



# Guía didáctica

## ASIGNATURA: Minería de datos

**Título:** *Máster Universitario en Big Data y Ciencia de Datos*

**Materia:** *Ciencia de Datos*

**Créditos:** *6 ECTS*

**Código:** *05MBID*

**Curso:** *Octubre 2023-2024*

# Índice

Índice.....	2
1. Organización general.....	3
1.1. Datos de la asignatura .....	3
1.2. Equipo docente .....	3
1.3. Introducción a la asignatura .....	3
1.4. Competencias y resultados de aprendizaje .....	4
2. Contenidos/temario .....	5
3. Metodología .....	5
4. Actividades formativas .....	6
5. Planificación de las sesiones.....	7
6. Evaluación .....	7
6.1. Sistema de evaluación .....	7
6.2. Sistema de calificación .....	9
7. Bibliografía.....	9
7.1. Bibliografía de referencia .....	9
7.2. Bibliografía complementaria .....	10
7.3. Videos docentes .....	10

# 1. Organización general

## 1.1. Datos de la asignatura

<b>MATERIA</b>	<b><i>Ciencia de Datos</i></b>
<b>ASIGNATURA</b>	<i>Minería de datos</i> <b>6 ECTS</b>
<b>Carácter</b>	Obligatoria
<b>Cuatrimestre</b>	Primero
<b>Idioma en que se imparte</b>	Castellano
<b>Requisitos previos</b>	No existen
<b>Dedicación al estudio por ECTS</b>	<b>25 horas</b>

## 1.2. Equipo docente

<b>Profesor</b>	<b>D. Jose Murillo</b> <a href="mailto:josea.murillo@professor.universidadviu.com">josea.murillo@professor.universidadviu.com</a>
-----------------	--

## 1.3. Introducción a la asignatura

El objetivo de este curso es introducir al alumnado los conceptos básicos de la minería de datos. Estas técnicas permiten describir y modelar conjuntos de datos complejos y procedimientos automáticos encuadrados bajo la categoría de “minería de datos”; también conocida como Knowledge Discovery in Databases (KDD).

Estos procedimientos sirven para extraer información relevante, generalmente implícita y desconocida de los datos. Este mecanismo está diseñado para detectar patrones y regularidades en dichos datos y que puedan ser usados para predecir respuestas ante nuevas situaciones. Además de conseguir detectar los campos clave que determinan las características del problema, realizar segmentaciones de forma automática, lanzar pronósticos, detectar excepciones, proponer escenarios; en definitiva, anticipar el futuro de forma objetiva y basada en datos reales.

Los objetivos propios de la asignatura Minería de Datos son:

- Introducir al alumno al análisis inteligente de datos en entornos Big Data.
- Que el alumno conozca y sea capaz de aplicar un proceso completo de extracción de conocimiento de una base de datos, con sus diferentes etapas.

- Que el alumno conozca las técnicas provenientes tanto de la estadística como de la Inteligencia Artificial para el análisis de datos, que conozca las diferentes utilidades de estas en cuanto a la adecuación de los diferentes problemas a tratar y que distinga claramente los diferentes resultados (conocimiento) que se podrían alcanzar.

## 1.4. Competencias y resultados de aprendizaje

### COMPETENCIAS GENERALES

CG.1.- Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en el desarrollo y/o aplicación de ideas, a menudo en un contexto de investigación

CG.2.- Que los estudiantes sepan aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio

CG.3.- Que los estudiantes sepan comunicar sus conclusiones y los conocimientos y razones últimas que las sustentan a públicos especializados y no especializados de un modo claro y sin ambigüedades

CG.4.- Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo.

### COMPETENCIAS ESPECÍFICAS DE LA ASIGNATURA

CE02- Utilizar técnicas y herramientas de programación especializada en analítica y procesamiento de datos en entornos de Big Data

CE04- Resolver problemas reales de clasificación, modelización e interpretación de un conjunto de datos en el contexto de Big Data

CE06- Utilizar conjuntamente métodos estadísticos avanzados, recursos informáticos y de diagnóstico estratégico para la elaboración de soluciones Big Data de alto valor añadido

CE07- Aplicar los principales métodos de minería de datos para la extracción de características y la contextualización de información para la toma de decisiones

CE08- Utilizar técnicas de aprendizaje automático en la detección de potenciales dependencias entre un conjunto de variables

CE09- Interpretar los resultados obtenidos del análisis de datos a los que se han aplicado técnicas de aprendizaje automático

CE10- Aplicar técnicas de generación de visualizaciones de datos para el análisis y comunicación de resultados

CE11- Elaborar modelos de decisión estratégica basados en técnicas de clasificación y tratado de datos en función de los requisitos y necesidades del usuario final

CE15- Identificar la solución Big Data óptima para un problema, en términos de eficiencia, eficacia e interpretación de resultados

CE16- Diseñar estrategias de visualización de resultados y presentación de conclusiones obtenidos en el proceso de análisis de datos en un contexto de toma de decisiones

## RESULTADOS DE APRENDIZAJE

Al finalizar esta asignatura se espera que el estudiante sea capaz de:

RA.1.- Saber identificar los distintos tipos de problemas de análisis de datos que pueden darse en un entorno de big data.

RA.2.- Aplicar adecuadamente las principales técnicas (tanto estadísticas como de machine learning), herramientas y tecnologías que se puedan utilizar en la solución.

## 2. Contenidos/temario

- Tema 1: Introducción al proceso KDD.
  - Proyectos de minería de datos y clasificación de los sistemas y tipos de modelos de data mining.
  - Catálogo de recursos técnicos: La Inteligencia Artificial, el Aprendizaje automático (Machine Learning), el aprendizaje estadístico, la Minería de Datos y el nuevo reto del Big Data.
  - El proceso KDD (Knowledge Discovery in Databases).
- Tema 2: Representación gráfica de variables.
- Tema 3: Análisis y transformación de variables.
- Tema 4: Técnicas de agrupamiento.
- Tema 5: Modelos predictivos.
- Tema 6: Comparación, selección y evaluación de modelos.
  - Matriz de confusión.
  - Métricas.
  - Costes.
  - Curvas ROC
- Tema 7: Aplicaciones relacionadas con KDD.
  - Sistemas de recomendación

## 3. Metodología

La metodología de la Universidad Internacional de Valencia (VIU) se caracteriza por una apuesta decidida en un modelo de carácter e-presencial. Así, siguiendo lo estipulado en el calendario de actividades docentes del Título, se impartirán en directo un conjunto de sesiones, que, además, quedarán grabadas para su posterior visionado por parte de aquellos estudiantes que lo necesiten. En todo caso, se recomienda acudir, en la medida de lo posible, a dichas sesiones, facilitando así el intercambio de experiencias y dudas con el docente.

En lo que se refiere a las metodologías específicas de enseñanza-aprendizaje, serán aplicadas por el docente en función de los contenidos de la asignatura y de las necesidades pedagógicas de los estudiantes. De manera general, se impartirán contenidos teóricos y, en el ámbito de las clases prácticas se podrá realizar la resolución de problemas, el estudio de casos y/o la simulación.

Por otro lado, la Universidad y sus docentes ofrecen un acompañamiento continuo al estudiante, poniendo a su disposición foros de dudas y tutorías para resolver las consultas de carácter académico que el estudiante pueda tener. Es importante señalar que resulta fundamental el trabajo autónomo del estudiante para lograr una adecuada consecución de los objetivos formativos previstos para la asignatura.

## 4. Actividades formativas

Durante el desarrollo de cada una de las asignaturas se programan una serie de actividades de aprendizaje que ayudan a los estudiantes a consolidar los conocimientos trabajados.

A continuación, se relacionan las actividades que forman parte de la asignatura:

### 1. Actividades de carácter teórico

Se trata de un conjunto de actividades guiadas por el profesor de la asignatura destinadas a la adquisición por parte de los estudiantes de los contenidos teóricos de la misma. Estas actividades, diseñadas de manera integral, se complementan entre sí y están directamente relacionadas con los materiales teóricos que se ponen a disposición del estudiante (manual, SCORM y material complementario). Estas actividades se desglosan en las siguientes categorías:

- a. Clases expositivas
- b. Sesiones con expertos en el aula
- c. Observación y evaluación de recursos didácticos audiovisuales
- d. Estudio y seguimiento de material interactivo

### 2. Actividades de carácter práctico

Se trata de un conjunto de actividades guiadas y supervisadas por el profesor de la asignatura vinculadas con la adquisición por parte de los estudiantes de los resultados de aprendizaje y competencias de carácter más práctico. Estas actividades, diseñadas con visión de conjunto, están relacionadas entre sí para ofrecer al estudiante una formación completa e integral.

### 3. Tutorías

Se trata de sesiones, tanto de carácter síncrono como asíncrono (e-mail), individuales o colectivas, en las que el profesor comparte información sobre el progreso académico del estudiante y en las que se resuelven dudas y se dan orientaciones específicas ante dificultades concretas en el desarrollo de la asignatura.

### 4. Trabajo autónomo

Se trata de un conjunto de actividades que el estudiante desarrolla autónomamente y que están enfocadas a lograr un aprendizaje significativo y a superar la evaluación de la asignatura. La realización de estas actividades es indispensable para adquirir las competencias y se encuentran entroncadas en el aprendizaje autónomo que consagra la actual ordenación de enseñanzas universitarias. Esta actividad, por su definición, tiene carácter asíncrono.

### 5. Prueba objetiva final

Como parte de la evaluación de cada una de las asignaturas (a excepción de las prácticas y el Trabajo fin de título), se realiza una prueba (examen final). Esta prueba se realiza en

tiempo real (con los medios de control antifraude especificados) y tiene como objetivo evidenciar el nivel de adquisición de conocimientos y desarrollo de competencias por parte de los estudiantes. Esta actividad, por su definición, tiene carácter síncrono.

## 5. Planificación de las sesiones

(Ver anexo)

## 6. Evaluación

### 6.1. Sistema de evaluación

El Modelo de Evaluación de estudiantes en la Universidad se sustenta en los principios del Espacio Europeo de Educación Superior (EEES), y está adaptado a la estructura de formación virtual propia de esta Universidad. De este modo, se dirige a la evaluación de competencias.

Sistema de Evaluación	Ponderación
<b>Portafolio*</b>	<b>60%</b>
Actividad Evaluable 1 (*)	10
Actividad Evaluable 2 (*)	30
Participación en Foro (**)	10
Actividad Evaluativa Continua (**)	10
Sistema de Evaluación	Ponderación
<b>Prueba final*</b>	<b>40 %</b>
<i>La primera parte es un examen tipo test y uno o dos ejercicios prácticos.</i>	

**\* Es requisito indispensable para superar la asignatura aprobar cada apartado (portafolio y prueba final) con un mínimo de 5 para ponderar las calificaciones.**

**\*\* Sólo disponibles en primera convocatoria.**

Los enunciados y especificaciones propias de las distintas actividades serán aportados por el docente, a través del Campus Virtual, a lo largo de la impartición de la asignatura.

Atendiendo a la Normativa de Evaluación de la Universidad, se tendrá en cuenta que la utilización de **contenido de autoría ajena** al propio estudiante debe ser citada adecuadamente en los trabajos entregados. Los casos de plagio serán sancionados con suspenso (0) de la actividad en la que se detecte. Asimismo, el uso de **medios fraudulentos durante las pruebas**

**de evaluación** implicará un suspenso (0) y podrá implicar la apertura de un expediente disciplinario.



## 6.2. Sistema de calificación

La calificación de la asignatura se establecerá en los siguientes cálculos y términos:

Nivel de aprendizaje	Calificación numérica	Calificación cualitativa
Muy competente	9,0 - 10	Sobresaliente
Competente	7,0 - 8,9	Notable
Aceptable	5,0 - 6,9	Aprobado
Aún no competente	0,0 - 4,9	Suspenso

Sin detrimento de lo anterior, el estudiante dispondrá de una **rúbrica simplificada** en el aula que mostrará los aspectos que valorará el docente, como así también los **niveles de desempeño que tendrá en cuenta para calificar las actividades vinculadas a cada resultado de aprendizaje**.

La mención de «**Matrícula de Honor**» podrá ser otorgada a estudiantes que hayan obtenido una calificación igual o superior a 9.0. Su número no podrá exceder del cinco por ciento de los estudiantes matriculados en una materia en el correspondiente curso académico, salvo que el número de estudiantes matriculados sea inferior a 20, en cuyo caso se podrá conceder una sola «Matrícula de Honor».

## 7. Bibliografía

### 7.1. Bibliografía de referencia

- Fayyad, U., Piatetsky-Shapiro, G., Smyth, P. (1996). The KDD Process for Extracting Useful Knowledge from Volumes of Data. Communications of the ACM, November 1996/ Vol 39, Nº 11, 27–34.
- Siegel, E. (2013). Analítica predictiva. Predecir el futuro utilizando Big Data. Anaya Multimedia-Anaya Interactiva.
- Mayer-Schönberger, V., Cukier, K. (2013). Big data. La revolución de los datos masivos. Turner

## 7.2. Bibliografía complementaria

- Agrawal, D., Das, S., Abbadi, A. E. (2010). Big Data and Cloud Computing: New Wine or Just New Bottles?. Proc. of the VLDB 2010, Vol. 3, No. 2.
- Mayer-Schönberger, V., Cukier, K. (2013). Big data. La revolución de los datos masivos. Turner. Olivas, J. A. (2011). Búsqueda eficaz de información en la Web, Edulp, La Plata, Argentina. <http://hdl.handle.net/10915/18401>.
- Piatetsky-Shapiro, G., Frawley, W. (1991). Knowledge Discovery in Databases. AAAI/MIT Press, Cambridge MA.
- Hastie, Trevor and Tibshirani, Robert and Friedman, Jerome (2009), The elements of statistical learning: data mining, inference, and prediction, Springer Science & Business Media

## 7.3. Videos docentes

- Jupyter Notebook for data Science: Publishing, Packt. Jupyter Notebook for Data Science. PACKT Publishing, 2018, [video.alexanderstreet.com/watch/jupyter-notebook-for-data-science](https://video.alexanderstreet.com/watch/jupyter-notebook-for-data-science). Accessed 17 May 2021.
- Minería de datos por José Ángel Olivas Varela