



# Minería de Datos

## Guía de Aprendizaje – Información al estudiante

### 1. Datos Descriptivos

<b>Asignatura</b>	Minería de Datos
<b>Materia</b>	--
<b>Departamento responsable</b>	Inteligencia Artificial
<b>Créditos ECTS</b>	3
<b>Carácter</b>	Optativa
<b>Titulación</b>	Grado en Ingeniería Informática
<b>Curso</b>	4º
<b>Especialidad</b>	No aplica

<b>Curso académico</b>	2014-2015
<b>Semestre en que se imparte</b>	Semestre 2
<b>Semestre principal</b>	
<b>Idioma en que se imparte</b>	Español
<b>Página Web</b>	<a href="http://www.dia.fi.upm.es/">http://www.dia.fi.upm.es/</a>



## 2. Profesorado

NOMBRE Y APELLIDO	DESPACHO	Correo electrónico
Pedro Larrañaga (Coord.)	2208	pedro.larranaga@fi.upm.es
Concepción Bielza Lozoya	2210	mcbielza@fi.upm.es

## 3. Conocimientos previos requeridos para poder seguir con normalidad la asignatura

<b>Asignaturas superadas</b>	<ul style="list-style-type: none"><li>• Probabilidades y Estadística</li><li>•</li></ul>
<b>Otros resultados de aprendizaje necesarios</b>	<ul style="list-style-type: none"><li>•</li></ul>

## 4. Objetivos de Aprendizaje

<b>COMPETENCIAS ASIGNADAS A LA ASIGNATURA Y SU NIVEL DE ADQUISICIÓN</b>		
<b>Código</b>	<b>Competencia</b>	<b>Nivel</b>
CG1	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	3
CG3	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	4
CG4	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	2
CG12	Comprensión amplia de las técnicas y métodos aplicables en una especialización concreta, así como de sus límites	2
CG13	Apreciación de los límites del conocimiento actual y de la aplicación práctica de la tecnología más reciente	2
CGI2	Comprender el procedimiento, valor y límites del método científico en el campo de la Informática, siendo capaz de identificar, localizar y obtener datos requeridos en un trabajo de investigación, de diseñar y guiar investigaciones analíticas, de modelado y experimentales, así como de evaluar datos de una manera crítica y extraer conclusiones	3
CEIA2	Capacidad de conectar la tecnología puntera en Inteligencia Artificial con las necesidades de los clientes	3
CEIA4	Capacidad de interpretar los modelos de clasificación supervisada y no supervisada obtenidos al aplicar las técnicas de Aprendizaje Automático para un conjunto de datos	4
CEIA10	Identificación de áreas de aplicación en las que se pueda utilizar las técnicas y métodos de la Inteligencia Artificial	1

LEYENDA: Nivel de adquisición 1: Conocimiento  
 Nivel de adquisición 2: Comprensión  
 Nivel de adquisición 3: Aplicación  
 Nivel de adquisición 4: Análisis y síntesis



<b>RESULTADOS DE APRENDIZAJE DE LA ASIGNATURA</b>			
<b>Código</b>	<b>Resultado de aprendizaje</b>	<b>Competencias asociadas</b>	<b>Nivel de adquisición</b>
RA1	Ser capaz de modelizar problemas reales de clasificación mediante paradigmas computacionales.	CEIA4, CG1	1
RA2	Ser capaz de exponer los modelos y resultados obtenidos, tanto de manera oral como escrita, de manera estructurada, aportando fuentes documentales, limitaciones del estudio, diferentes perspectivas sobre la bondad del modelo e ideas futuras.	CG3, CG4, CGI2	4
RA3	Ser capaz de utilizar software actual y entender las hipótesis asumidas al aplicarse sobre diferentes conjuntos de datos y con diferentes metodologías.	CG12, CEIA2, CEIA10	3
RA4	Ser capaz de manejar fuentes bibliográficas en el área, incluyendo manuales, documentación online, y artículos científicos.	CG13	4

## 5. Sistema de evaluación de la asignatura

INDICADORES DE LOGRO		
Ref	Indicador	Relacionado con RA
I1	Redactar un trabajo escrito sobre una aplicación práctica y exponerlo oralmente.	RA2, RA4
I2	Utilizar algoritmos de clasificación supervisada.	RA1, RA3
I3	Utilizar algoritmos de clasificación no supervisada.	RA1, RA3

EVALUACION SUMATIVA			
Breve descripción de las actividades evaluables	Momento	Lugar	Peso en la calif.
Realizar individualmente un trabajo escrito y exponerlo oralmente sobre la temática de clasificación supervisada, utilizando el software WEKA	Semana 13	Biblioteca y Aula	60%
Realizar individualmente un trabajo escrito y exponerlo oralmente sobre la temática de clasificación no supervisada, utilizando el software WEKA	Semana 16	Biblioteca y Aula	40%
<b>Total:</b>			<b>100%</b>



## CRITERIOS DE CALIFICACIÓN

La calificación vendrá dada a partir de la evaluación de:

1. El trabajo escrito sobre clasificación supervisada (temas 2-4) proporcionado por el alumno y su exposición oral en el aula. Se utilizará el software WEKA. Este trabajo tendrá un peso del 60%.
2. El trabajo escrito sobre clasificación no supervisada (tema 5) proporcionado por el alumno y su exposición oral en el aula. Se utilizará el software WEKA. Este trabajo tendrá un peso del 40%.

Estos pesos se utilizan para la nota final siempre que el alumno se presente a ambas partes y obtenga una nota mínima de 3 sobre 10 en cada parte. Si un alumno no se presenta a una de las dos partes, aparecerá como suspenso en la nota final y deberá entregar la parte no presentada en la convocatoria de Julio. Si habiéndose presentado a ambas partes una de ellas no superara el mínimo de 3 (su nota es inferior a 3), se presentará de nuevo a esta parte en Julio. Si, en cambio, ambas son superiores o iguales a 3, pero la nota final no resulta un aprobado (mayor o igual que 5 sobre 10), el alumno tendrá que presentarse en Julio a una de las partes o a las dos, lo que él decida.

No podrá haber dos alumnos que utilicen los mismos datos y con el mismo objetivo.

Está terminantemente prohibido copiar o dejar copiar. Su detección implicará el suspenso inmediato y la repetición de los trabajos con datos diferentes.

Para la convocatoria extraordinaria de Julio las evaluaciones serán análogas (dos exposiciones orales y sus respectivos trabajos escritos realizados de forma individual). Si el estudiante decide realizar sólo una prueba final, se efectuará en la fecha de examen final fijada por jefatura de estudios y consistirá (como en la evaluación continua) en las dos exposiciones orales y sus respectivos trabajos escritos realizados de forma individual. La solicitud para hacer sólo una prueba final se tiene que realizar por email al coordinador de la asignatura antes de la primera entrega (semana 13).

## 6. Contenidos y Actividades de Aprendizaje

CONTENIDOS ESPECÍFICOS		
Bloque / Tema / Capítulo	Apartado	Indicadores Relacionados
<b>Tema 1: Introducción a la ingeniería de datos</b>	1.1 Reconocimiento de patrones	I1, I2, I3
	1.2 Preproceso de datos	I1
<b>Tema 2: Generalidades sobre clasificación supervisada</b>	2.1 Métodos de evaluación	I1, I2
	2.2 Selección de variables	I1, I2
<b>Tema 3: Clasificadores no probabilísticos</b>	3.1 Vecinos más cercanos	I1, I2
	3.2 Árboles de clasificación	I1, I2
	3.3 Inducción de reglas	I1, I2
	3.4 Máquinas de vector soporte	I1, I2
	3.5 Redes neuronales	I1, I2
<b>Tema 4: Clasificadores probabilísticos</b>	4.1 Análisis discriminante	I1, I2
	4.2 Regresión logística	I1, I2
	4.3 Clasificadores bayesianos	I1, I2
	4.4 Metaclasificadores	I1, I2
<b>Tema 5: Clasificación no supervisada</b>	5.1 Clasificación ascendente jerárquica	I1, I3
	5.2 Métodos particionales	I1, I3
	5.3 Clustering probabilista	I1, I3

## 7. Breve descripción de las modalidades organizativas utilizadas y de los métodos de enseñanza empleados

Tabla 7. Modalidades organizativas de la enseñanza








MODALIDADES ORGANIZATIVAS DE LA ENSEÑANZA		
Escenario	Modalidad	Finalidad
	Clases Teóricas	<i>Hablar a los estudiantes</i>
	Seminarios-Talleres	<i>Construir conocimiento a través de la interacción y la actividad de los estudiantes</i>
	Clases Prácticas	<i>Mostrar a los estudiantes cómo deben actuar</i>
	Prácticas Externas	<i>Completar la formación de los alumnos en un contexto profesional</i>
	Tutorías	<i>Atención personalizada a los estudiantes</i>
	Trabajo en grupo	<i>Hacer que los estudiantes aprendan entre ellos</i>
	Trabajo autónomo	<i>Desarrollar la capacidad de autoaprendizaje</i>



Tabla 8. Métodos de enseñanza

MÉTODOS DE ENSEÑANZA		
	Método	Finalidad
	Método Expositivo/Lección Magistral	Transmitir conocimientos y activar procesos cognitivos en el estudiante
	Estudio de Casos	Adquisición de aprendizajes mediante el análisis de casos reales o simulados
	Resolución de Ejercicios y Problemas	Ejercitar, ensayar y poner en práctica los conocimientos previos
	Aprendizaje Basado en Problemas (ABP)	Desarrollar aprendizajes activos a través de la resolución de problemas
	Aprendizaje orientado a Proyectos	Realización de un proyecto para la resolución de un problema, aplicando habilidades y conocimientos adquiridos
	Aprendizaje Cooperativo	Desarrollar aprendizajes activos y significativos de forma cooperativa
	Contrato de Aprendizaje	Desarrollar el aprendizaje autónomo

Se conoce como método expositivo "la presentación de un tema lógicamente estructurado con la finalidad de facilitar información organizada siguiendo criterios adecuados a la finalidad pretendida". Esta metodología -también conocida como lección (lecture)- se centra fundamentalmente en la exposición verbal por parte del profesor de los contenidos sobre la materia objeto de estudio. El término "lección magistral" se suele utilizar para denominar un tipo específico de lección impartida por un profesor en ocasiones especiales.

Análisis intensivo y completo de un hecho, problema o suceso real con la finalidad de conocerlo, interpretarlo, resolverlo, generar hipótesis, contrastar datos, reflexionar, completar conocimientos, diagnosticarlo y, en ocasiones, entrenarse en los posibles procedimientos alternativos de solución.

Situaciones en las que se solicita a los estudiantes que desarrollen las soluciones adecuadas o correctas mediante la ejercitación de rutinas, la aplicación de fórmulas o algoritmos, la aplicación de procedimientos de transformación de la información disponible y la interpretación de los resultados. Se suele utilizar como complemento de la lección magistral.

Método de enseñanza-aprendizaje cuyo punto de partida es un problema que, diseñado por el profesor, el estudiante ha de resolver para desarrollar determinadas competencias previamente definidas.

Método de enseñanza-aprendizaje en el que los estudiantes llevan a cabo la realización de un proyecto en un tiempo determinado para resolver un problema o abordar una tarea mediante la planificación, diseño y realización de una serie de actividades, y todo ello a partir del desarrollo y aplicación de aprendizajes adquiridos y del uso efectivo de recursos.

Enfoque interactivo de organización del trabajo en el aula en el cual los alumnos son responsables de su aprendizaje y del de sus compañeros en una estrategia de corresponsabilidad para alcanzar metas e incentivos grupales. Es tanto un método, a utilizar entre otros, como un enfoque global de la enseñanza, una filosofía.

Un acuerdo establecido entre el profesor y el estudiante para la consecución de unos aprendizajes a través de una propuesta de trabajo autónomo, con una supervisión por parte del profesor y durante un período determinado. En el contrato de aprendizaje es básico un acuerdo formalizado, una relación de contraprestación recíproca, una implicación personal y un marco temporal de ejecución.



### **BREVE DESCRIPCIÓN DE LAS MODALIDADES ORGANIZATIVAS UTILIZADAS Y METODOS DE ENSEÑANZA EMPLEADOS**

<b>CLASES DE TEORIA</b>	Durante una clase de teoría o lección magistral, el profesor realiza una exposición verbal de los contenidos sobre la materia objeto de estudio, mediante la cual suministra a los alumnos información esencial y organizada procedente de diversas fuentes con unos objetivos específicos predefinidos (motivar al alumno, exponer los contenidos sobre un tema, explicar conocimientos, efectuar demostraciones teóricas, presentar experiencias, etc.) pudiendo utilizar para ello, además de la exposición oral, otros recursos didácticos (audiovisuales, documentos, etc.).
<b>CLASES DE PROBLEMAS</b>	...
<b>PRÁCTICAS</b>	El profesor plantea un supuesto práctico para cuya solución el alumno ha de aplicar los conocimientos adquiridos en las clases de teoría y las sesiones de trabajo personal.
<b>TRABAJOS AUTONOMOS</b>	El alumno aprenderá a trabajar de forma autónoma y autodirigida con el fin de preparar tanto las presentaciones orales a realizar como la elaboración de un trabajo escrito. En este tipo de trabajo autónomo, el alumno recopilará información de fuentes bibliográficas y/o Internet, estudiará posibles ideas novedosas sobre investigación futura y preparará la presentación y redacción del trabajo escrito.
<b>TRABAJOS EN GRUPO</b>	...
<b>TUTORÍAS</b>	Los alumnos, de forma individual o en grupo, podrán solicitar al profesor tantas sesiones de tutorías como precisen con el fin de alcanzar el máximo grado de aprovechamiento en la asignatura

## 8. Recursos didácticos

RECURSOS DIDÁCTICOS	
<b>BIBLIOGRAFÍA</b>	<ul style="list-style-type: none"><li>• E. Alpaydin. Introduction to Machine Learning. MIT Press. 2004.</li><li>• R. Duda, P.E. Hart, D.G. Stork. Pattern Classification. Wiley. 2001.</li><li>• J. Hernández-Orallo, M.J. Ramírez, C. Ferri. Introducción a la Minería de Datos. Pearson Educación. 2004.</li><li>• L. Kuncheva. Combining Pattern Classifiers. Wiley. 2004.</li><li>• L. Méndez del Río. Más Allá del <i>Business Intelligence</i>. 16 Experiencias de Éxito. Ediciones Gestión 2000. 2006.</li><li>• A. Webb. Statistical Pattern Recognition. Wiley. 2002.</li><li>• I. Witten, E. Frank, M. Hall. Data Mining. Morgan Kaufmann. 3ª ed. 2011.</li></ul>
<b>RECURSOS WEB</b>	Software WEKA: <a href="http://www.cs.waikato.ac.nz/ml/weka/">http://www.cs.waikato.ac.nz/ml/weka/</a> Repositorios de datos: <ul style="list-style-type: none"><li>• <a href="http://archive.ics.uci.edu/ml/">http://archive.ics.uci.edu/ml/</a></li><li>• <a href="http://www.stats4stem.org/data-sets.html">http://www.stats4stem.org/data-sets.html</a></li><li>• <a href="http://www.inside-r.org/howto/finding-data-internet">http://www.inside-r.org/howto/finding-data-internet</a></li></ul>
	Sitio Moodle de la asignatura
<b>EQUIPAMIENTO</b>	Aula Asignada Biblioteca



### 9. Cronograma de trabajo de la asignatura

Semana	Actividades en Aula	Actividades en Laboratorio	Trabajo Individual	Trabajo en Grupo	Actividades de Evaluación	Otros
Semana 1 (2 horas)	• Clases teóricas del tema 1 (2 horas)	•	•	•	•	•
Semana 2 (3 horas)	• Clases teóricas del tema 2.1 (2 horas)	•	• Estudio tema 1 (1 hora)	•	•	•
Semana 3 (4 horas)	• Clases teóricas del tema 2.2 y 3.1 (2 horas)	•	• Estudio tema 2 (2 horas)	•	•	•
Semana 4 (6 horas)	• Clases teóricas del tema 3.2 (2 horas)	•	• Estudio tema 2 (3 horas) • Elaboración de trabajo individual (1 hora)	•	•	
Semana 5 (8 horas)	• Clases teóricas del tema 3.3 y 3.4 (2 horas)	•	• Estudio tema 3 (2 horas) • Elaboración de trabajo individual (2 horas)	• Tutorías en grupo (2 horas)	•	
Semana 6 (6 horas)	• Clases teóricas del tema 3.5 (2 horas)	•	• Estudio tema 3 (2 horas) • Elaboración de trabajo individual (2 horas)	•	•	•
Semana 7 (4 horas)	• Clases teóricas del tema 4.1 (2 horas)	•	• Elaboración de trabajo individual (2 horas)	•	•	
Semana 8 (7 horas)	• Clases teóricas del tema 4.2 (2 horas)	•	• Estudio tema 4 (1 hora) • Elaboración de trabajo individual (2 horas)	• Tutorías en grupo (2 horas)	•	•
Semana 9 (7 horas)	• Clases teóricas del tema 4.3 (2 horas)	•	• Estudio tema 4 (3 horas) • Elaboración de trabajo individual (2 horas)	•	•	
Semana 10 (5 horas)	• Clases teóricas del tema 4.4 (2 horas)	•	• Estudio tema 4 (3 horas)	•	•	•
Semana 11 (4 horas)	•	•	• Elaboración de trabajo individual (2 horas)	• Tutorías en grupo (2 horas)	•	•



Semana 12 (4 horas)	• Clases teóricas del tema 5.1 (2 horas)	•	• Elaboración de trabajo individual (2 horas)	•	•	•
Semana13 (4 horas)	•	•	• Estudio tema 5 (2 horas)	•	• Exposición oral individual y evaluación (2 horas)	•
Semana14 (7 horas)	• Clases teóricas del tema 5.2 (2 horas)	•	• Estudio tema 5 (3 horas) • Elaboración de trabajo individual (2 horas)	•	•	•
Semana15 (7 horas)	• Clases teóricas del tema 5.3 (2 horas)	•	• Estudio tema 5 (2 horas) • Elaboración de trabajo individual (1 horas)	• Tutorías en grupo (2 horas)	•	•
Semana 16 (3 horas)	•	•	• Elaboración de trabajo individual (1 horas)	•	• Exposición oral individual y evaluación (2 horas)	•
TOTAL	26 horas	•	43 horas	8 horas	4 horas	•

Nota: Para cada actividad se especifica la dedicación en horas que implica para el alumno. Esta distribución de esfuerzos debe entenderse para el "estudiante medio", por lo que si bien puede servir de orientación, no debe tomarse en ningún caso en sentido estricto a la hora de planificar su trabajo. Cada alumno deberá hacer su propia planificación para alcanzar los resultados de aprendizaje descritos en esta Guía y ajustar dicha planificación en un proceso iterativo en función de los resultados intermedios que vaya obteniendo



**POLITÉCNICA**



UNIVERSIDAD POLITÉCNICA DE MADRID  
**FACULTAD DE INFORMÁTICA**  
Campus de Montegancedo  
Boadilla del Monte. 28660 Madrid