



**Guía docente de la asignatura**

<b>Asignatura</b>	Minería de Datos		
<b>Materia</b>	Computación		
<b>Módulo</b>	Tecnologías Específicas		
<b>Titulación</b>	Grado en Ingeniería Informática, mención Computación		
<b>Plan</b>	545	<b>Código</b>	46970
<b>Periodo de impartición</b>	1 <sup>er</sup> Cuatrimestre	<b>Tipo/Carácter</b>	OB (Mención CO)
<b>Nivel/Ciclo</b>	Grado	<b>Curso</b>	4 <sup>o</sup>
<b>Créditos ECTS</b>	6		
<b>Lengua en que se imparte</b>	Español		
<b>Profesor/es responsable/s</b>	Carlos J. Alonso González Teodoro Calonge Cano		
<b>Datos de contacto (E-mail, teléfono...)</b>	<a href="mailto:calonso@infor.uva.es">calonso@infor.uva.es</a> , 983 185602 <a href="mailto:teodoro@infor.uva.es">teodoro@infor.uva.es</a> , 983 185603		
<b>Horario de tutorías</b>	Consultarlas en siguiente página web: Véase <a href="http://www.uva.es">www.uva.es</a> → Centros → Campus de Valladolid → Escuela Técnica Superior de Ingeniería Informática → Tutorías <a href="http://www.uva.es/opencms/portal/paginas/contenidoDinamico?funcion=D_Tutorias&amp;cod_dpto=041&amp;carpeta=/contenidos/departamentos/informatica/">http://www.uva.es/opencms/portal/paginas/contenidoDinamico?funcion=D_Tutorias&amp;cod_dpto=041&amp;carpeta=/contenidos/departamentos/informatica/</a>		
<b>Departamento</b>	Informática (ATC, CCIA, LSI)		

**1. Situación / Sentido de la Asignatura**

**1.1 Contextualización**

La asignatura de Minería de Datos está programada como asignatura obligatoria en el Semestre 1 del 4º curso de la titulación del Grado en Ingeniería Informática, mención Computación.

La minería de datos es un término relativamente moderno que integra numerosas técnicas de análisis de datos y de extracción de modelos. El objetivo es ser capaces de extraer patrones, de describir tendencias y regularidades, de predecir comportamientos y, en general, utilizar la información computerizada que nos rodea hoy en día, generalmente heterogénea y en grandes cantidades, para permitir a los individuos y a las organizaciones comprender y modelar de una manera más eficiente y precisa el contexto en el que deben actuar y tomar decisiones.



## 1.2 Relación con otras materias

Esta asignatura, tal y como está definida, tiene una relación clara con Técnicas de Aprendizaje Automático, y se asumirá que los alumnos han adquirido las competencias de esa asignatura.

De manera indirecta también tiene relación con Estructuras de Datos, Estadística, Bases de Datos y Programación.

## 1.3 Prerrequisitos

---

- Conocimientos de las técnicas de Aprendizaje Automático estudiadas en la asignatura de Técnicas de Aprendizaje Automático.
- Programación – Pues el alumno necesitará crear sus propios programas o funciones para manipular los datos, u obtener resultados.
- Conocimientos básicos de WEKA.
- Estadística – Como herramienta para poder manipular y condensar los rasgos más relevantes de un conjunto de datos de tamaño grande, ya sea por el número de muestras, o por la dimensionalidad de dichas muestras.





## 2. Competencias

### 2.1 Generales

Código	Descripción
CG4	Capacidad para definir, evaluar y seleccionar plataformas hardware y software para el desarrollo y la ejecución de sistemas, servicios y aplicaciones informáticas, de acuerdo con los conocimientos adquiridos según lo establecido en las competencias de formación especificadas a continuación en esta sección de la memoria.
CG9	Capacidad para resolver problemas con iniciativa, toma de decisiones, autonomía y creatividad. Capacidad para saber comunicar y transmitir los conocimientos, habilidades y destrezas de la profesión de Ingeniero Técnico en Informática.

### 2.2 Específicas

Código	Descripción
CI15	Conocimiento y aplicación de los principios fundamentales y técnicas básicas de los sistemas inteligentes y su aplicación práctica.
CO5	Capacidad para adquirir, obtener, formalizar y representar el conocimiento humano en una forma computable para la resolución de problemas mediante un sistema informático en cualquier ámbito de aplicación, particularmente los relacionados con aspectos de computación, percepción y actuación en ambientes o entornos inteligentes.
CO6	Capacidad para desarrollar y evaluar sistemas interactivos y de presentación de información compleja y su aplicación a la resolución de problemas de diseño de interacción persona computadora.
CO7	Capacidad para conocer y desarrollar técnicas de aprendizaje computacional y diseñar e implementar aplicaciones y sistemas que las utilicen, incluyendo las dedicadas a extracción automática de información y conocimiento a partir de grandes volúmenes de datos.

## 3. Objetivos

Código	Descripción
CO5.1, CI15	Comprender el ámbito de aplicación de las técnicas de minería de datos para la extracción y representación de conocimiento basada en datos.
CO7.1	Saber aplicar técnicas de selección, preparación y visualización de datos a problemas concretos y en dominios de interés práctico.
CO6.1	Emplear adecuadamente técnicas de evaluación y representación de resultados.
CO7.2	Ampliar las técnicas de aprendizaje automático conocidas y saber aplicarlas a problemas de minería de datos.
CO7.3	Usar eficazmente plataformas computacionales para resolver problemas concretos de minería de datos en áreas de interés práctico.



**4. Tabla de dedicación del estudiante a la asignatura**

ACTIVIDADES PRESENCIALES	HORAS	ACTIVIDADES NO PRESENCIALES	HORAS
Clases teórico-prácticas (T/M)	30	Estudio y trabajo autónomo individual	30
Clases prácticas de aula (A)		Estudio y trabajo autónomo grupal	60
Laboratorios (L)	20		
Prácticas externas, clínicas o de campo			
Seminarios (S)	10		
Tutorías grupales (TG)			
Evaluación			
<b>Total presencial</b>	<b>60</b>	<b>Total no presencial</b>	<b>90</b>



5. Bloques temáticos

PARTE TEÓRICA

Bloque 1: Introducción a la Minería de Datos

Carga de trabajo en créditos ECTS: 0,2

a. Contextualización y justificación

Como en toda disciplina en la que se va a iniciar un estudiante, hace falta describir cuál es el contexto de aquello que se le va a enseñar.

La Minería de Datos va a tratar básicamente del proceso de descubrimiento de información útil que se encuentra dentro de los datos que se han ido tomando. Por esto mismo, se ponen en juego técnicas de obtención de datos, de filtrado y adaptación, de extracción de la información, de evaluación y de representación de esa información, todo ello con el fin de que se obtenga alguna utilidad de ella.

Esta secuencia de pasos puede inducir varias metodologías para abordar el problema de la Minería de Datos, pero nunca hay que perder de vista que las técnicas (de procesado, extracción, representación...) y metodologías a desarrollar, van a depender de la naturaleza del problema que se considere.

No sólo deberíamos centrarnos en lo que es la Minería de Datos en sí; también sería deseable que se pensara en lo que se espera obtener de ella, de cuáles son los retos a los que se enfrenta, y los nuevos ámbitos de aplicación a los que se abre (datos biológicos, genoma, multimedia, mercado, redes sociales, ...)

b. Objetivos de aprendizaje

Código	Descripción
CO5.1	Comprender el ámbito de aplicación de las técnicas de minería de datos para la extracción y representación de conocimiento basada en datos.

c. Contenidos

Tema 1: Introducción. Etapas del proceso de minería de datos.

d. Métodos docentes

Ver Anexo: Métodos docentes.

e. Plan de trabajo

Ver cronograma apartado 8. f.

Evaluación

Ver apartado 7.





### g. Bibliografía básica

---

- José Hernández Orallo, M.José Ramírez Quintana, Cèsar Ferri Ramírez. Introducción a la Minería de Datos. Ed. Pearson. 2004.
- Ian H. Witten, Eibe Frank and Mark A. Hall. Data Mining: practical machine learning tools and techniques (third Edition). Morgan Kaufmann, 2011.
- Jiawei Han, Micheline Kamber, Jian Pei. Data Mining – Concepts and Techniques (third edition). Morgan Kaufmann, 2012.

### i. Recursos necesarios

---

Se proporcionarán transparencias de la asignatura. Este material estará disponible en Moodle. Se utilizará este medio también para comunicar al alumno información relativa a la asignatura así como detalles de los trabajos propuestos o publicación de calificaciones parciales.





**Bloque 2: Preprocesamiento y Evaluación**

Carga de trabajo en créditos ECTS: 2.0

**a. Contextualización y justificación**

Por definición la Minería de Datos ha de enfrentarse al problema de extraer información útil de entre una cantidad muy grande de datos. Dichos datos han de prepararse adecuadamente y con rigor científico. En toda esta tarea la Estadística será una de las herramientas básicas en nuestro estudio.

En el primer tema se centra en la preparación y selección de los datos con los que se va a trabajar. Técnicas como la representación gráfica, o el uso de la estadística descriptiva puede ayudar a caracterizar las partes obvias y evidentes de esos datos, con lo que poder centrarnos en aquellos aspectos más ocultos.

A la vista de la naturaleza del problema podremos elegir algún método para extraer información de ellos, pero hay veces en las que los datos con los que vamos a trabajar no tienen el formato que la herramienta puede utilizar. Por ello surge la necesidad de adaptar los datos, ya sea por un simple escalado, o por métodos más complejos como la codificación, discretización. Se estudiarán también los métodos básicos de selección de características y reducción de la dimensionalidad.

Saltándonos el paso fundamental (la extracción del conocimiento), y suponiendo que ya tenemos una de las posibles representaciones del conocimiento, llega el momento de evaluar ese resultado. De nuevo necesitamos de las herramientas estadísticas para poder hacer esto. Nos centraremos en la evaluación de clasificadores, ya conocidos por los alumnos, y en los test de hipótesis.

**b. Objetivos de aprendizaje**

Código	Descripción
CC7.1	Saber aplicar técnicas de selección, preparación y visualización de datos a problemas concretos y en dominios de interés práctico.
CC6.1	Emplear adecuadamente técnicas de evaluación y representación de resultados.
CC7.3	Usar eficazmente plataformas computacionales para resolver problemas concretos de minería de datos en áreas de interés práctico.

**c. Contenidos**

**Tema 2: Preparación y selección de datos.**

**Tema 3: Selección de características y reducción de la dimensionalidad**

**Tema 3: Evaluación de resultados.**

**d. Métodos docentes**

Ver Anexo: Métodos docentes.

**e. Plan de trabajo**

Ver cronograma apartado 8.



**f. Evaluación**

---

Ver apartado 7.

**g. Bibliografía básica**

---

- José Hernández Orallo, M. José Ramírez Quintana, Cèsar Ferri Ramírez. Introducción a la Minería de Datos. Ed. Pearson. 2004.
- Ian H. Witten, Eibe Frank and Mark A. Hall. Data Mining: practical machine learning tools and techniques (third Edition). Morgan Kaufmann, 2011.
- Jiawei Han, Micheline Kamber, Jian Pei. Data Mining – Concepts and Techniques (third edition). Morgan Kaufmann, 2012.

**h. Recursos necesarios**

---

Se proporcionarán transparencias de la asignatura. Este material estará disponible en Moodle. Se utilizará este medio también para comunicar al alumno información relativa a la asignatura así como detalles de los trabajos propuestos o publicación de calificaciones parciales.

**Bloque 3: Extracción de conocimiento**Carga de trabajo en créditos ECTS: 

3.8
-----

**a. Contextualización y justificación**

---

Este bloque trata de describir el funcionamiento y conveniencia de las distintas técnicas de la fase central de la Minería de Datos, así como dar oportunidad de aplicarlas mediante las correspondientes prácticas de laboratorio.

La Extracción de Conocimiento a partir de datos tiene como objetivo descubrir y describir patrones que entre otras cosas deben de ser válidos, novedosos, interesantes y, si es posible, comprensibles. Muchas de las técnicas de Minería de Datos se basan en las del aprendizaje automático y del reconocimiento de formas o patrones

Asumimos que el alumno ha cursado a la asignatura de Aprendizaje Automático y conoce las técnicas de clasificación. Presentaremos los grupos de técnicas que completan los métodos básicos de la Minería de Datos, cada uno de ellos enfocado a un tipo distinto de problema: agrupamiento ("clustering"), la búsqueda eficiente de patrones frecuentes (reglas de asociación), la regresión y los sistemas multiclasificadores básicos.

**b. Objetivos de aprendizaje**

---

Código	Descripción
CC7.2	Ampliar las técnicas de aprendizaje automático conocidas y saber aplicarlas a problemas de minería de datos.
CC7.3	Usar eficazmente plataformas computacionales para resolver problemas concretos de minería de datos en áreas de interés práctico.

**c. Contenidos**

---





- Tema 4: Clustering.
- Tema 5: Reglas de asociación.
- Tema 6: Regresión.
- Tema 7: Multiclasificadores.

**d. Métodos docentes**

Ver Anexo: Métodos docentes.

**e. Plan de trabajo**

Ver cronograma apartado 8. f.

**Evaluación**

Ver apartado 7.

**g. Bibliografía básica**

- José Hernández Orallo, M.José Ramírez Quintana, Cèsar Ferri Ramírez. Introducción a la Minería de Datos. Ed. Pearson. 2004.
- Ian H. Witten, Eibe Frank and Mark A. Hall. Data Mining: practical machine learning tools and techniques (third Edition). Morgan Kaufmann, 2011.
- Jiawei Han, Micheline Kamber, Jian Pei. Data Mining – Concepts and Techniques (third edition). Morgan Kaufmann, 2012.

**h. Recursos necesarios**

Se proporcionarán transparencias de la asignatura. Este material estará disponible en Moodle. Se utilizará este medio también para comunicar al alumno información relativa a la asignatura así como detalles de los trabajos propuestos o publicación de calificaciones parciales.

**6. Temporalización (por bloques temáticos)**

BLOQUE TEMÁTICO	CARGA ECTS	PERIODO PREVISTO DE DESARROLLO
1.- Introducción a la Minería de Datos	0.2	Semana 1
2.- Preprocesamiento y Evaluación	2.8	Semana 2 – 7



3.- Extracción de conocimiento	3.8	Semana8 – 15
--------------------------------	-----	--------------

**7. Tabla resumen de los instrumentos, procedimientos y sistemas de evaluación/calificación**

INSTRUMENTO/PROCEDIMIENTO	PESO EN LA NOTA FINAL	OBSERVACIONES
Los alumnos realizarán y expondrán varios trabajos prácticos.	100%	

El alumno deberá realizar y aprobar todos y cada uno de los trabajos prácticos propuestos para aprobar la asignatura. Todos los trabajos se podrán recuperar en convocatoria extraordinaria.

**8. Consideraciones finales**

**Anexo: Métodos docentes**

- Clase magistral participativa, para la exposición de los puntos básicos del temario y la resolución de problemas.
- Clases de laboratorio. Utilizarán herramientas específicas para la asignatura. El profesor las explicará en las sesiones de laboratorio, dará las indicaciones para resolver cada práctica y resolverá las dudas que vayan surgiendo.
- Aprendizaje basado en problemas. Se utilizará como método docente fundamental para el aprendizaje de los conceptos de la asignatura.