



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

**103000391 - Descubrimiento de conocimiento en bases de datos**

### PLAN DE ESTUDIOS

10AK - Master Universitario en Software y Sistemas

### CURSO ACADÉMICO Y SEMESTRE

2017-18 - Primer semestre

## Índice

---

### Guía de Aprendizaje

1. Datos descriptivos .....	1
2. Profesorado .....	1
3. Competencias y resultados de aprendizaje .....	2
4. Descripción de la asignatura y temario .....	3
5. Cronograma .....	5
6. Actividades y criterios de evaluación .....	8
7. Recursos didácticos .....	10
8. Otra información .....	11

## 1. Datos descriptivos

---

### 1.1 Datos de la asignatura

<b>Nombre de la Asignatura</b>	103000391 - Descubrimiento de conocimiento en bases de datos
<b>Nº de Créditos</b>	4 ECTS
<b>Carácter</b>	Optativa
<b>Curso</b>	Primer curso
<b>Semestre</b>	Primer semestre
<b>Período de impartición</b>	Septiembre-Enero
<b>Idioma de impartición</b>	Castellano
<b>Titulación</b>	10AK - Master Universitario en Software y Sistemas
<b>Centro en el que se imparte</b>	Escuela Tecnica Superior de Ingenieros Informaticos
<b>Curso Académico</b>	2017-18

## 2. Profesorado

---

### 2.1 Profesorado implicado en la docencia

<b>Nombre</b>	<b>Despacho</b>	<b>Correo electrónico</b>	<b>Horario de tutorías*</b>
Juan Pedro Caraca-Valente Hernandez (Coordinador/a)	D4301	juanpedro.caracavalente@u pm.es	L - 09:00 - 12:00 X - 09:00 - 12:00
Aurora Perez Perez	D4301	aurora.perez@upm.es	M - 11:00 - 13:30 X - 10:00 - 13:30

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

## 3. Competencias y resultados de aprendizaje

---

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

CEM2 - Analizar y sintetizar soluciones a problemas que requieran aproximaciones novedosas para la definición de la infraestructura computacional que permita el procesamiento y el análisis de datos de diversa naturaleza

CEM7 - Evaluar y aplicar las diversas teorías matemáticas y estadísticas, y los procesos, métodos y técnicas disponibles para la extracción y descubrimiento de conocimiento a partir de grandes volúmenes de datos

CG13 - Apreciación de los límites del conocimiento actual y de la aplicación práctica de la tecnología más reciente.

CG7 - Especificación y realización de tareas informáticas complejas, poco definidas o no familiares

CG8 - Planteamiento y resolución de problemas también en áreas nuevas y emergentes de su disciplina

CG9 - Aplicación de los métodos de resolución de problemas más recientes o innovadores y que puedan implicar el uso de otras disciplinas

CGI20 - Adquirir conocimientos científicos avanzados del campo de la informática que le permitan generar nuevas ideas dentro de una línea de investigación.

CGI23 - Capacidad de leer y comprender publicaciones dentro de su ámbito de estudio/investigación, así como su catalogación y valor científico

### 3.2 Resultados del aprendizaje al cursar la asignatura

RA68 - Ser capaz de analizar un dominio para determinar la relevancia de sus características temporales y las tareas de descubrimiento de conocimiento que se podrían plantear

RA70 - Ser capaz de realizar una evaluación completa del funcionamiento y utilidad de un proyecto de este tipo.

RA69 - Ser capaz de utilizar las técnicas de descubrimiento de conocimiento y su aplicabilidad en cada caso

## 4. Descripción de la asignatura y temario

---

### 4.1 Descripción de la asignatura

Las técnicas de Descubrimiento de Conocimiento (o de Data Mining) en grandes volúmenes de información son ampliamente utilizadas hoy en día en diferentes dominios como la medicina, entornos bancarios, sistemas industriales, etc. y tienen una gran variedad de aplicaciones como por ejemplo el análisis de datos, la detección de fraude, el análisis de riesgos, campañas de marketing, etc.

En esta asignatura se revisarán todas las etapas del proceso de Descubrimiento de Conocimiento y se enumerarán las técnicas más importantes para cada etapa. Se hará hincapié en las técnicas de limpieza y preprocesado de los datos, frecuentemente olvidadas.

A continuación se abordarán las principales técnicas de Data Mining incluyendo técnicas de Clasificación y Clustering. Los algoritmos genéticos han tenido un gran auge en los últimos años, surgiendo numerosas aplicaciones de dichos algoritmos en el área de Descubrimiento de Conocimiento, por lo que también serán tratados en profundidad en esta asignatura.

Recientemente se ha producido una tendencia a la incorporación de información de carácter temporal en grandes Bases de Datos para conservar información histórica, para poder analizar la evolución de una variable o conocer los instantes en que un dato es válido. Además, hay dominios donde la información se presenta mayoritariamente en forma de Series Temporales las cuales requieren un tratamiento muy especializado. Por ello, se abordarán las técnicas de descubrimiento de información en series temporales ya que este tipo de datos suponen un desafío a las técnicas tradicionales de Data Mining y requieren el uso de nuevas soluciones.

## 4.2 Temario de la asignatura

1. Introducción
  - 1.1. Reseña Histórica
  - 1.2. Conceptos Básicos
2. Proceso de Descubrimiento de Conocimiento
  - 2.1. Etapas del Proceso
  - 2.2. Preprocesamiento de los Datos
3. Herramientas de KDD
  - 3.1. Visión general
  - 3.2. Una herramienta de KDD: WEKA
4. Técnicas de Data Mining
  - 4.1. Clasificación
  - 4.2. Clustering
  - 4.3. Algoritmos Genéticos
  - 4.4. Data Mining en Series Temporales
5. Evaluación de Resultados
  - 5.1. Importancia y Objetivos
  - 5.2. Técnicas de evaluación y validación

## 5. Cronograma

### 5.1 Cronograma de la asignatura\*

Semana	Actividad Presencial en Aula	Actividad Presencial en Laboratorio	Otra Actividad Presencial	Actividades de Evaluación
1	<b>1. Introducción</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral			
2	<b>2.1 Etapas del Proceso KDD</b> Duración: 01:45 LM: Actividad del tipo Lección Magistral  <b>Supuesto práctico sobre el Proceso KDD</b> Duración: 00:15 AC: Actividad del tipo Acciones Cooperativas			
3	<b>2.2 Preprocesamiento de los datos</b> Duración: 01:45 LM: Actividad del tipo Lección Magistral  <b>Supuesto práctico sobre Procesamiento de Datos</b> Duración: 00:15 AC: Actividad del tipo Acciones Cooperativas			
4	<b>3.1 Herramientas KDD: Introducción</b> Duración: 00:20 LM: Actividad del tipo Lección Magistral  <b>3.2 WEKA</b> Duración: 01:40 LM: Actividad del tipo Lección Magistral			
5	<b>3.1 Trabajo práctico con WEKA</b> Duración: 01:20 PR: Actividad del tipo Clase de Problemas  <b>Analisis de Dominios y Proceso KDD</b> Duración: 00:40 AC: Actividad del tipo Acciones Cooperativas			
6	<b>4.1 Técnicas de Clasificación</b> Duración: 01:45 LM: Actividad del tipo Lección Magistral  <b>Supuesto práctico sobre Técnicas de Clasificación</b> Duración: 00:15 AC: Actividad del tipo Acciones Cooperativas		<b>Tutorías colectivas</b> Duración: 01:00 OT: Otras actividades formativas	

7	<p><b>4.1 Técnicas de Clasificación</b> Duración: 01:45 LM: Actividad del tipo Lección Magistral</p> <p><b>Supuesto práctico sobre Técnicas de Clasificación</b> Duración: 00:15 AC: Actividad del tipo Acciones Cooperativas</p>			<p><b>Entrega de la Práctica Fase 1: Análisis del dominio, estudio de los datos y establecimiento de objetivos</b> TG: Técnica del tipo Trabajo en Grupo Evaluación continua Duración: 00:00</p>
8	<p><b>4.1 Técnicas de Clasificación</b> Duración: 01:45 LM: Actividad del tipo Lección Magistral</p> <p><b>Supuesto práctico sobre Técnicas de Clasificación</b> Duración: 00:15 AC: Actividad del tipo Acciones Cooperativas</p>			
9	<p><b>4.2 Técnicas de Clustering</b> Duración: 01:45 LM: Actividad del tipo Lección Magistral</p> <p><b>Supuesto práctico sobre Técnicas de Clustering</b> Duración: 00:15 AC: Actividad del tipo Acciones Cooperativas</p>			
10	<p><b>4.2 Técnicas de Clustering</b> Duración: 01:45 LM: Actividad del tipo Lección Magistral</p> <p><b>Supuesto práctico sobre Técnicas de Clustering</b> Duración: 00:15 AC: Actividad del tipo Acciones Cooperativas</p>		<p><b>Tutorías colectivas</b> Duración: 01:00 OT: Otras actividades formativas</p>	
11	<p><b>4.2 Técnicas de Clustering</b> Duración: 01:45 LM: Actividad del tipo Lección Magistral</p> <p><b>Supuesto práctico sobre Técnicas de Clustering</b> Duración: 00:15 AC: Actividad del tipo Acciones Cooperativas</p>			
12	<p><b>4.3 Algoritmos Genéticos</b> Duración: 01:45 LM: Actividad del tipo Lección Magistral</p> <p><b>Supuesto práctico sobre Algoritmos Genéticos</b> Duración: 00:15 AC: Actividad del tipo Acciones Cooperativas</p>			<p><b>Entrega de la Práctica Fase 2: Aplicación de algoritmos de Data Mining y análisis de sus limitaciones y posibles mejoras</b> TG: Técnica del tipo Trabajo en Grupo Evaluación continua Duración: 00:00</p>

13	<p><b>4.3 Algoritmos Genéticos</b> Duración: 01:45 LM: Actividad del tipo Lección Magistral</p> <p><b>Supuesto práctico sobre Algoritmos Genéticos</b> Duración: 00:15 AC: Actividad del tipo Acciones Cooperativas</p>			
14	<p><b>4.4 Data Mining Temporal</b> Duración: 01:45 LM: Actividad del tipo Lección Magistral</p> <p><b>Supuesto práctico sobre Técnicas de Data Mining en Series Temporales</b> Duración: 00:15 AC: Actividad del tipo Acciones Cooperativas</p>			
15	<p><b>4.4 Data Mining Temporal</b> Duración: 01:45 LM: Actividad del tipo Lección Magistral</p> <p><b>Supuesto práctico sobre Técnicas de Abstracción Temporal</b> Duración: 00:15 AC: Actividad del tipo Acciones Cooperativas</p>		<p><b>Tutorías colectivas</b> Duración: 01:00 OT: Otras actividades formativas</p>	
16	<p><b>5 Evaluación de Resultados</b> Duración: 01:45 LM: Actividad del tipo Lección Magistral</p> <p><b>Supuesto práctico sobre Análisis y Evaluación de Resultados</b> Duración: 00:15 AC: Actividad del tipo Acciones Cooperativas</p>			
17				<p><b>Entrega de la Práctica Fase 3: Evaluación de los resultados obtenidos</b> TG: Técnica del tipo Trabajo en GrupoEvaluación continua Duración: 00:00</p> <p><b>Entrega de la Práctica Completa</b> TG: Técnica del tipo Trabajo en GrupoEvaluación sólo prueba final Duración: 00:00</p> <p><b>Presentación Oral de la Práctica</b> PG: Técnica del tipo Presentación en GrupoEvaluación continua y sólo prueba final Duración: 02:00</p> <p><b>Asistencia a clase, participación y ejercicios evaluables</b> TI: Técnica del tipo Trabajo IndividualEvaluación continua Duración: 00:00</p>

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

## 6. Actividades y criterios de evaluación

### 6.1 Actividades de evaluación de la asignatura

#### 6.1.1 Evaluación continua

Sem.	Descripción	Modalidad	Tipo	Duración	Peso en la nota	Nota mínima	Competencias evaluadas
7	Entrega de la Práctica Fase 1: Análisis del dominio, estudio de los datos y establecimiento de objetivos	TG: Técnica del tipo Trabajo en Grupo	Presencial	00:00	10%	/ 10	CG8 CG7
12	Entrega de la Práctica Fase 2: Aplicación de algoritmos de Data Mining y análisis de sus limitaciones y posibles mejoras	TG: Técnica del tipo Trabajo en Grupo	Presencial	00:00	20%	/ 10	CEM2 CG13 CGI20
17	Entrega de la Práctica Fase 3: Evaluación de los resultados obtenidos	TG: Técnica del tipo Trabajo en Grupo	Presencial	00:00	10%	/ 10	CG9 CG13 CGI23 CEM7
17	Presentación Oral de la Práctica	PG: Técnica del tipo Presentación en Grupo	Presencial	02:00	30%	/ 10	CG9 CGI20 CGI23
17	Asistencia a clase, participación y ejercicios evaluables	TI: Técnica del tipo Trabajo Individual	Presencial	00:00	30%	/ 10	CGI20 CGI23

#### 6.1.2 Evaluación sólo prueba final

Sem.	Descripción	Modalidad	Tipo	Duración	Peso en la nota	Nota mínima	Competencias evaluadas
17	Entrega de la Práctica Completa	TG: Técnica del tipo Trabajo en Grupo	Presencial	00:00	70%	/ 10	CG8 CG9 CEM2 CG7 CG13 CGI20 CGI23 CEM7

17	Presentación Oral de la Práctica	PG: Técnica del tipo Presentación en Grupo	Presencial	02:00	30%	/ 10	CG9 CGI20 CGI23
----	----------------------------------	--	------------	-------	-----	------	-----------------------

### 6.1.3 Evaluación convocatoria extraordinaria

No se ha definido la evaluación extraordinaria.

## 6.2 Criterios de Evaluación

Para la evaluación de la asignatura se tendrá en cuenta, por una parte, la asistencia y participación en clase y, por otra, la práctica de la asignatura.

Para la Evaluación Continua se considerarán la asistencia a clase, la participación activa del alumno y los ejercicios evaluables que se planteen en clase.

La práctica de la asignatura se evaluará de acuerdo con las tres fases que se describen a continuación y los pesos correspondientes.

### Práctica

La parte práctica consta de un trabajo que se realizará en grupos de 2 personas, pudiendo realizarse excepcionalmente de forma individual. El trabajo se realizará de forma incremental y se presentará en las siguientes entregas:

- Fase 1: los alumnos elegirán un dominio al que tengan acceso, analizarán sus características y establecerán los objetivos a lograr. Redactarán una memoria indicando las distintas tareas que se llevarían a cabo en cada etapa del proceso de Descubrimiento de Conocimiento de acuerdo a las necesidades específicas del dominio y a los objetivos definidos.
- Fase 2: mediante el uso de una herramienta software de Descubrimiento de Conocimiento, se aplicarán algoritmos de Data Mining a los datos de cada dominio. Además, el alumno analizará las limitaciones de los algoritmos disponibles en la herramienta y las posibles mejoras.
- Fase 3: se realizará un plan de evaluación para la valoración de los resultados obtenidos y se ejecutará dicho plan.

Las 3 entregas de ejercicios prácticos son de carácter obligatorio y se evaluarán según los pesos asignados en la tabla del apartado anterior (evaluación sumativa).

La práctica completa será presentada en clase. Cada grupo dispondrá de 15 minutos para la presentación oral más 5 minutos de preguntas.

#### Normas de calificación

La asignatura se evaluará sobre 10 puntos, repartidos en 3 puntos para la evaluación continua y 7 para la práctica. Para superar la asignatura será necesario asistir al menos a un 70% de las clases y obtener una nota no inferior a 5 puntos en la calificación final.

Las fechas tanto de entrega como de publicación de notas de cada parte de la práctica se publicarán en sus enunciados.

En las convocatorias extraordinarias se podrán entregar aquellas partes de la práctica que queden pendientes. No se volverá a realizar la evaluación continua, por lo que la nota de la asignatura se obtendrá de la práctica. No obstante, se podrá plantear un examen que sustituya la evaluación continua en estas convocatorias.

## 7. Recursos didácticos

---

### 7.1 Recursos didácticos de la asignatura

Nombre	Tipo	Observaciones
WEKA	Recursos web	Página web oficial de la herramienta de Data Mining WEKA, con tutoriales y posibilidad de descarga gratuita <a href="http://www.cs.waikato.ac.nz/ml/weka/">http://www.cs.waikato.ac.nz/ml/weka/</a>
Data Mining: Concepts and Techniques	Bibliografía	Libro sobre técnicas de Data Mining. J.Han y M. Kamber. Ed. Morgan Kauffman, 2006.
Data Mining: Concepts, Models, Methods, and Algorithms	Bibliografía	Libro sobre técnicas de Data Mining. M. Kantardzic (eds.), John Wiley & Sons, 2003
From Data Mining to Knowledge Discovery in Databases	Bibliografía	Libro con artículos fundacionales sobre muchas de las técnicas actuales de Data Mining U. Fayyad, G. Piatetsky-Shapiro y P. Smyth, 1996

Página web de la asignatura	Recursos web	<a href="http://www.dlsiis.fi.upm.es/master_muss/asigDCBD.html">http://www.dlsiis.fi.upm.es/master_muss/asigDCBD.html</a>
-----------------------------	--------------	---

## 8. Otra información

---

### 8.1 Otra información sobre la asignatura

Las clases se impartirán en Castellano y la documentación (incluido las transparencias usadas en clase) se proporcionará en Inglés