



Asignatura: Data Mining
Código: 16792
Centro: Facultad de Ciencias Económicas y Empresariales
Titulación: Administración y Dirección de Empresas
Nivel: Grado
Tipo: Optativa
Nº. de Créditos: 6

1. ASIGNATURA / COURSE TITLE

DATA MINING

1.1. Código / Course number

16792

1.2. Materia/ Content area

Econometría / Econometrics

1.3. Tipo / Course type

Optativa / Optional

1.4. Nivel / Course level

Grado / Bachelor

1.5. Curso / Year

4º/4th

1.6. Semestre / Semester

Primero / First

1.7. Número de créditos / Credit allotment

6 créditos ECTS / 6 ECTS credits

1.8. Requisitos previos / Prerequisites

Para el correcto seguimiento de la asignatura se recomienda que el alumno esté familiarizado con los conceptos básicos estudiados en las siguientes asignaturas:

- Estadística Descriptiva
- Estadística Teórica

Es conveniente, además, que el alumno tenga un conocimiento suficiente (al nivel de usuario) de *Microsoft Word* (procesador de textos) y alguna



Asignatura: Data Mining
Código: 16792
Centro: Facultad de Ciencias Económicas y Empresariales
Titulación: Administración y Dirección de Empresas
Nivel: Grado
Tipo: Optativa
Nº. de Créditos: 6

experiencia de uso de los paquetes *Ms. Excel* (hoja de cálculo), *MATLAB* y *SPSS*.

1.9. Requisitos mínimos de asistencia a las sesiones presenciales/ **Minimum attendance requirement**

No existe un mínimo de asistencia. No obstante, la asistencia a las actividades presenciales se considera necesaria para la adquisición de competencias e importante para el aprendizaje y la comprensión de conceptos, por lo que es altamente recomendable la asistencia tanto a clases teóricas, prácticas, seminarios, etc.

1.10. Datos del equipo docente / **Faculty data**

El equipo docente está integrado por profesores del siguiente departamento:
/ **The faculty is composed of professors from the following department:**

Departamento de Economía Aplicada. UDI de Estadística

Módulo E-12

Facultad de Ciencias Económicas y Empresariales

C/ Francisco Tomás y Valiente, 5

Universidad Autónoma de Madrid

28049 Madrid

Secretaría: Despacho E-12-311

Tel.: (+34) 91 497 4676

Fax: (+34) 91 497 4676

Web:

http://www.uam.es/ss/Satellite/Economicas/es/1234888141300/subhomeDepartamento/Economia_Aplicada:UDI_de_Estadistica.htm

Los profesores concretos encargados de la docencia de cada asignatura aparecen en los horarios de cada curso académico, disponibles en la siguiente página web: / **The concrete professors in charge of teaching each subject can be seen in each academic course schedule, which is available at the following web page:**

http://www.uam.es/ss/Satellite/Economicas/es/1242650730114/contenidoFiscal/Horarios_y_aulas.htm?idenlace=1242661251796

Los coordinadores de cada asignatura pueden consultarse en la misma página web. / **Each subject coordinator can be seen also at the same web page.**



Asignatura: Data Mining
Código: 16792
Centro: Facultad de Ciencias Económicas y Empresariales
Titulación: Administración y Dirección de Empresas
Nivel: Grado
Tipo: Optativa
Nº. de Créditos: 6

1.11. Objetivos del curso / Course objectives

El objetivo general de esta asignatura es presentar, mediante casos prácticos, técnicas clásicas y modernas de Data Mining (Minería de Datos). En concreto se analizarán técnicas de representación de la información, predicción y clasificación desde una perspectiva aplicada.

Código	Competencias genéricas
CG1	• Capacidad teórica de análisis y síntesis
CG2	• Capacidad de aprendizaje
CG3	• Capacidad creativa para encontrar nuevas ideas y soluciones
CG4	• Capacidad para detectar oportunidades y amenazas
CG5	• Capacidad para rendir bajo presión
CG6	• Capacidad de negociación
CG7	• Capacidad para tomar decisiones
CG8	• Capacidad crítica y de autocrítica
CG9	• Capacidad para trabajar en equipo de carácter interdisciplinar
CG11	• Capacidad de liderazgo: habilidad para convencer, influir y motivar a otros
CG12	• Capacidad para utilizar nuevas herramientas informáticas y de análisis de datos
CG15	• Habilidad para trabajar en un contexto de carácter internacional
CG16	• Habilidad para la búsqueda, identificación y análisis de las fuentes de información pertinentes al ámbito de estudio
CG17	• Habilidades de comunicación a través de internet y, manejo de herramientas multimedia para la comunicación a distancia
CG18	• Habilidades para la presentación en público de trabajos, ideas e informes
CG19	• Iniciativa y espíritu emprendedor
CG20	• Preocupación por la calidad y el trabajo bien hecho
CG23	• Saber gestionar eficazmente el tiempo

Código	Competencias específicas (CE)
CE1	• Adquirir la perspectiva histórica en la comprensión de los fenómenos económicos, jurídicos y sociales que conforman el entorno empresarial
CE2	• Analizar, valorar y sintetizar la complejidad de las situaciones empresariales y su posible evolución a partir de sistemas reales de información
CE4	• Comprender y saber aplicar las herramientas básicas e instrumentos de naturaleza cuantitativa precisas para la obtención, diagnóstico, análisis de la información empresarial y de su entorno económico y social
CE9	• Diferenciar el contexto económico, internacional, nacional, regional y sectorial, que rodea a la empresa así como interpretar su impacto en la misma
CE10	• Diseñar y gestionar proyectos empresariales
CE11	• Dominar el diseño de las estructuras organizativas y saber identificar todas las variables que intervienen en la toma de decisiones empresariales
CE15	• Organización y planificación de los conocimientos adquiridos de forma que configuren una plataforma para la adquisición de conocimientos avanzados para la práctica de la alta dirección empresarial o para la investigación en



Asignatura: Data Mining
Código: 16792
Centro: Facultad de Ciencias Económicas y Empresariales
Titulación: Administración y Dirección de Empresas
Nivel: Grado
Tipo: Optativa
Nº. de Créditos: 6

	el área empresarial
CE21	<ul style="list-style-type: none">Redactar de forma correcta informes y documentos internos y externos empresariales
CE22	<ul style="list-style-type: none">Reunir, analizar, interpretar y presentar los datos procedentes de la investigación de mercados a los que se dirigen las empresas
CE23	<ul style="list-style-type: none">Saber formular, evaluar y seleccionar las estrategias empresariales
CE25	<ul style="list-style-type: none">Tomar decisiones empresariales aplicando diferentes modelos teóricos

En concreto, los resultados del aprendizaje se establecen como la capacidad del alumno para el desarrollo de las siguientes competencias:

- 1) Destacar la capacidad del alumno para desarrollar y aplicar ideas originales en el área de ayuda a la toma de decisiones, utilizando para ello las herramientas y modelos de análisis de datos. Evaluación: resolución de casos prácticos en el área de la Minería de Datos.
- 2) Que el estudiante adquiera la habilidad de integrar conocimientos, afrontar la complejidad, así como formular juicios a partir de información incierta por medio de herramientas de análisis de datos, todo ello a partir de funciones de utilidad socialmente responsable. Para ello, habrán de realizar trabajos en los que el alumno utilice herramientas específicas para la toma de decisiones.
- 3) Que el estudiante sea capaz de continuar su formación en el campo de la toma de decisión, a partir del manejo de publicaciones punteras. Por ello se valorará la lectura y realización de trabajos innovadores que incluyan temáticas novedosas de interés para el estudiante.

El objetivo general de la asignatura es crear en el estudiante una capacidad de aproximación a determinados problemas del mundo económico y empresarial de forma operativa a través del conocimiento avanzado de herramientas estadísticas apropiadas.

La materia tiene una clara vocación empírica, orientada a especificar soluciones de tipo estadístico e informático a problemas reales de la gestión empresarial, a partir de la información disponible.

Esta asignatura permitirá al alumno comprender los elementos de los procedimientos de aprendizaje basados en muestras para el reconocimiento de patrones, siendo capaz de identificar aplicaciones del ámbito profesional susceptibles de ser resueltas satisfactoriamente mediante estas técnicas.

La asignatura de Data Mining se enmarca en el contexto de habilidades y competencias que han de fomentarse en el alumno al final de su formación ya que habrá adquirido experiencia en la utilización de las técnicas de procesamiento de la información y tendrá recursos suficientes para implementar otras modalidades no experimentadas explícitamente.



Asignatura: Data Mining
Código: 16792
Centro: Facultad de Ciencias Económicas y Empresariales
Titulación: Administración y Dirección de Empresas
Nivel: Grado
Tipo: Optativa
Nº. de Créditos: 6

1.12. Contenidos del programa / [Course contents](#)

PROGRAMA SINTÉTICO:

1. Introducción a la Minería de datos.
2. Clasificación Automática: modelos y técnicas
3. Identificación de grupos.
4. Identificación de relaciones.
5. Reducción de la dimensionalidad y representación de la información.

PROGRAMA DETALLADO:

1. Introducción a la Minería de Datos.

- 1.1. Introducción a la metodología CRISP-DM.
- 1.2. Conocimiento del negocio y definición del problema a analizar.
- 1.3. Conocimiento de los datos: búsqueda y captura de datos.
- 1.4. Preparación de los datos: limpieza e integración de diversas bases de datos.
- 1.5. Modelización de los datos: análisis exploratorio de la información.
- 1.6. Evaluación del resultado y revisión del proceso.
- 1.7. Diseño de estrategias de negocio y presentación de resultados.

2. Clasificación Automática: modelos y técnicas.

- 2.1. Clasificación mediante vecinos más cercanos
- 2.2. Clasificación mediante árboles de decisión: principales técnicas
- 2.3. Clasificación mediante reglas de decisión: principales técnicas
- 2.4. Clasificación mediante redes neuronales: principales técnicas.

3. Identificación de grupos.

- 3.1. Clasificación de técnicas de cluster.
- 3.2. Cluster jerárquico.
- 3.3. Cluster no jerárquico.
- 3.4. Cluster Kohonen.

4. Identificación de relaciones.

- 4.1. Concepto de asociación entre atributos
- 4.2. El problema de la cesta de la compra
- 4.3. Diferentes enfoques para la búsqueda de reglas de asociación



Asignatura: Data Mining
Código: 16792
Centro: Facultad de Ciencias Económicas y Empresariales
Titulación: Administración y Dirección de Empresas
Nivel: Grado
Tipo: Optativa
Nº. de Créditos: 6

4.4. Principales técnicas para la búsqueda de reglas de asociación

5. Reducción de la dimensionalidad y representación de la información.

5.1. Análisis de componentes principales

5.2. Análisis de correspondencias

1.13. Referencias de consulta / Course bibliography

General

Título	Learning from data
Autor	V.S. Cherkassky, F. Mulier
Editorial, año	John Wiley & Sons, 1998
Título	Neural Networks: A Comprehensive Foundation
Autor	S. Haykin
Editorial, año	Prentice Hall, 2ª edición, 1998
Título	Pattern Recognition and Machine Learning
Autor	C. Bishop
Editorial, año	Springer, 2006
Título	An Introduction to Support Vector Machines And Other Kernel Based Methods
Autor	N. Cristianini, J. Shawe-Taylor
Editorial, año	Cambridge University Press, 2000
Título	Pattern Classification
Autor	R.O. Duda, P.E. Hart, D.G. Store
Editorial, año	John Wiley & Sons, 2000
Título	Statistical Pattern Recognition
Autor	A.R. Webb
Editorial, año	John Wiley & Sons, 2002
Título	The elements of Statistical Learning: Data Mining, Inference and Prediction
Autor	A. Hastie, R. Tibshirani, J. Friedman
Editorial, año	Springer, 2001
Título	Introducción a la Minería de Datos
Autor	Hernández-Orallo y otros
Editorial, año	Pearson Education, 2004



Asignatura: Data Mining
Código: 16792
Centro: Facultad de Ciencias Económicas y Empresariales
Titulación: Administración y Dirección de Empresas
Nivel: Grado
Tipo: Optativa
Nº. de Créditos: 6

Direcciones web de interés

1. <http://www.ph.tn.tudelft.nl/PRInfo/>
2. <http://cgm.cs.mcgill.ca/~godfried/teaching/pr-web.html>
3. <http://www.aaai.org/AITopics/html/pattern.html>
4. <http://www.iapr.org/>

2. Métodos Docentes / Teaching methodology

Las sesiones presenciales se dividirán en:

- Sesiones teóricas, en las que se presentarán los contenidos teóricos de cada tema. Las clases comenzarán con una breve introducción de los contenidos que se pretenden transmitir en la clase, así como un breve comentario a los conceptos vistos con anterioridad y que sirven de enlace a los que se pretenden desarrollar. El desarrollo de la clase se llevará a cabo con medios audiovisuales, textos y transparencias...que permitan un adecuado nivel de motivación e interés en los alumnos.
- Sesiones prácticas en los laboratorios de informática, donde se pondrán en práctica los conceptos vistos.

Las sesiones de tutorías se organizan:

- Tutorías presenciales destinadas a grupos de alumnos previamente definidos por el profesor. Estas tutorías están destinadas a afianzar los conocimientos, para el seguimiento de los trabajos y la resolución de dudas, así como para la discusión de casos y puesta en común de resultados.
- Tutorías virtuales a través de la plataforma Moodle. En estas se plantearán ejercicios tipo test y su resolución exigirá una preparación previa por parte del alumno.

Como trabajo personal del alumno se deberán realizar una serie de prácticas individuales y/o en grupo, que se irán solicitando a lo largo del curso.



Asignatura: Data Mining
Código: 16792
Centro: Facultad de Ciencias Económicas y Empresariales
Titulación: Administración y Dirección de Empresas
Nivel: Grado
Tipo: Optativa
Nº. de Créditos: 6

3. Tiempo de trabajo del estudiante / Student workload

TAREAS	Nº horas	%
Actividades presenciales	57,0	38,0%
CLASES TEÓRICO-PRÁCTICAS (3 horas/semana x 15 semanas)	45,0	
ASISTENCIA A TUTORIAS	8	
REALIZACIÓN PRUEBAS	4	
Actividades no presenciales (trabajo autónomo del estudiante)	93,0	62,0%
ACTIVIDADES RELACIONADAS: JORNADAS SEMINARIOS,ETC	4	
PREPARACIÓN DE ACTIVIDADES PRÁCTICAS (ejercicios, trabajos, etc.)	40	
ESTUDIO SEMANAL Y PREPARACIÓN DE PRUEBAS	49	
Carga total de horas de trabajo: 25 horas x 6 créditos ECTS	150	100%

4. Métodos de evaluación y porcentaje en la calificación final / Evaluation procedures and weight of components in the final grade

Convocatoria ordinaria

La evaluación se llevará a cabo atendiendo a los siguientes aspectos:

1. Al final de curso se realizará un examen de **carácter obligatorio**. Dicho examen sería una **prueba tipo test** sobre los contenidos de las prácticas realizadas en el curso. El test tendrá una evaluación cuantitativa de 0 a 10 y **debe aprobarse** si se desea aprobar la asignatura. Ponderará respecto a la nota final en un 45 %.
2. Las **prácticas** son obligatorias, deben **entregarse puntualmente para poder aprobar el curso**. Las prácticas se realizarán en grupos de alumnos, cuyo número será determinado por el profesor, y tendrán una evaluación cualitativa: Liberada (cumple con los requisitos establecidos) y No Liberada. Al final del curso, el 35 % de la nota se calculará sobre la proporción de prácticas liberadas respecto del total de prácticas.



Asignatura: Data Mining
Código: 16792
Centro: Facultad de Ciencias Económicas y Empresariales
Titulación: Administración y Dirección de Empresas
Nivel: Grado
Tipo: Optativa
Nº. de Créditos: 6

3. Se consideran **ejercicios** las entregas semanales de las prácticas. Las **entregas son obligatorias** y el alumno recibirá la evaluación cualitativa de liberatoria o no liberatoria. Al final del curso, el 20 % de la nota se calculará sobre la proporción de ejercicios liberatorios respecto del total de ejercicios planteados.

Tanto los casos prácticos como los ejercicios de clase deberán ser entregados para su evaluación en las fechas que se vayan notificando durante el curso, no pudiendo realizarse todos a posteriori de cara a presentarse a la convocatoria extraordinaria.

Convocatoria extraordinaria

La calificación final es resultado de ponderar en un 55% la evaluación de las prácticas y los ejercicios realizados por el estudiante durante el curso y en un 45% la evaluación del examen de la convocatoria extraordinaria.

Alumnos de segunda matrícula

A aquellos alumnos que estén matriculados por segunda vez se les aplicarán los mismos criterios de evaluación que a los de primera matrícula. Como norma general no se guardará ninguna calificación obtenida cuando cursó la asignatura en primera matrícula salvo que la Junta de Facultad establezca algún otro criterio al respecto.

5. Cronograma* / Course calendar

Semana	Contenido	Horas presenciales	Horas no presenciales del estudiante
1	Presentación de la guía docente de la asignatura. Introducción a la Minería de Datos.	3	5
2	Continúa el tema 1. Con el desarrollo de ejemplos y aplicaciones	3	5
3	Tema 2: Clasificación del vecino más cercanos: Introducción, ejemplo y aplicaciones	3	5
4	Tema 2: Clasificación mediante árboles de decisión	3	5
5	Tema 2: Clasificación mediante reglas de decisión: principales técnicas	3	5
6	Prácticas de árboles y reglas de decisión. Planteamiento de casos prácticos similares a los explicados en clase para su resolución por los alumnos	3	9



Asignatura: Data Mining
Código: 16792
Centro: Facultad de Ciencias Económicas y Empresariales
Titulación: Administración y Dirección de Empresas
Nivel: Grado
Tipo: Optativa
Nº. de Créditos: 6

Semana	Contenido	Horas presenciales	Horas no presenciales del estudiante
7	Tema 2: Clasificación mediante Redes Neuronales	6	9
8	Prácticas de clasificación, estimación y predicción con Redes Neuronales. Planteamiento de casos prácticos similares a los explicados en clase para su resolución por los alumnos	6	9
9	Tema 3: Identificación de grupos	3	5
10	Prácticas de Análisis de Conglomerados. Planteamiento de casos prácticos similares a los explicados en clase para su resolución por los alumnos	5	5
11	Tema 4: Identificación de relaciones, principios y aplicaciones	3	5
12	Tema 4: Diferentes enfoques y técnicas para la búsqueda de reglas de asociación	4	5
13	Prácticas de reglas de asociación. Planteamiento de casos prácticos similares a los explicados en clase para su resolución por los alumnos	4	7
14	Tema 5: Reducción de la Dimensionalidad y representación de la información	4	7
15	Prácticas de Componentes Principales y Análisis de Correspondencias. Planteamiento de casos prácticos similares a los explicados en clase para su resolución por los alumnos	4	7
	Total Horas	57	93

*Este cronograma tiene carácter orientativo.