

**ANX-PR/CL/001-01**  
**GUÍA DE APRENDIZAJE**

**ASIGNATURA**

Descubrimiento de conocimiento en bases de datos

**CURSO ACADÉMICO - SEMESTRE**

2016-17 - Primer semestre

## Datos Descriptivos

---

<b>Nombre de la Asignatura</b>	Descubrimiento de conocimiento en bases de datos
<b>Titulación</b>	10AK - Master Universitario en Software y Sistemas
<b>Centro responsable de la titulación</b>	Escuela Técnica Superior de Ingenieros Informáticos
<b>Semestre/s de impartición</b>	Primer semestre
<b>Módulos</b>	Modulo sistemas
<b>Materias</b>	Análisis y procesado de datos e información
<b>Carácter</b>	Optativa
<b>Código UPM</b>	103000391
<b>Nombre en inglés</b>	Knowledge discovery in temporal data

## Datos Generales

---

<b>Créditos</b>	4	<b>Curso</b>	1
<b>Curso Académico</b>	2016-17	<b>Período de impartición</b>	Septiembre-Enero
<b>Idioma de impartición</b>	Castellano	<b>Otros idiomas de impartición</b>	

## Requisitos Previos Obligatorios

---

### Asignaturas Previas Requeridas

El plan de estudios Master Universitario en Software y Sistemas no tiene definidas asignaturas previas superadas para esta asignatura.

### Otros Requisitos

El plan de estudios Master Universitario en Software y Sistemas no tiene definidos otros requisitos para esta asignatura.

## Conocimientos Previos

---

### Asignaturas Previas Recomendadas

El coordinador de la asignatura no ha definido asignaturas previas recomendadas.

### Otros Conocimientos Previos Recomendados

El coordinador de la asignatura no ha definido otros conocimientos previos recomendados.

## Competencias

---

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

## Resultados de Aprendizaje

---

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

## Profesorado

---

### Profesorado

Nombre	Despacho	e-mail	Tutorías
Caraca-Valente Hernandez, Juan Pedro (Coordinador/a)	D4301	juanpedro.caracavalente@upm.es	X - 09:00 - 12:00 V - 10:00 - 13:00
Perez Perez, Aurora	D4301	aurora.perez@upm.es	X - 10:00 - 13:00 V - 10:00 - 13:00

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

## 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.

## Temario

---

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



CAMPUS  
DE EXCELENCIA  
INTERNACIONAL

**UNIVERSIDAD POLITÉCNICA DE MADRID**  
Escuela Técnica Superior de Ingenieros Informáticos  
**PROCESO DE COORDINACIÓN DE LAS ENSEÑANZAS**

ANX-PR/CL/001-01: GUÍA DE APRENDIZAJE



Código PR/CL/001

## Cronograma

**Horas totales:** 37 horas

**Horas presenciales:** 37 horas (35.6%)

**Peso total de actividades de evaluación continua:**  
100%

**Peso total de actividades de evaluación sólo prueba final:**  
100%

Semana	Actividad Presencial en Aula	Actividad Presencial en Laboratorio	Otra Actividad Presencial	Actividades Evaluación
Semana 1	<b>1. Introducción</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral			
Semana 2	<b>2.1 Etapas del Proceso KDD</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral			
Semana 3	<b>2.2 Preprocesamiento de los datos</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral			
Semana 4	<b>3.1 Herramientas KDD: Introducción</b> Duración: 01:00 LM: Actividad del tipo Lección Magistral <b>3.2 WEKA</b> Duración: 01:00 LM: Actividad del tipo Lección Magistral			
Semana 5	<b>3.1 WEKA</b> Duración: 01:00 PR: Actividad del tipo Clase de Problemas <b>Análisis de Dominios y Proceso KDD</b> Duración: 01:00 AC: Actividad del tipo Acciones Cooperativas			
Semana 6	<b>4.1 Técnicas de Clasificación</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral		<b>Tutorías colectivas</b> Duración: 01:00 OT: Otras actividades formativas	
Semana 7	<b>4.1 Técnicas de Clasificación</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral			<b>Entrega de la Práctica Fase 1: Análisis del dominio, estudio de los datos y establecimiento de objetivos</b> Duración: 00:00 TG: Técnica del tipo Trabajo en Grupo Evaluación continua Actividad presencial
Semana 8	<b>4.1 Técnicas de Clasificación</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral			
Semana 9	<b>4.2 Técnicas de Clustering</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral			

Semana 10	<b>4.2 Técnicas de Clustering</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral		<b>Tutorías colectivas</b> Duración: 01:00 OT: Otras actividades formativas	
Semana 11	<b>4.2 Técnicas de Clustering</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral			
Semana 12	<b>4.3 Algoritmos Genéticos</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral			<b>Entrega de la Práctica Fase 2: Aplicación de algoritmos de Data Mining y análisis de sus limitaciones y posibles mejoras</b> Duración: 00:00 TG: Técnica del tipo Trabajo en Grupo Evaluación continua Actividad presencial
Semana 13	<b>4.3 Algoritmos Genéticos</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral			
Semana 14	<b>4.4 Data Mining Temporal</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral			
Semana 15	<b>4.4 Data Mining Temporal</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral		<b>Tutorías colectivas</b> Duración: 01:00 OT: Otras actividades formativas	
Semana 16	<b>5 Evaluación de Resultados</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral			
Semana 17				<b>Entrega de la Práctica Fase 3: Evaluación de los resultados obtenidos</b> Duración: 00:00 TG: Técnica del tipo Trabajo en Grupo Evaluación continua Actividad presencial <b>Entrega de la Práctica Completa</b> Duración: 00:00 TG: Técnica del tipo Trabajo en Grupo Evaluación sólo prueba final Actividad presencial <b>Presentación Oral de la Práctica</b> Duración: 02:00 PG: Técnica del tipo Presentación en Grupo Evaluación continua y sólo prueba final Actividad presencial <b>Asistencia a clase, participación y ejercicios evaluables</b> Duración: 00:00 TI: Técnica del tipo Trabajo Individual Evaluación continua Actividad presencial



**Nota.-** El cronograma sigue una planificación teórica de la asignatura que puede sufrir modificaciones durante el curso.

**Nota 2.-** Para poder calcular correctamente la dedicación de un alumno, la duración de las actividades que se repiten en el tiempo (por ejemplo, subgrupos de prácticas") únicamente se indican la primera vez que se definen.

## Actividades de Evaluación

Semana	Descripción	Duración	Tipo evaluación	Técnica evaluativa	Presencial	Peso	Nota mínima	Competencias evaluadas
7	Entrega de la Práctica Fase 1: Análisis del dominio, estudio de los datos y establecimiento de objetivos	00:00	Evaluación continua	TG: Técnica del tipo Trabajo en Grupo	Sí	10%		CG7, CG8
12	Entrega de la Práctica Fase 2: Aplicación de algoritmos de Data Mining y análisis de sus limitaciones y posibles mejoras	00:00	Evaluación continua	TG: Técnica del tipo Trabajo en Grupo	Sí	20%		CGI20, CEM2, CGI13
17	Entrega de la Práctica Fase 3: Evaluación de los resultados obtenidos	00:00	Evaluación continua	TG: Técnica del tipo Trabajo en Grupo	Sí	10%		CGI23, CGI13, CEM7, CG9
17	Entrega de la Práctica Completa	00:00	Evaluación sólo prueba final	TG: Técnica del tipo Trabajo en Grupo	Sí	70%		CG9, CG7, CEM7, CGI20, CGI23, CG8, CGI13, CEM2
17	Presentación Oral de la Práctica	02:00	Evaluación continua y sólo prueba final	PG: Técnica del tipo Presentación en Grupo	Sí	30%		CGI23, CG9, CGI20
17	Asistencia a clase, participación y ejercicios evaluables	00:00	Evaluación continua	TI: Técnica del tipo Trabajo Individual	Sí	30%		CGI20, CGI23

## 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.

## Recursos Didácticos

---

Descripción	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>

## Otra Información

---

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