



POLITÉCNICA

CAMPUS  
DE EXCELENCIA  
INTERNACIONAL

PROCESO DE  
COORDINACIÓN DE LAS  
ENSEÑANZAS PR/CL/001



E.T.S. de Ingenieros  
Informaticos

# ANX-PR/CL/001-01

## GUÍA DE APRENDIZAJE

### ASIGNATURA

**105000169 - Minería de datos**

### PLAN DE ESTUDIOS

10II - Grado en Ingeniería Informática

### CURSO ACADÉMICO Y SEMESTRE

2017-18 - Segundo semestre

## Índice

---

### Guía de Aprendizaje

1. Datos descriptivos .....	1
2. Profesorado .....	1
3. Conocimientos previos recomendados .....	2
4. Competencias y resultados de aprendizaje .....	2
5. Descripción de la asignatura y temario .....	4
6. Cronograma .....	6
7. Actividades y criterios de evaluación .....	8
8. Recursos didácticos .....	10

## 1. Datos descriptivos

---

### 1.1 Datos de la asignatura

<b>Nombre de la Asignatura</b>	105000169 - Minería de datos
<b>Nº de Créditos</b>	3 ECTS
<b>Carácter</b>	105000169
<b>Curso</b>	Cuarto curso
<b>Semestre</b>	Octavo semestre
<b>Período de impartición</b>	Febrero-Junio
<b>Idioma de impartición</b>	Castellano
<b>Titulación</b>	10II - Grado en Ingeniería Informática
<b>Centro en el que se imparte</b>	Escuela Técnica Superior de Ingenieros Informaticos
<b>Curso Académico</b>	2017-18

## 2. Profesorado

---

### 2.1 Profesorado implicado en la docencia

Nombre	Despacho	Correo electrónico	Horario de tutorías*
Emilio Serrano Fernandez	2201	emilio.serrano@upm.es	L - 15:00 - 16:00 M - 11:00 - 13:00 X - 15:00 - 16:00
Pedro Maria Larrañaga Mugica (Coordinador/a)	2208	pedro.larranaga@upm.es	M - 15:00 - 19:00 V - 15:00 - 17:00
Maria Concepcion Bielza Lozoya	2210	c.bielza@upm.es	M - 15:00 - 19:00 V - 15:00 - 17:00

\* Las horas de tutoría son orientativas y pueden sufrir modificaciones. Se deberá confirmar los horarios de tutorías con el profesorado.

## 3. Conocimientos previos recomendados

---

### 3.1 Asignaturas previas que se recomienda haber cursado

- Probabilidades y estadística I

### 3.2 Otros conocimientos previos recomendados para cursar la asignatura

El plan de estudios Grado en Ingeniería Informática no tiene definidos otros conocimientos previos para esta asignatura.

## 4. Competencias y resultados de aprendizaje

---

### 4.1 Competencias que adquiere el estudiante al cursar la asignatura

CG-1/21 - Capacidad de resolución de problemas aplicando conocimientos de matemáticas, ciencias e ingeniería.

CG-13/CE55 - Capacidad de comunicarse de forma efectiva con los compañeros, usuarios (potenciales) y el público en general acerca de cuestiones reales y problemas relacionados con la especialización elegida.

CG-19 - Capacidad de usar las tecnologías de la información y la comunicación.

CG-2/CE45 - Capacidad para el aprendizaje autónomo y la actualización de conocimientos, y reconocimiento de su necesidad en el área de la informática.

CG-5 - Capacidad de gestión de la información.

CG-6 - Capacidad de abstracción, análisis y síntesis

Ce 12/16 - Conocer los campos de aplicación de la informática, y tener una apreciación de la necesidad de poseer unos conocimientos técnicos profundos en ciertas áreas de aplicación; apreciación del grado de esta necesidad en, por lo menos, una situación.

Ce 14/15 - Conocer el software, el hardware y las aplicaciones existentes en el mercado, así como el uso de sus elementos, y capacidad para familiarizarse con nuevas aplicaciones informáticas.

Ce 17 - Conocer los temas informáticos avanzados de modo que permita a los alumnos vislumbrar y entender las fronteras de la disciplina, por medio de la inclusión de experiencias de aprendizaje que dirigen a los alumnos desde los temas elementales a los temas avanzados o los temas de los que se nutren los novísimos desarrollos.

Ce 44 - Conocimiento de tecnologías punteras relevantes y su aplicación.

## 4.2 Resultados del aprendizaje al cursar la asignatura

RA277 - Dado un problema real elegir la tecnología informática existente en el mercado mas apropiada para su solución y diseñar su desarrollo e integración, analizando la viabilidad de su solución, lo que se puede y no se puede conseguir a través del estado actual de desarrollo de la tecnología usada, y lo que se espera que avance en el futuro.

RA278 - Desarrollar la solución matemática y algorítmica mas apropiada a un problema informático que requiera un tratamiento especialmente complejo, analizando y exponiendo su viabilidad.

RA280 - Obtención de las competencias lingüísticas comunicativas (comprensión, expresión, etc.) habladas y escritas en entornos académicos/profesionales nacionales/internacionales.

RA281 - Obtención de las técnicas necesarias para la realización de un informe o memoria sobre un trabajo realizado en un entorno socio-lingüístico nacional/internacional.

## 5. Descripción de la asignatura y temario

---

### 5.1 Descripción de la asignatura

La Minería de Datos trata de construir sistemas informáticos cuando no existe experiencia humana o cuando no es fácilmente explicable. Así, los datos se transforman en conocimiento. Entre las muchas aplicaciones exitosas pueden citarse el reconocimiento del habla o de texto manuscrito, navegación autónoma de robots, recuperación de información documental, filtrado cooperativo, sistemas de diagnóstico, análisis de microarrays de ADN, etc.

Este curso expone varios métodos de Minería de Datos desde el punto de vista práctico. El objetivo es enfrentarse a un conjunto de datos sobre los que construir modelos clasificatorios con la ayuda del software WEKA.

### 5.2 Temario de la asignatura

1. Introducción a la minería de datos
  - 1.1. Reconocimiento de patrones
  - 1.2. Preproceso de datos
2. Clasificación supervisada
  - 2.1. Métodos de evaluación
  - 2.2. Vecinos más cercanos
  - 2.3. Clasificadores Bayesianos
  - 2.4. Regresión logística
  - 2.5. Árboles de clasificación
  - 2.6. Inducción de reglas
  - 2.7. Redes neuronales
  - 2.8. Selección de variables
  - 2.9. Metaclasificadores
3. Clasificación no supervisada
  - 3.1. Métodos particionales
  - 3.2. Clasificación ascendente jerárquica

### 3.3. Clustering probabilista

## 6. Cronograma

### 6.1 Cronograma de la asignatura\*

Semana	Actividad Presencial en Aula	Actividad Presencial en Laboratorio	Otra Actividad Presencial	Actividades de Evaluación
1	<b>Clases teóricas del tema 1</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral			
2	<b>Clases teóricas del tema 1 y 2.1</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral			
3		<b>Clases prácticas del tema 1 y 2.1</b> Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio		
4	<b>Clases teóricas del tema 2.2 y 2.3</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral			
5		<b>Clases prácticas del tema 2.2 y 2.3</b> Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio		
6	<b>Clases teóricas del tema 2.4 y 2.5</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral			
7		<b>Clases prácticas del tema 2.4 y 2.5</b> Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio		
8	<b>Clases teóricas del tema 2.6 y 2.7</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral			
9		<b>Clases prácticas del tema 2.6 y 2.7</b> Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio		
10	<b>Clases teóricas del tema 2.8 y 2.9</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral			
11		<b>Clases prácticas del tema 2.8 y 2.9</b> Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio		
12				<b>Entrega y exposición oral del trabajo individual sobre clasificación supervisada</b> PI: Técnica del tipo Presentación Individual Evaluación continua Duración: 02:00



13	<b>Clases teóricas del tema 3</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral			
14		<b>Clases prácticas del tema 3</b> Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio		
15				<b>Entrega y exposición oral del trabajo individual sobre clasificación no supervisada</b> PI: Técnica del tipo Presentación IndividualEvaluación continua Duración: 02:00
16				<b>Examen tipo test sobre los contenidos de la asignatura</b> EX: Técnica del tipo Examen EscritoEvaluación continua Duración: 01:00
17				<b>Entrega y exposición oral del trabajo individual sobre clasificación supervisada y no supervisada, así como examen tipo test</b> PI: Técnica del tipo Presentación IndividualEvaluación sólo prueba final Duración: 02:00

\* El cronograma sigue una planificación teórica de la asignatura y puede sufrir modificaciones durante el curso.

## 7. Actividades y criterios de evaluación

### 7.1 Actividades de evaluación de la asignatura

#### 7.1.1 Evaluación continua

Sem.	Descripción	Modalidad	Tipo	Duración	Peso en la nota	Nota mínima	Competencias evaluadas
12	Entrega y exposición oral del trabajo individual sobre clasificación supervisada	PI: Técnica del tipo Presentación Individual	Presencial	02:00	35%	3 / 10	CG-1/21 CG-6 CG-13/CE55 Ce 12/16 Ce 14/15 Ce 17 Ce 44
15	Entrega y exposición oral del trabajo individual sobre clasificación no supervisada	PI: Técnica del tipo Presentación Individual	Presencial	02:00	25%	3 / 10	CG-5 CG-2/CE45 CG-6 CG-13/CE55
16	Examen tipo test sobre los contenidos de la asignatura	EX: Técnica del tipo Examen Escrito	Presencial	01:00	40%	5 / 10	CG-1/21 Ce 12/16 Ce 44

#### 7.1.2 Evaluación sólo prueba final

Sem.	Descripción	Modalidad	Tipo	Duración	Peso en la nota	Nota mínima	Competencias evaluadas
17	Entrega y exposición oral del trabajo individual sobre clasificación supervisada y no supervisada, así como examen tipo test	PI: Técnica del tipo Presentación Individual	Presencial	02:00	100%	5 / 10	CG-1/21 CG-5 CG-2/CE45 CG-6 CG-13/CE55 Ce 12/16 Ce 14/15 Ce 17 Ce 44

#### 7.1.3 Evaluación convocatoria extraordinaria

No se ha definido la evaluación extraordinaria.

## 7.2 Criterios de Evaluación

La calificación vendrá dada a partir de la evaluación de:

1. El trabajo escrito sobre clasificación supervisada (tema 2) proporcionado por el alumno y su exposición oral en el aula. Se utilizará el software WEKA. Este trabajo tendrá un peso del 35%.
2. El trabajo escrito sobre clasificación no supervisada (tema 3) proporcionado por el alumno y su exposición oral en el aula. Se utilizará el software WEKA. Este trabajo tendrá un peso del 25%.
3. El examen tipo test sobre los contenidos de la asignatura. Tendrá un peso del 40%.

Estos pesos se utilizan para la nota final siempre que el alumno se presente a las tres partes y obtenga una nota mínima de 3 sobre 10 en la primera y segunda parte, y de un 5 sobre 10 en la tercera parte. Si un alumno no se presenta a una de las tres partes, deberá presentar esta parte en la prueba final de esta convocatoria.

Si no se superan los mínimos previamente mencionados en alguna de las tres partes, deberá presentarse de nuevo a esa(s) parte(s) en Julio. Si, en cambio, todas superan los mínimos establecidos, pero la nota final no resulta un aprobado (mayor o igual que 5 sobre 10), el alumno tendrá que presentarse en Julio a cualquiera de las partes (una, dos o las tres), lo que él decida, teniendo en cuenta que para aprobar la nota final ha de ser mayor o igual a 5 sobre 10.

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, y en el examen de tipo test, evaluándose con los mismos pesos que

en la evaluación continua. La solicitud para hacer sólo una prueba final se tiene que realizar por email al coordinador de la asignatura antes de la primera entrega.

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, junto con el examen de tipo test).

## 8. Recursos didácticos

---

### 8.1 Recursos didácticos de la asignatura

Nombre	Tipo	Observaciones
Libro 1	Bibliografía	E. Alpaydin. Introduction to Machine Learning. MIT Press
Libro 3	Bibliografía	J. Hernández-Orallo, M.J. Ramírez, C. Ferri. Introducción a la Minería de Datos. Pearson Educación. 2004.
Libro 4	Bibliografía	L. Kuncheva. Combining Pattern Classifiers. Wiley. 2004
Libro 7	Bibliografía	I. Witten, E. Frank. Data Mining. Morgan Kaufmann. 2ª ed. 2005.
software	Recursos web	Software WEKA: <a href="http://www.cs.waikato.ac.nz/ml/weka">http://www.cs.waikato.ac.nz/ml/weka</a>
Repositorios de datos	Recursos web	- <a href="http://www.inside-r.org/howto/finding-data-internet">http://www.inside-r.org/howto/finding-data-internet</a> - <a href="https://www.quora.com/Where-can-I-find-large-datasets-open-to-the-public">https://www.quora.com/Where-can-I-find-large-datasets-open-to-the-public</a>