



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

105001038 - Descubrimiento De Conocimiento En Datos Complejos

PLAN DE ESTUDIOS

10CD - Grado En Ciencia De Datos E Inteligencia Artificial

CURSO ACADÉMICO Y SEMESTRE

2023/24 - 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.....	2
5. Cronograma.....	5
6. Actividades y criterios de evaluación.....	7
7. Recursos didácticos.....	10
8. Otra información.....	11

1. Datos descriptivos

1.1. Datos de la asignatura

Nombre de la asignatura	105001038 - Descubrimiento de Conocimiento en Datos Complejos
No de créditos	3 ECTS
Carácter	Obligatoria
Curso	Cuarto curso
Semestre	Séptimo semestre
Período de impartición	Septiembre-Enero
Idioma de impartición	Castellano
Titulación	10CD - Grado en Ciencia de Datos e Inteligencia Artificial
Centro responsable de la titulación	10 - Escuela Tecnica Superior De Ingenieros Informaticos
Curso académico	2023-24

2. Profesorado

2.1. Profesorado implicado en la docencia

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

* 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

CB05 - Que los estudiantes hayan desarrollado aquellas habilidades de aprendizaje necesarias para emprender estudios posteriores con un alto grado de autonomía

CE10 - Capacidad para aplicar las metodologías y las técnicas adecuadas de análisis y explotación de datos sobre datos disponibles, incluidos los poco estructurados o de estructura compleja (como los que contienen series temporales, los provenientes de redes sociales, etc.), para descubrir nuevas relaciones y proporcionar conocimiento y una comprensión intuitiva precisa y profunda sobre problemas científicos o procesos organizacionales reales y así respaldar la toma de decisiones.

3.2. Resultados del aprendizaje

RA175 - Ser capaz de extraer conocimiento de datos complejos mediante técnicas de tratamiento numérico o simbólico, seleccionando técnicas existentes o modificándolas en función de las necesidades de cada dominio

RA174 - Analizar problemas reales con datos de estructura compleja o poco estructurados (como los que contienen series temporales, los provenientes de redes sociales, etc.), resolver los problemas de preprocesamiento asociados a estos datos y elegir y aplicar las técnicas adecuadas para el análisis de los mismos con el fin de descubrir nuevo conocimiento implícito en ellos.

4. Descripción de la asignatura y temario

4.1. Descripción de la asignatura

Las técnicas de Descubrimiento de Conocimiento (o Data Mining) en grandes volúmenes de información son ampliamente utilizadas hoy en día en diferentes dominios como la medicina, entornos bancarios, sistemas industriales, etc. con una amplia variedad de aplicaciones como análisis de datos, detección de fraudes, análisis de riesgos, campañas de marketing, etc. En este curso se revisarán todas las etapas del proceso de Descubrimiento del Conocimiento, y se enumerarán las técnicas más importantes para cada etapa, poniendo el foco de atención en el descubrimiento a partir de datos complejos y series temporales.

Se explorarán áreas del Descubrimiento del Conocimiento menos conocidas, pero que cada vez son más

importante, y que implican el manejo de información compleja, como series temporales, grafos, redes sociales, etc. y, por tanto, requieren un tratamiento especializado. Hay dominios donde la información se presenta mayoritariamente en forma de Series Temporales, por ejemplo, dominios médicos como electrocardiografía o audiometría, dominios financieros, etc. Las series temporales representa un desafío para las técnicas tradicionales de Minería de Datos y a menudo requieren el uso de soluciones novedosas. Discutiremos las técnicas numéricas tradicionales para el tratamiento de series de temporales, así como otros enfoques novedosos, prestando especial atención a los enfoques simbólicos.

4.2. Temario de la asignatura

1. Introducción
 - 1.1. Tipos de datos, series temporales
 - 1.2. Conceptos básicos
2. El Proceso de descubrimiento de conocimiento
 - 2.1. Etapas del proceso de descubrimiento del conocimiento
 - 2.2. Preprocesamiento de datos
3. Técnicas Numéricas: Enfoque Clásico
 - 3.1. Introducción a las distancias de las series temporales
 - 3.2. Transformadas y Transformaciones
 - 3.3. Descubrimiento de patrones
4. Técnicas numéricas: Predicción
 - 4.1. Autocorrelación
 - 4.2. Deep Learning

4.3. Caso práctico: predicción en series temporales

5. Técnicas Simbólicas para Series Temporales

5.1. Técnicas simbólicas independientes del dominio

5.2. Abstracción temporal dependiente del dominio

5.3. Caso de estudio

5. Cronograma

5.1. Cronograma de la asignatura *

Sem	Actividad en aula	Actividad en laboratorio	Tele-enseñanza	Actividades de evaluación
1	1. Introducción 1.1. Tipos de datos, series temporales 1.2. Conceptos básicos Duración: 02:00 LM: Actividad del tipo Lección Magistral			
2	2.El Proceso de Descubrimiento de Conocimiento 2.1. Etapas del proceso de descubrimiento del conocimiento Duración: 02:00 LM: Actividad del tipo Lección Magistral			
3	2.2. Preprocesamiento de datos Duración: 01:15 LM: Actividad del tipo Lección Magistral			Trabajo Práctico. Preprocesamiento TI: Técnica del tipo Trabajo Individual Evaluación continua Presencial Duración: 00:45
4	3. Técnicas Numéricas: Enfoque Clásico. 3.1. Introducción a las distancias de las series temporales Duración: 01:30 LM: Actividad del tipo Lección Magistral			Generación de Series Temporales: Componentes subyacentes en una serie temporal TI: Técnica del tipo Trabajo Individual Evaluación continua Presencial Duración: 00:30
5	Domain Analysis and KDD Process Duración: 00:20 AC: Actividad del tipo Acciones Cooperativas 3.2. Transformadas y Transformaciones Duración: 01:40 LM: Actividad del tipo Lección Magistral			
6	3.2. Transformadas y Transformaciones Duración: 01:00 LM: Actividad del tipo Lección Magistral 3.3. Descubrimiento de patrones Duración: 01:00 LM: Actividad del tipo Lección Magistral			Proyecto. Etapa 1. Análisis del dominio, estudio de los datos y definición de objetivos TG: Técnica del tipo Trabajo en Grupo Evaluación continua No presencial Duración: 00:20
7	4. Técnicas numéricas: pronóstico. 4.1. Autocorrelación: Introducción y Modelos ARIMA básicos Duración: 01:30 LM: Actividad del tipo Lección Magistral			Caso práctico: Modelos ARIMA básicos con R TI: Técnica del tipo Trabajo Individual Evaluación continua Presencial Duración: 00:30

8	<p>4.1. Autocorrelación: otros modelos ARIMA Duración: 01:30 LM: Actividad del tipo Lección Magistral</p>			<p>Caso práctico: ARIMA estacional o multivariable con R TI: Técnica del tipo Trabajo Individual Evaluación continua Presencial Duración: 00:30</p>
9	<p>4.2. Aprendizaje profundo Duración: 02:00 LM: Actividad del tipo Lección Magistral</p>			<p>Proyecto. Etapa 2. Aplicación de Técnicas de minería. TG: Técnica del tipo Trabajo en Grupo Evaluación continua No presencial Duración: 00:20</p>
10	<p>Discusión en Grupo Duración: 00:30 OT: Otras actividades formativas</p>			<p>Técnicas de predicción en Series Temporales PI: Técnica del tipo Presentación Individual Evaluación continua Presencial Duración: 01:30</p>
11	<p>4.3. Caso práctico: Predicción en Series Temporales Duración: 02:00 PR: Actividad del tipo Clase de Problemas</p>			
12	<p>5. Técnicas Simbólicas para Series Temporales. 5.1. Técnicas simbólicas independientes del dominio Duración: 02:00 LM: Actividad del tipo Lección Magistral</p>			
13	<p>5.2. Abstracción temporal dependiente del dominio Duración: 01:20 LM: Actividad del tipo Lección Magistral</p>			<p>Aplicación de SAX u otras técnicas de simbólicas TI: Técnica del tipo Trabajo Individual Evaluación continua Presencial Duración: 00:40</p>
14	<p>5.3. Caso de estudio: Abstracción Temporal Duración: 01:20 PR: Actividad del tipo Clase de Problemas</p>			<p>Aplicación de la Abstracción Temporal TI: Técnica del tipo Trabajo Individual Evaluación continua Presencial Duración: 00:40</p>
15				<p>Presentación del Proyecto completo de Descubrimiento de Conocimiento PG: Técnica del tipo Presentación en Grupo Evaluación continua Presencial Duración: 02:00</p>
16				
17				<p>Proyecto completo de Descubrimiento de Conocimiento PG: Técnica del tipo Presentación en Grupo Evaluación sólo prueba final Presencial Duración: 02:00</p>

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.

6. Actividades y criterios de evaluación

6.1. Actividades de evaluación de la asignatura

6.1.1. Evaluación (progresiva)

Sem.	Descripción	Modalidad	Tipo	Duración	Peso en la nota	Nota mínima	Competencias evaluadas
3	Trabajo Práctico. Preprocesamiento	TI: Técnica del tipo Trabajo Individual	Presencial	00:45	5%	0 / 10	CB05 CE10
4	Generación de Series Temporales: Componentes subyacentes en una serie temporal	TI: Técnica del tipo Trabajo Individual	Presencial	00:30	5%	0 / 10	CE10 CB05
6	Proyecto. Etapa 1. Análisis del dominio, estudio de los datos y definición de objetivos	TG: Técnica del tipo Trabajo en Grupo	No Presencial	00:20	10%	0 / 10	CE10 CB05
7	Caso práctico: Modelos ARIMA básicos con R	TI: Técnica del tipo Trabajo Individual	Presencial	00:30	5%	0 / 10	CE10 CB05
8	Caso práctico: ARIMA estacional o multivariable con R	TI: Técnica del tipo Trabajo Individual	Presencial	00:30	5%	0 / 10	CE10 CB05
9	Proyecto. Etapa 2. Aplicación de Técnicas de minería.	TG: Técnica del tipo Trabajo en Grupo	No Presencial	00:20	10%	0 / 10	CE10 CB05
10	Técnicas de predicción en Series Temporales	PI: Técnica del tipo Presentación Individual	Presencial	01:30	10%	0 / 10	CE10 CB05
13	Aplicación de SAX u otras técnicas de simbólicas	TI: Técnica del tipo Trabajo Individual	Presencial	00:40	5%	0 / 10	CE10 CB05

14	Aplicación de la Abstracción Temporal	TI: Técnica del tipo Trabajo Individual	Presencial	00:40	5%	0 / 10	CE10 CB05
15	Presentación del Proyecto completo de Descubrimiento de Conocimiento	PG: Técnica del tipo Presentación en Grupo	Presencial	02:00	50%	0 / 10	CE10 CB05

6.1.2. Prueba evaluación global

Sem	Descripción	Modalidad	Tipo	Duración	Peso en la nota	Nota mínima	Competencias evaluadas
17	Proyecto completo de Descubrimiento de Conocimiento	PG: Técnica del tipo Presentación en Grupo	Presencial	02:00	100%	5 / 10	CE10 CB05

6.1.3. Evaluación convocatoria extraordinaria

Descripción	Modalidad	Tipo	Duración	Peso en la nota	Nota mínima	Competencias evaluadas
Proyecto completo de Descubrimiento de Conocimiento	PG: Técnica del tipo Presentación en Grupo	Presencial	00:00	100%	5 / 10	CE10 CB05

6.2. Criterios de evaluación

Para la evaluación de esta asignatura tendremos en cuenta, por un lado, las Actividades de Evaluación Progresivas y, por otro lado, el Proyecto de Minería de Datos.

Para la **Evaluación Progresiva** consideraremos:

- la asistencia a clase
- la participación activa del alumno
- las tareas que se realizarán en cada clase aplicando los conceptos explicados anteriormente. Esta será la parte de más peso en la Evaluación Progresiva.

Como se realizan durante el curso, las actividades de evaluación progresiva no son recuperables, y no pueden repetirse para la Evaluación Final ni para la convocatoria extraordinaria.

El **Proyecto de Minería de Datos** será evaluado de acuerdo a las fases que se describen a continuación y sus pesos correspondientes.

Este proyecto se realizará individualmente o en grupos. El trabajo se realizará de forma incremental y se entregará en las siguientes fases:

- Fase 1: los estudiantes elegirán un dominio en el que estén interesados y para el cual tengan acceso a los datos, analizarán sus características y establecerán los objetivos a alcanzar a través del Proyecto de Minería de Datos. Escribirán un informe describiendo el dominio que han seleccionado, el análisis realizado sobre los datos y haciendo una propuesta inicial de los objetivos del proyecto.
- Fase 2: aplicarán algoritmos de minería a los datos seleccionados, usando para ello las herramientas software de Descubrimiento de Conocimiento o los lenguajes de programación que elijan. Los resultados deben ser evaluados y analizados. Además, los estudiantes analizarán las limitaciones de los algoritmos utilizados y las posibles mejoras.

Las dos entregas del Proyecto de Descubrimiento de Conocimiento son obligatorias y serán evaluadas según los pesos asignados a cada fase.

El Proyecto final se presentará en clase. La presentación de cada grupo irá seguida de un turno de preguntas en que se establecerá un pequeño debate sobre el trabajo presentado.

Calificación.

- Evaluación Progresiva: La asignatura se evaluará sobre 10 puntos, divididos en 4 puntos para Actividades de Evaluación Progresiva y 7 para el Proyecto de Minería de Datos (en total son más de 10 puntos; con ello se pretende incentivar la constancia en la realización de las actividades diarias). Las fechas concretas para la entrega de cada parte del Proyecto de Descubrimiento de Conocimiento se publicarán con la suficiente antelación.

- **Evaluación Final** y la **Convocatoria extraordinaria**: Sólo se considerará el Proyecto de Descubrimiento de

Conocimiento. Se deberán entregar las partes del Proyecto de Minería de Datos que estén pendientes. No se podrán repetir las Actividades de Evaluación Progresiva, pero se tendrá en cuenta la calificación obtenida en caso de que el estudiante las realizara durante el curso.

7. Recursos didácticos

7.1. Recursos didácticos de la asignatura

Nombre	Tipo	Observaciones
Time Series Analysis and Its Applications With R Examples - 4th Edition	Bibliografía	Libro sobre Análisis de Series Temporales de Shumway and Stoffer
A review on time series data mining	Bibliografía	Engineering Applications of Artificial Intelligence 24 (2011) 164?181
Data Mining: Concepts and Techniques	Bibliografía	Libro sobre Técnicas de Minería de Datos. J.Han y M. Kamber. Ed. Morgan Kauffman, 2006.
Data Mining: Concepts, Models, Methods, and Algorithms	Bibliografía	Libro sobre Técnicas de Minería de Datos. M. Kantardzic (eds.), John Wiley & Sons, 2003
From Data Mining to Knowledge Discovery in Databases	Bibliografía	Artículo que sienta las bases de la Minería de Datos. U. Fayyad, G. Piatetsky-Shapiro y P. Smyth, 1996
WEKA	Recursos web	Official webpage of the Data Mining Tool WEKA, with tutorials and free download http://www.cs.waikato.ac.nz/ml/weka/
Moodle	Otros	

8. Otra información

8.1. Otra información sobre la asignatura

Durante el curso, se utilizarán tantos 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.