



centro adscrito a:



UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH

GUIA DOCENT D'ANÀLISI PREDICTIVA I MINERIA DE DADES 2026-27

DADES GENERALS

Número:	Anàlisi predictiva i mineria de dades
Codi:	801409 (MKTTECH)
Curs:	2026-27
Titulació:	Grau en Màrqueting, Innovació i Tecnologia
Nº de crèdits (ECTS):	6
Ubicació en el pla d'estudis:	3r. Curs, 1r. quadrimestre
Departament:	Mètodes Quantitatius
Responsable departament:	Dr. Iván Romero
Data de l'última revisió:	2026
Professor Responsable:	Dr. Ignasi Gómez-Sebastià

DESCRIPCIÓ GENERAL

La mineria de dades és una ciència que combina coneixements estadístics, computacionals i d'Intel·ligència Artificial per extreure coneixement d'alt valor a partir de grans volums de dades d'escàs valor. En un món on la quantitat de dades disponibles creix de forma exponencial, la capacitat de descobrir patrons, predir comportaments i classificar col·lectius resulta crucial per a la presa de decisions estratègiques.

L'assignatura s'estructura en cinc blocs temàtics que mostren el cicle complet d'un projecte de mineria de dades: des de la comprensió del problema i les dades involucrades a mètodes predictius i classificatoris, passant pel pre-processament i l'anàlisi exploratòria de dades. Al llarg del curs s'emprarà el llenguatge Python com a eina principal, amb llibreries com *pandes*, *scikit-learn*, *matplotlib* i *seaborn*.

El primer bloc proporciona el marc conceptual de la mineria de dades i el machine learning, motivant aquestes ciències dins del marc del màrqueting digital. S'introdueixen els conceptes de dada i característica, així diferents tipus de variables. A continuació, s'exploren tècniques estadístiques i de visualització per explorar satisfactòriament les diferents variables.

El segon bloc aborda el pre-processament de dades, fase crítica que determina la qualitat i rellevància de qualsevol model posterior. S'estudien tècniques de neteja, transformació i tractament de valors absents, incloent tècniques capdavanteres com el *one-hot encoding* o les transformacions *min-max*. S'introdueixen nocions d'integració i agrupació de dades així com els conceptes de variable objectiu i variables predictores.

El tercer bloc cobreix els models de regressió, tant basats en estadística com basats en intel·ligència artificial. S'explica en profunditat la morfologia d'un projecte de mineria de dades enfocat a regressió. Es fa esment a conceptes capdavaners com *Nested Cross Validation* i *PCR*. S'expliquen mètriques per avaluar la qualitat del model i es realitza un projecte aplicant els conceptes introduïts.

El quart bloc cobreix els models de classificació i clustering. S'explica en profunditat la morfologia d'un projecte de mineria de dades enfocat a classificació. Es reforça el concepte de *Nested Cross Validation* i s'introdueix el concepte de *PCA*. Es fa esment als models probabilístics i en com poden ser utilitzats per obtenir un mètode de pre-ordre de les dades. S'expliquen mètriques per avaluar la qualitat del model i es realitza un projecte aplicant els conceptes introduïts.

2. OBJECTIUS

En finalitzar el curs l'estudiant serà capaç de:

- Integrar coneixement expert de Màrqueting amb conceptes d'analítica de dades i machine learning.
- Familiaritzar-se amb les mesures bàsiques de l'estadística descriptiva per a la descripció d'un conjunt de dades.
- Oferir una primera aproximació a la representació i interpretació de les dades recollides i resumides, ja sigui en forma matemàtica o gràfica.
- Adquirir habilitats per interpretar la distribució de dades i detectar possibles valors atípics o anomalies.
- Incorporar l'ús de Python que permetrà donar una visió més pràctica i moderna de l'exploració de les dades.
- Desenvolupar habilitats crítiques per avaluar la qualitat i la confiabilitat de les dades.
- Fomentar la capacitat per comunicar de manera efectiva els requeriments i els resultats d'una anàlisi estadística, facilitant la comunicació i coordinació amb un expert en analítica de dades.
- Aplicar els conceptes apresos en situacions del món real, com la interpretació de dades d'enquestes, estudis científics, o anàlisi de dades de màrqueting.
- Tenir present i avaluar l'impacte del biaix en els diferents models.

3. RESULTATS DE L'APRENTATGE

- Diferenciar entre els conceptes de dada i característica, recalcant la seva importància en la presa de decisions i en la resolució de problemes.
- Aplicar tècniques computacionals per a l'anàlisi i la interpretació de les dades.
- Aplicar tècniques de regressió a dades de màrqueting per poder predir valors futurs.
- Aplicar tècniques de classificació a dades de màrqueting per poder predir comportaments futurs.

4. CONTINGUTS

TEMA 1. L'ANÀLISI DESCRIPTIVA EN L'ERA DEL MÀRQUETING DIGITAL

Resultats específics de l'aprenentatge

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

- Entendre el paper crucial dels mètodes estadístics i computacionals en l'era del màrqueting digital.
- Diferenciar entre dada en brut i característica pre-processada.
- Entendre els conceptes bàsics dels algorismes de Machine Learning, incloent:
 - Les diferències entre l'aprenentatge supervisat i no supervisat.
 - Les diferències entre els problemes de regressió, classificació i clustering.
- Entendre el cicle de vida d'un projecte de mineria de dades en base a l'estàndard CRISP-DM.
- Entendre les causes dels biaixos en els models de Machine Learning i els seus riscos en entorns de Màrqueting.

Contingut

- 1.1. Motivació de l'analítica descriptiva per al màrqueting digital.
- 1.2. De la dada a la característica.
- 1.3. Diferents tipus de variables.
- 1.4. Introducció als algorismes de Machine Learning.
 - a. Aprenentatge supervisat i no supervisat.
 - b. Classificació, regressió i clustering.
 - c. Exemples reals de projectes i identificació de característiques.
- 1.5. El cicle CRISP-DM per a un projecte de mineria de dades.
- 1.6. El biaix en els models de Machine Learning.

TEMA 2. PRE-PROCESSAT I PREPARACIÓ DE DADES

Resultats específics de l'aprenentatge

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

- Explorar un conjunt de dades detectant els problemes típics que impedeixen analitzar-nos satisfactòriament mitjançant tècniques d'aprenentatge automàtic.
- Solucionar els problemes detectats en el punt anterior.
- Transformar, integrar, agrupar i normalitzar les dades perquè puguin ser processats satisfactòriament mitjançant tècniques d'aprenentatge automàtic.

Contingut

- 2.1. Motivació en entorns de màrqueting i problemes més comuns.
- 2.2. Anàlisi exploratòria del conjunt de dades.
- 2.3. Resolució de problemes I: Valors faltants.
- 2.4. Resolució de problemes II: Valors duplicats.
- 2.5. Resolució de problemes II: Valors extrems i outliers.
- 2.6. Transformació de variables aplicada al màrqueting.
- 2.7. Integració de dades.
- 2.8. Agrupació de dades.
- 2.9. Entenent les variables: Cura objectiu, variables predictores i mostra.
- 2.10. Normalització de dades.

TEMA 3. MODELS DE REGRESSIÓ

Resultats específics de l'aprenentatge

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

- Aplicar models de regressió a un projecte de mineria de dades.
- Avaluar la bondat d'un model de regressió i aplicar-lo de forma crítica per prendre decisions estratègiques.
- Entendre com es poden ajustar automàticament els hiper-paràmetres d'un model de regressió.
- Identificar les variables que tenen més impacte en el model de regressió.

Contingut

- 3.1. Regressió lineal.
- 3.2. Regressió logística.
- 3.3. Regressió amb DNNs.
- 3.4. Concepte de *Nested Cross Validation*.

- 3.5. PCR per identificar variables rellevants.
- 3.6. Avaluació de la qualitat del Model.
- 3.7. Estudi del biaix en els models de regressió.

TEMA 4. MODELS DE CLASSIFICACIÓ

Resultats específics de l'aprenentatge

L'estudiant després d'estudiar els temes d'aquest bloc i realitzar els exercicis, serà capaç de:

- Aplicar models de classificació supervisada a un projecte de mineria de dades.
- Aplicar models de classificació no supervisada un projecte de mineria de dades.
- Avaluar la bondat d'un model de classificació (supervisada o no supervisada) i aplicar-lo de forma crítica per prendre decisions estratègiques.
- Entendre com es poden ajustar automàticament els hiper-paràmetres d'un model de classificació supervisada.
- Identificar les variables que tenen més impacte en el model de classificació supervisada.

Contingut

- 4.1. Arbres de decisió i models explicatius supervisats.
- 4.2. Random Forest i models no explicatius supervisats.
- 4.3. Concepte de *Nested Cross Validation* i balanceig de classes.
- 4.4. PCA per identificar variables rellevants.
- 4.5. Avaluació de la qualitat d'un model de classificació supervisada.
- 4.6. Models probabilístics i construcció de pre-ordre.
- 4.7. Models no supervisats: Clustering.
- 4.8. Models no supervisats: Segmentació.
- 4.9. Avaluació de la qualitat d'un model de classificació no supervisada.

5. METODOLOGIA D'ENSENYAMENT I APRENTATGE

L'assignatura Minería de Dades s'estructura en 12 sessions presencials teoricopràctiques més la sessió d'examen final. El seu caràcter eminentment pràctic (orientada al Grau en Màrqueting, Innovació i Tecnologia i contextualitzada en l'ecosistema del màrqueting digital) exigeix que cada sessió combini la presentació de marcs conceptuals i algorítmics

amb l'aplicació immediata a Python sobre datasets reals de màrqueting, fent especial èmfasi en la interpretació de resultats i l'extracció de conclusions accionables per al negoci.

Part teòrica

El professorat introdueix els continguts del bloc temàtic corresponent recolzant-se en el material docent i els notebooks de Python disponibles al Campus Virtual. La presentació parteix sempre de la motivació en entorns de màrqueting digital (segmentació de clients, predicció de comportament, detecció de patrons de compra) per contextualitzar cada tècnica estadística o algoritme de machine learning. Es fomentarà la participació activa mitjançant la discussió crítica de resultats, l'avaluació del biaix en els models i la connexió entre l'anàlisi tècnica i el coneixement expert de màrqueting.

Part pràctica

Un mínim de 10 sessions inclourà una part pràctica obligatòria en la qual l'estudiant — individualment o en grups de fins a 5 membres— treballarà amb els continguts de la sessió sobre datasets reals a Jupyter Notebook. Les tipologies d'activitats pràctiques inclouen, entre d'altres:

- Anàlisi descriptiva i exploració de dades de màrqueting: identificació de tipus de variables, detecció de patrons i anomalies.
- Preprocessament de dades: detecció i correcció de valors faltants, duplicats i outliers; transformacions, one-hot encoding, normalització min-max.
- Construcció, entrenament i avaluació de models de regressió (lineal, logística, DNNs) amb Nested Cross Validation i PCR.
- Construcció i avaluació de models de classificació supervisada (arbres de decisió, Random Forest) i no supervisada (clustering, segmentació), amb PCA i balanceig de classes.
- Avaluació crítica del biaix en els models i trasllat de resultats tècnics a llenguatge i decisions de màrqueting.
- Activitats avaluable basades en casos d'ús reals de màrqueting, desenvolupades íntegrament durant la sessió i lliurades a través de Classlife.

Es fa especial èmfasi en la capacitat de l'estudiant per aportar coneixement expert de màrqueting al llarg de tot el projecte de mineria de dades, actuant com a pont entre el departament d'anàlisi i la resta d'àrees de l'organització.

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 60% de la nota s'obté de l'avaluació contínua i el 40% 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 40% + Nota Avaluació Continuada x 60%**
- 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.

Grup presencial:

Tipus d'activitat	Descripció	% Avaluació contínua	
Lliuraments:			60%
Activitat 1	Tema 1: Introducció i Teoria. Desenvolupada a classe. Lliurament per Classlife	17%	
Activitat 2	Tema 2: Pre-processament de dades Desenvolupada a classe. Lliurament per Classlife	16%	
Activitat 3	Tema 2: Normalització de dades Desenvolupada a classe. Lliurament per Classlife	16%	
Activitat 4	Tema 3: Models de regressió Desenvolupada a classe. Lliurament per Classlife	17%	
Activitat 5	Tema 4: Classificació supervisada Desenvolupada a classe. Lliurament per Classlife	17%	
Activitat 6	Tema 4: Classificació no supervisada Desenvolupada a classe. Lliurament per Classlife	17%	
Examen final			40%
	Examen final 10% Teoria en Paper 90% Pràctica en Python	100%	

7. BIBLIOGRAFIA

7.1. BIBLIOGRAFIA BÀSICA

Dolnicar, S., Grün, B., & Leisch, F. (2018). Market segmentation analysis. In Market segmentation analysis: Understanding it, doing it, and making It useful (pp. 11-22). Singapore: Springer Singapore.

Géron, A. (2022). Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow. "O'Reilly Media, Inc."

Katsov, Ilya. (2017). Introduction to algorithmic marketing: Artificial intelligence for marketing operations.