



centro adscrito a:



UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH

GUÍA DOCENTE DE ANÁLISIS PREDICTIVO Y MINERÍA DE DATOS 2026-27

DATOS GENERALES

Número:	Análisis predictivo y minería de Datos
Código:	801409 (MKTTECH)
Curso:	2026-27
Titulación:	Grado en Marketing, Innovación y Tecnología
Nº de créditos (ECTS):	6
Ubicación en el plan de estudios:	3º. Curso, 1º. cuatrimestre
Departamento:	Métodos Cuantitativos
Responsable departamento:	Dr. Iván Romero
Fecha de la última revisión:	2026
Profesor Responsable:	Dr. Ignasi Gómez-Sebastià

1. DESCRIPCIÓN GENERAL

La minería de datos es una ciencia que combina conocimientos estadísticos, computacionales y de Inteligencia Artificial para extraer conocimiento de alto valor a partir de grandes volúmenes de datos de escaso valor. En un mundo donde la cantidad de datos disponibles crece de forma exponencial, la capacidad de descubrir patrones, predecir comportamientos y clasificar colectivos resulta crucial para la toma de decisiones estratégicas.

La asignatura se estructura en cinco bloques temáticos que muestran el ciclo completo de un proyecto de minería de datos: desde la comprensión del problema y los datos involucrados a métodos predictivos y clasificatorios, pasando por el pre-procesamiento y el análisis exploratorio de datos. A lo largo del curso se empleará el lenguaje Python como herramienta principal, con librerías como *pandas*, *scikit-learn*, *matplotlib* y *seaborn*.

El primer bloque proporciona el marco conceptual de la minería de datos y el machine learning, motivando dichas ciencias dentro del marco del marketing digital. Se introducen los conceptos de dato y característica, así diferentes tipos de variables. A continuación, se exploran técnicas estadísticas y de visualización para explorar satisfactoriamente las distintas variables.

El segundo bloque aborda el pre-procesamiento de datos, fase crítica que determina la calidad y relevancia de cualquier modelo posterior. Se estudian técnicas de limpieza, transformación y tratamiento de valores ausentes, incluyendo técnicas punteras como el *one-hot encoding* o las transformaciones *min-max*. Se introducen nociones de integración y agrupación de datos así como los conceptos de variable objetivo y variables predictoras.

El tercer bloque cubre los modelos de regresión, tanto basados en estadística como basados en inteligencia artificial. Se explica en profundidad la morfología de un proyecto de minería de datos enfocado a regresión. Se hace hincapié en conceptos punteros como *Nested Cross Validation* y *PCR*. Se explican métricas para evaluar la calidad del modelo y se realiza un proyecto aplicando los conceptos introducidos.

El cuarto bloque cubre los modelos de clasificación y clustering. Se explica en profundidad la morfología de un proyecto de minería de datos enfocado a clasificación. Se refuerza el concepto de *Nested Cross Validation* y se introduce el concepto de *PCA*. Se hace hincapié en los modelos probabilísticos y en como pueden ser utilizados para obtener un método de pre-orden de los datos. Se explican métricas para evaluar la calidad del modelo y se realiza un proyecto aplicando los conceptos introducidos.

2. OBJETIVOS

Al finalizar el curso el estudiante será capaz de:

- Integrar conocimiento experto de Marketing con conceptos de analítica de datos y machine learning.
- Familiarizarse con las medidas básicas de la estadística descriptiva para la descripción de un conjunto de datos.
- Ofrecer una primera aproximación a la representación e interpretación de los datos recogidos y resumidos, ya sea en forma matemática o gráfica.
- Adquirir habilidades para interpretar la distribución de datos y detectar posibles valores atípicos o anomalías.
- Incorporar el uso de Python que permitirá dar una visión más práctica y moderna de la exploración de los datos
- Desarrollar habilidades críticas para evaluar la calidad y la confiabilidad de los datos.
- Fomentar la capacidad para comunicar de manera efectiva los requerimientos y los resultados de un análisis estadístico, facilitando la comunicación y coordinación con un experto en analítica de datos.
- Aplicar los conceptos aprendidos en situaciones del mundo real, como la interpretación de datos de encuestas, estudios científicos, o análisis de datos de marketing.
- Tener presente y evaluar el impacto del sesgo en los distintos modelos.

3. RESULTADOS DEL APRENDIZAJE

- Diferenciar entre los conceptos de dato y característica, recalando su importancia en la toma de decisiones y en la resolución de problemas.
- Aplicar técnicas computacionales para el análisis y la interpretación de los datos.
- Aplicar técnicas de regresión a datos de marketing para poder predecir valores futuros.
- Aplicar técnicas de clasificación a datos de marketing para poder predecir comportamientos futuros.

4. CONTENIDOS

TEMA 1. EL ANÁLISIS DESCRIPTIVO EN LA ERA DEL MARKETING DIGITAL

Resultados específicos del aprendizaje

El estudiante después de estudiar el tema y realizar los ejercicios, será capaz de:

- Entender el papel crucial de los métodos estadísticos y computacionales en la era del marketing digital.
- Diferenciar entre dato en bruto y característica pre-procesada.
- Entender los conceptos básicos de los algoritmos de Machine Learning, incluyendo:
 - Las diferencias entre el aprendizaje supervisado y no supervisado.
 - Las diferencias entre los problemas de regresión, clasificación y clustering.
- Entender el ciclo de vida de un proyecto de minería de datos en base al estándar CRISP-DM.
- Entender las causas de los sesgos en los modelos de Machine Learning y sus riesgos en entornos de Marketing.

Contenido

- 1.1. Motivación de la analítica descriptiva para el marketing digital.
- 1.2. Del dato a la característica.
- 1.3. Diferentes tipos de variables.
- 1.4. Introducción a los algoritmos de Machine Learning.
 - a. Aprendizaje supervisado y no supervisado.
 - b. Clasificación, regresión y clustering.
 - c. Ejemplos reales de proyectos e identificación de características.
- 1.5. El ciclo CRISP-DM para un proyecto de minería de datos.
- 1.6. El sesgo en los modelos de Machine Learning.

TEMA 2. PRE-PROCESADO Y PREPARACIÓN DE DATOS

Resultados específicos del aprendizaje

El estudiante después de estudiar el tema y realizar los ejercicios, será capaz de:

- Explorar un conjunto de datos detectando los problemas típicos que impiden analizarlos satisfactoriamente mediante técnicas de aprendizaje automático.
- Solucionar los problemas detectados en el punto anterior.
- Transformar, integrar, agrupar y normalizar los datos para que puedan ser procesados satisfactoriamente mediante técnicas de aprendizaje automático.

Contenido

- 2.1. Motivación en entornos de marketing y problemas más comunes.
- 2.2. Análisis exploratorio del conjunto de datos.
- 2.3. Resolución de problemas I: Valores faltantes.
- 2.4. Resolución de problemas II: Valores duplicados.
- 2.5. Resolución de problemas II: Valores extremos y outliers.
- 2.6. Transformación de variables aplicada al marketing.
- 2.7. Integración de datos.
- 2.8. Agrupación de datos.
- 2.9. Entendiendo las variables: Variable objetivo, variables predictoras y muestra.
- 2.10. Normalización de datos.

TEMA 3. MODELOS DE REGRESIÓN

Resultados específicos del aprendizaje

El estudiante después de estudiar el tema y realizar los ejercicios, será capaz de:

- Aplicar modelos de regresión a un proyecto de minería de datos.
- Evaluar la bondad de un modelo de regresión y aplicarlo de forma crítica para tomar decisiones estratégicas.
- Entender como se pueden ajustar automáticamente los hiper-parámetros de un modelo de regresión.
- Identificar las variables que tienen mayor impacto en el modelo de regresión.

Contenido

- 3.1. Regresión lineal.
- 3.2. Regresión logística.
- 3.3. Regresión con DNNs.
- 3.4. Concepto de *Nested Cross Validation*.

- 3.5. PCR para identificar variables relevantes.
- 3.6. Evaluación de la calidad del Modelo.
- 3.7. Estudio del sesgo en los modelos de regresión.

TEMA 4. MODELOS DE CLASIFICACIÓN

Resultados específicos del aprendizaje

El estudiante después de estudiar los temas de este bloque y realizar los ejercicios, será capaz de:

- Aplicar modelos de clasificación supervisada a un proyecto de minería de datos.
- Aplicar modelos de clasificación no supervisada un proyecto de minería de datos.
- Evaluar la bondad de un modelo de clasificación (supervisada o no supervisada) y aplicarlo de forma crítica para tomar decisiones estratégicas.
- Entender como se pueden ajustar automáticamente los hiper-parámetros de un modelo de clasificación supervisada.
- Identificar las variables que tienen mayor impacto en el modelo de clasificación supervisada.

Contenido

- 4.1. Árboles de decisión y modelos explicativos supervisados.
- 4.2. Random Forest y modelos no explicativos supervisados.
- 4.3. Concepto de *Nested Cross Validation* y balanceo de clases.
- 4.4. PCA para identificar variables relevantes.
- 4.5. Evaluación de la calidad de un modelo de clasificación supervisada.
- 4.6. Modelos probabilísticos y construcción de pre-orden.
- 4.7. Modelos no supervisados: Clustering.
- 4.8. Modelos no supervisados: Segmentación.
- 4.9. Evaluación de la calidad de un modelo de clasificación no supervisada.

5. METODOLOGÍA DE ENSEÑANZA Y APRENDIZAJE

La asignatura Análisis predictivo y minería de Datos se estructura en 12 sesiones presenciales teórico-prácticas más la sesión de examen final. Su carácter eminentemente práctico (orientada al Grado en Marketing, Innovación y Tecnología y contextualizada en el ecosistema del marketing digital) exige que cada sesión combine la presentación de marcos conceptuales y algorítmicos con la aplicación inmediata en Python sobre datasets reales de marketing, haciendo especial énfasis en la interpretación de resultados y la extracción de conclusiones accionables para el negocio.

Parte teórica

El profesorado introduce los contenidos del bloque temático correspondiente apoyándose en el material docente y los notebooks de Python disponibles en el Campus Virtual. La presentación parte siempre de la motivación en entornos de marketing digital (segmentación de clientes, predicción de comportamiento, detección de patrones de compra) para contextualizar cada técnica estadística o algoritmo de machine learning. Se fomentará la participación activa mediante la discusión crítica de resultados, la evaluación del sesgo en los modelos y la conexión entre el análisis técnico y el conocimiento experto de marketing.

Parte práctica

Un mínimo de 10 sesiones incluirá una parte práctica obligatoria en la que el estudiante — individualmente o en grupos de hasta 5 miembros— trabajará con los contenidos de la sesión sobre datasets reales en Jupyter Notebook. Las tipologías de actividades prácticas incluyen, entre otras:

- Análisis descriptivo y exploración de datos de marketing: identificación de tipos de variables, detección de patrones y anomalías.
- Preprocesamiento de datos: detección y corrección de valores faltantes, duplicados y outliers; transformaciones, one-hot encoding, normalización min-max.
- Construcción, entrenamiento y evaluación de modelos de regresión (lineal, logística, DNNs) con Nested Cross Validation y PCR.
- Construcción y evaluación de modelos de clasificación supervisada (árboles de decisión, Random Forest) y no supervisada (clustering, segmentación), con PCA y balanceo de clases.
- Evaluación crítica del sesgo en los modelos y traslado de resultados técnicos a lenguaje y decisiones de marketing.
- Actividades evaluables basadas en casos de uso reales de marketing, desarrolladas íntegramente durante la sesión y entregadas a través de Classlife.

Se hace especial énfasis en la capacidad del estudiante para aportar conocimiento experto de marketing a lo largo de todo el proyecto de minería de datos, actuando como puente entre el departamento de analítica y el resto de áreas de la organización.

6. EVALUACIÓN

Las tareas y actividades evaluativas se ajustarán al contenido del material docente expuesto en clase y facilitado en el Campus para comprobar que el alumnado los ha consolidado. De acuerdo con el Plan Bolonia, el modelo premia el esfuerzo constante y continuado del estudiantado. Un 60% de la nota se obtiene de la evaluación continua y el 40% restante, del examen final presencial. El examen final tiene dos convocatorias.

La nota final de la asignatura (NF) se calculará a partir de la siguiente fórmula:

- **NF = Nota Examen Final x 40% + Nota Evaluación Continuada x 60%**
- Nota mínima del examen final para calcular la NF será de 40 puntos sobre 100.
- La asignatura queda aprobada con una NF igual o superior a 50 puntos sobre 100.

Grupo presencial:

Tipo de actividad	Descripción	% Evaluación continua	
Entregas:			60%
Actividad 1	Tema 1: Introducción y Teoría. Desarrollada en clase. Entrega por Classlife	17%	
Actividad 2	Tema 2: Pre-procesado de datos Desarrollada en clase. Entrega por Classlife	16%	
Actividad 3	Tema 2: Normalización de datos Desarrollada en clase. Entrega por Classlife	16%	
Actividad 4	Tema 3: Modelos de regresión Desarrollada en clase. Entrega por Classlife	17%	
Actividad 5	Tema 4: Clasificación supervisada Desarrollada en clase. Entrega por Classlife	17%	
Actividad 6	Tema 4: Clasificación no supervisada Desarrollada en clase. Entrega por Classlife	17%	
Examen final			40%
	Examen final 10% Teoría en Papel 90% Práctica en Python	100%	

7. BIBLIOGRAFÍA

7.1. BIBLIOGRAFIA BÁSICA

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.