom a Python, seguint el cicle complet d'un projecte de mineria de dades (CRISP-DM).

La modalitat s'articula en 6 sessions presencials en aula física, 6 sessions asíncrones (o síncrones quan així ho determini el professorat) i 1 sessió d'examen final presencial, sumant un total de 13 sessions.

Sessions presencials (6 sessions):

Cada sessió combina una part expositiva i una part pràctica d'un mínim d'hora i mitja. En la part teòrica, el professorat introdueix els conceptes, algorismes i marcs metodològics del bloc temàtic corresponent recolzant-se en el material docent i els notebooks disponibles en el campus virtual. S'utilitzaran casos d'ús empresarials reals (detecció de frau, segmentació de clients, sistemes de recomanació, predicció de churn) per contextualitzar cada tècnica i facilitar la transferència a l'entorn professional. Es fomentarà la participació activa mitjançant la discussió crítica de resultats, la comparació entre algorismes i la reflexió sobre les implicacions ètiques i limitacions dels models. En la part pràctica, l'estudiant (de forma individual o en grup) implementarà, entrenarà i avaluarà models a Python. L'ordre d'ambdues parts podrà variar segons criteri pedagògic del professorat. Almenys 4 de les 6 sessions presencials inclouran obligatòriament part pràctica amb activitat avaluable. No s'avisarà amb antelació del moment exacte en què tindrà lloc l'activitat pràctica, per tal d'afavorir l'assistència completa a cada sessió.

Sessions asíncrones (6 sessions):

Les sessions asíncrones s'articularen mitjançant materials docents, notebooks de Python estructurats per tema i recursos publicats al campus virtual, preparats específicament per als continguts de l'assignatura. Almenys 4 d'aquestes 6 sessions inclouran una part pràctica d'un mínim d'hora i mitja que l'estudiant podrà desenvolupar de forma asíncrona, o de

forma síncrona si així ho decideix el professorat. Almenys 4 sessions asíncrones comptaran amb algun tipus d'activitat avaluable.

Les principals activitats que es realitzaran al llarg de les sessions presencials i asíncrones són:

- Estudi i preparació dels continguts a partir del material docent i notebooks del campus virtual.
- Classes expositives (presencials o asíncrones) sobre fonaments de machine learning, preprocessament, modelatge supervisat i no supervisat, i sistemes de recomanació.
- Homologació, neteja i preprocessament de datasets amb pandas i scikit-learn (EDA, tractament de valors absents, codificació, escalat).
- Implementació, entrenament i avaluació de models supervisats (classificació i regressió) i no supervisats (clustering, regles d'associació) a Python.
- Construcció de pipelins complets de modelatge amb scikit-learn, aplicant validació creuada i ajust d'hiperparàmetres.
- Interpretació i comunicació de resultats tècnics en llenguatge de negoci, adaptats a audiències no especialitzades.
- Lliuraments pràctics per bloc temàtic, amb elaboració prèvia autònoma i resolució o presentació obligatòria en sessió.

Es treballarà amb notebooks de Python estructurats per tema, disponibles al campus virtual, perquè l'estudiant pugui consolidar l'aprenentatge autònom fora de les sessions amb exercicis equivalents als treballats a classe. La progressió tècnica anirà des del preprocessament i els models supervisats fins a les tècniques no supervisades i els sistemes de recomanació.

L'ordre en què es desenvolupen la part teòrica i la part pràctica dins de cada sessió el determinarà el professor segons criteris pedagògics, podent variar entre sessions. No s'avisarà amb antelació del moment exacte en què tindrà lloc l'activitat pràctica, per tal d'afavorir l'assistència completa a cada sessió.

6. AVALUACIÓ

Les tasques i activitats avaluatives s'ajustaran al contingut del material docent exposat a classe i facilitat al Campus per comprovar que l'alumnat els ha consolidat. D'acord amb el Pla Bolonya, el model premia l'esforç constant i continuat de l'estudiantat. Un 40% de la

nota s'obté de l'avaluació contínua i el 60% restant, de l'examen final presencial. L'examen final té dues convocatòries.

La nota final de l'assignatura (NF) es calcularà a partir de la següent fórmula:

- **NF = Nota Examen Final x 60% + Nota Avaluació Continuada x 40%**
- Nota mínima de l'examen final per calcular la NF serà de 40 punts sobre 100.
- L'assignatura queda aprovada amb una NF igual o superior a 50 punts sobre 100.

Tipus d'activitat	Descripció	% Avaluació contínua	
Lliuraments per tema:			20%
Ex. Tema 1	Lliurament per Classlife	10%	
Ex. Tema 2	Lliurament per Classlife	20%	
Ex. Tema 3	Lliurament per Classlife	20%	
Ex. Tema 4	Lliurament per Classlife	20%	
Ex. Tema 5	Lliurament per Classlife	20%	
Ex. Tema 6	Lliurament per Classlife	10%	
Examen Parcial I			10%
	Examen Parcial I	100%	
Examen Parcial II			10%
	Examen Parcial II	100%	
Examen final			60%
	Examen final	100%	

7. BIBLIOGRAFIA

7.1. BIBLIOGRAFIA Bàsica

- Géron, A. (2025). Hands-On Machine Learning with Scikit-Learn and Pytorch. O'Reilly Media.
- Raschka, S. i Mirjalili, V. (2022). Machine Learning with PyTorch and Scikit-Learn. Packt Publishing.
- Tan, P.-N., Steinbach, M., Karpapne, A., & Kumar, V. (2019). Introduction to data mining (2nd ed.). Pearson.