



UNIVERSIDAD
POLITÉCNICA
DE MADRID

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

2021/22 - 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.....	12
8. Otra información.....	12

1. Datos descriptivos

1.1. Datos de la asignatura

Nombre de la asignatura	103000391 - Descubrimiento de Conocimiento en Bases de Datos
No de créditos	4 ECTS
Carácter	Optativa
Curso	Primer curso
Semestre	Primer semestre
Período de impartición	Septiembre-Enero
Idioma de impartición	Castellano
Titulación	10AK - Master Universitario en Software y Sistemas
Centro responsable de la titulación	10 - Escuela Tecnica Superior De Ingenieros Informaticos
Curso académico	2021-22

2. Profesorado

2.1. Profesorado implicado en la docencia

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

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

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.

CG12 - Comprensión amplia de las técnicas y métodos aplicables en una especialización concreta, así como de sus límites

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

CG14 - Conocimiento y comprensión de la informática necesaria para la creación de modelos de información, y de los sistemas y procesos complejos

CG17 - Habilidades de gestión y capacidad de liderar un equipo que puede estar integrado por disciplinas y niveles distintos.

CG19 - Aproximación sistemática a la gestión de riesgos.

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.

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.

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

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

This course will be taught in English primarily, but can be changed to Spanish if all the students prefer that option.

Knowledge Discovery techniques (or Data Mining) in large volumes of information are widely used today in different domains such as medicine, banking environments, industrial systems, etc. with a wide variety of applications such as data analysis, fraud detection, risk analysis, marketing campaigns, etc.

In this course all the stages of the Knowledge Discovery process will be reviewed and the most important techniques for each stage will be listed. Emphasis will be placed on techniques for data cleaning and preprocessing that, despite their importance, are often forgotten.

Next, the main techniques of Data Mining including Classification and Clustering techniques will be addressed. The genetic algorithms have had a great boom in recent years, arising numerous applications of such algorithms in the area of Knowledge Discovery, so they will also be treated in depth.

In this subject we also want to explore areas of Knowledge Discovery less known, but equally important. There are domains where information is presented mostly in the form of Time Series which require a very specialized treatment. Examples of these are medical domains such as Electrocardiography or Audiometry, financial domains, etc. Time series are a challenge to the traditional techniques of Data Mining and often require the use of novel solutions. Special emphasis will be made on Temporal Abstraction techniques.

4.2. Temario de la asignatura

1. Introduction
 - 1.1. Data Types, Time Series
 - 1.2. Basic Concepts
2. Knowledge Discovery Process
 - 2.1. Knowledge Discovery Process Stages
 - 2.2. Data Preprocessing for basic data types and time series
3. KDD Tools
 - 3.1. Background
 - 3.2. A KDD Tool: WEKA
4. Data Mining Techniques
 - 4.1. Classification
 - 4.2. Clustering
 - 4.3. Time Series Techniques
 - 4.4. Genetic Algorithms
5. Evaluation
 - 5.1. Objectives
 - 5.2. Evaluation Techniques

5. Cronograma

5.1. Cronograma de la asignatura *

Sem	Actividad presencial en aula	Actividad presencial en laboratorio	Tele-enseñanza	Actividades de evaluación
1	1. Introduction Duración: 02:00 LM: Actividad del tipo Lección Magistral		1. Introduction Duración: 02:00 LM: Actividad del tipo Lección Magistral	
2	2.1 Knowledge Discovery Process Duración: 01:45 LM: Actividad del tipo Lección Magistral Case Study: Knowledge Discovery Process Duración: 00:15 AC: Actividad del tipo Acciones Cooperativas		2.1 Knowledge Discovery Process Duración: 01:45 LM: Actividad del tipo Lección Magistral Case Study: Knowledge Discovery Process Duración: 00:15 AC: Actividad del tipo Acciones Cooperativas	
3	2.2 Data Preprocessing for basic data types and time series Duración: 01:45 LM: Actividad del tipo Lección Magistral Case Study: Data Preprocessing Duración: 00:15 AC: Actividad del tipo Acciones Cooperativas		2.2 Data Preprocessing for basic data types and time series Duración: 01:45 LM: Actividad del tipo Lección Magistral Case Study: Data Preprocessing Duración: 00:15 AC: Actividad del tipo Acciones Cooperativas	
4	3 KDD Tools Duración: 00:20 LM: Actividad del tipo Lección Magistral 3.1 Background and 3.2 WEKA Duración: 01:40 LM: Actividad del tipo Lección Magistral		3 KDD Tools Duración: 00:20 LM: Actividad del tipo Lección Magistral 3.1 Background and 3.2 WEKA Duración: 01:40 LM: Actividad del tipo Lección Magistral	
5	3.2 Case Study: WEKA Duración: 01:20 PR: Actividad del tipo Clase de Problemas Domain Analysis and KDD Process Duración: 00:40 AC: Actividad del tipo Acciones Cooperativas		3.2 Case Study: WEKA Duración: 01:20 PR: Actividad del tipo Clase de Problemas Domain Analysis and KDD Process Duración: 00:40 AC: Actividad del tipo Acciones Cooperativas	
6	4.1 Classification Techniques Duración: 01:15 LM: Actividad del tipo Lección Magistral Group Discussion Duración: 00:45 OT: Otras actividades formativas		4.1 Classification Techniques Duración: 01:15 LM: Actividad del tipo Lección Magistral Group Discussion Duración: 00:45 OT: Otras actividades formativas	

7	<p>4.1 Classification Techniques Duración: 01:45 LM: Actividad del tipo Lección Magistral</p> <p>Case Study: Classification Techniques Duración: 00:15 AC: Actividad del tipo Acciones Cooperativas</p>		<p>4.1 Classification Techniques Duración: 01:45 LM: Actividad del tipo Lección Magistral</p> <p>Case Study: Classification Techniques Duración: 00:15 AC: Actividad del tipo Acciones Cooperativas</p>	<p>Project Stage 1: Domain Analysis, Data study, Objective definition TG: Técnica del tipo Trabajo en Grupo Evaluación continua No presencial Duración: 00:20</p>
8	<p>4.2 Clustering Techniques Duración: 01:45 LM: Actividad del tipo Lección Magistral</p> <p>Case Study: Clustering Techniques Duración: 00:15 AC: Actividad del tipo Acciones Cooperativas</p>		<p>4.2 Clustering Techniques Duración: 01:45 LM: Actividad del tipo Lección Magistral</p> <p>Case Study: Clustering Techniques Duración: 00:15 AC: Actividad del tipo Acciones Cooperativas</p>	
9	<p>4.2 Clustering Techniques Duración: 01:45 LM: Actividad del tipo Lección Magistral</p> <p>Case Study: Clustering Techniques Duración: 00:15 AC: Actividad del tipo Acciones Cooperativas</p>		<p>4.2 Clustering Techniques Duración: 01:45 LM: Actividad del tipo Lección Magistral</p> <p>Case Study: Clustering Techniques Duración: 00:15 AC: Actividad del tipo Acciones Cooperativas</p>	
10	<p>4.3 Time Series Data Mining Duración: 01:45 LM: Actividad del tipo Lección Magistral</p> <p>Case Study: Time Series Data Mining Duración: 00:15 AC: Actividad del tipo Acciones Cooperativas</p>		<p>4.3 Time Series Data Mining Duración: 01:45 LM: Actividad del tipo Lección Magistral</p> <p>Case Study: Time Series Data Mining Duración: 00:15 AC: Actividad del tipo Acciones Cooperativas</p>	
11	<p>4.3 Time Series Data Mining Duración: 01:45 LM: Actividad del tipo Lección Magistral</p> <p>Case Study: Time Series Data Mining Duración: 00:15 AC: Actividad del tipo Acciones Cooperativas</p>		<p>4.3 Time Series Data Mining Duración: 01:45 LM: Actividad del tipo Lección Magistral</p> <p>Case Study: Time Series Data Mining Duración: 00:15 AC: Actividad del tipo Acciones Cooperativas</p>	
12	<p>4.3 Time Series Data Mining Duración: 01:45 LM: Actividad del tipo Lección Magistral</p> <p>Case Study: Time Series Data Mining Duración: 00:15 AC: Actividad del tipo Acciones Cooperativas</p>		<p>4.3 Time Series Data Mining Duración: 01:45 LM: Actividad del tipo Lección Magistral</p> <p>Case Study: Time Series Data Mining Duración: 00:15 AC: Actividad del tipo Acciones Cooperativas</p>	<p>Project Stage 2: Application of Data Mining Techniques TG: Técnica del tipo Trabajo en Grupo Evaluación continua No presencial Duración: 00:20</p>
13	<p>4.4 Genetic Algorithms Duración: 01:45 LM: Actividad del tipo Lección Magistral</p> <p>Group Discussion Duración: 00:15 OT: Otras actividades formativas</p>		<p>4.4 Genetic Algorithms Duración: 01:45 LM: Actividad del tipo Lección Magistral</p> <p>Group Discussion Duración: 00:15 OT: Otras actividades formativas</p>	

14	5 Evaluation Duración: 01:45 LM: Actividad del tipo Lección Magistral Group Discussion Duración: 00:15 OT: Otras actividades formativas		5 Evaluation Duración: 01:45 LM: Actividad del tipo Lección Magistral Group Discussion Duración: 00:15 OT: Otras actividades formativas	
15	Project Presentations Duración: 02:00 OT: Otras actividades formativas			Project Stage 3: Evaluation TG: Técnica del tipo Trabajo en Grupo Evaluación continua No presencial Duración: 00:20 Project Presentation PG: Técnica del tipo Presentación en Grupo Evaluación continua y sólo prueba final No presencial Duración: 02:00
16				
17				Project complete TG: Técnica del tipo Trabajo en Grupo Evaluación sólo prueba final No presencial Duración: 01:00 Attendance to class, participation and evaluable exercises TI: Técnica del tipo Trabajo Individual Evaluación continua No presencial Duración: 00:00

Para el cálculo de los valores totales, se estima que por cada crédito ECTS el alumno dedicará dependiendo del plan de estudios, entre 26 y 27 horas de trabajo presencial y no presencial.

* El cronograma sigue una planificación teórica de la asignatura y puede sufrir modificaciones durante el curso derivadas de la situación creada por la COVID-19.

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	Project Stage 1: Domain Analysys, Data study, Objective definition	TG: Técnica del tipo Trabajo en Grupo	No Presencial	00:20	10%	/ 10	CG8 CG1 CG7 CG12 CG17
12	Project Stage 2: Application of Data Mining Techniques	TG: Técnica del tipo Trabajo en Grupo	No Presencial	00:20	10%	/ 10	CG4 CEM2 CG1 CG13 CG14 CG17 CGI20
15	Project Stage 3: Evaluation	TG: Técnica del tipo Trabajo en Grupo	No Presencial	00:20	10%	/ 10	CG9 CG19 CG1 CG13 CG17 CGI23 CEM7
15	Project Presentation	PG: Técnica del tipo Presentación en Grupo	No Presencial	02:00	40%	/ 10	CG9 CG3 CG4 CG17 CGI20 CGI23
17	Attendance to class, participation and evaluable exercises	TI: Técnica del tipo Trabajo Individual	No Presencial	00:00	30%	/ 10	CG8 CG1 CG12 CGI20 CGI23

6.1.2. Evaluación sólo prueba final

Sem	Descripción	Modalidad	Tipo	Duración	Peso en la nota	Nota mínima	Competencias evaluadas
-----	-------------	-----------	------	----------	-----------------	-------------	------------------------

15	Project Presentation	PG: Técnica del tipo Presentación en Grupo	No Presencial	02:00	40%	/ 10	CG9 CG3 CG4 CG17 CGI20 CGI23
17	Project complete	TG: Técnica del tipo Trabajo en Grupo	No Presencial	01:00	60%	/ 10	CG8 CG9 CEM2 CG19 CG1 CG7 CG12 CG13 CG14 CG17 CGI20 CGI23 CEM7

6.1.3. Evaluación convocatoria extraordinaria

Descripción	Modalidad	Tipo	Duración	Peso en la nota	Nota mínima	Competencias evaluadas
Project complete	TG: Técnica del tipo Trabajo en Grupo	Presencial	00:00	100%	5 / 10	CG4 CG8 CG9 CEM2 CG19 CG1 CG3 CG7 CG12 CG13 CG14 CG17 CGI20 CGI23 CEM7

6.2. Criterios de evaluación

For the evaluation of this subject we will take into account, on the one hand, the attendance and participation in class and, on the other hand, the Data Mining Project.

For the Continuous Evaluation the attendance to class (online or in class), active participation of the student and the evaluable exercises that are raised in class will be considered.

The Data Mining Project will be evaluated according to the three phases described below and the corresponding weights.

Data Mining Project

This project will be done individually or in groups of 2 people. The work will be done incrementally and will be presented in the following phases:

- Phase 1: students will choose a domain to which data they have access, analyze their characteristics and establish the objectives to be achieved through the Data Mining Project. They will write a report indicating the different tasks that would be carried out in each stage of the Knowledge Discovery process according to the specific needs of the domain and the objectives.
- Phase 2: through the use of a Knowledge Discovery software tool, Data Mining algorithms will be applied to the data of each domain. In addition, the student will analyze the limitations of the algorithms available in the tool and possible improvements.

- Phase 3: an evaluation plan will be made to assess the results obtained and the plan will be executed.

The 3 deliveries of the Data Mining Project are mandatory and will be evaluated according to the weights assigned in the table in the previous section (summative evaluation).

The Data Mining Project will be presented in class. Each group will have 15 minutes for the oral presentation plus 5 minutes of questions.

Qualification standards

The subject will be evaluated on 10 points, divided into 3 points for continuous assessment and 7 for the Data Mining Project. To pass the subject it will be necessary to attend at least 70% of the classes and obtain a final grade of no less than 5 points.

The dates for the delivery of each part of the Data Mining Project will be published at the beginning of the course.

In the extra call, those parts of the Data Mining Project that are pending may be delivered. Continuous assessment will not be repeated, so the grade of the subject will be obtained exclusively from the Data Mining Project.

7. Recursos didácticos

7.1. Recursos didácticos de la asignatura

Nombre	Tipo	Observaciones
WEKA	Recursos web	Official webpage of the Data Mining Tool WEKA, with tutorials and free download http://www.cs.waikato.ac.nz/ml/weka/
Data Mining: Concepts and Techniques	Bibliografía	Book about Data Mining Techniques. J.Han y M. Kamber. Ed. Morgan Kauffman, 2006.
Data Mining: Concepts, Models, Methods, and Algorithms	Bibliografía	Book about Data Mining Techniques. M. Kantardzic (eds.), John Wiley & Sons, 2003
From Data Mining to Knowledge Discovery in Databases	Bibliografía	Book about some foundational works on nowadays Data Mining Techniques U. Fayyad, G. Piatetsky-Shapiro y P. Smyth, 1996
Subject webpage	Recursos web	http://www.dlsiis.fi.upm.es/master_muss/asig DCBD.html

8. Otra información

8.1. Otra información sobre la asignatura

Classes will be taught in Spanish or English depending on the students. Documentation (including transparencies used in class) will be provided in English.

During the course, we will try to use as many the data files related to Sustainable Development Goals of UN as possible, specially number 13 Climate Action

Las clases se impartirán en Castellano o Ingles dependiendo del alumnado. La documentación (incluido las transparencias usadas en clase) se proporcionará en Inglés

Durante el curso, se utilizarán tantas ficheros de datos relacionados con Objetivos de Desarrollo Sostenible (ODS) de Naciones Unidas como sea posible, especialmente con el número 13 Acción por el Clima