



UNIVERSIDAD
POLITÉCNICA
DE MADRID

PROCESO DE
COORDINACIÓN DE LAS
ENSEÑANZAS PR/CL/001

Escuela Politécnica de
Enseñanza Superior

ANX-PR/CL/001-01

GUÍA DE APRENDIZAJE

ASIGNATURA

305000140 - Introducción A La Ciencia De Datos

PLAN DE ESTUDIOS

30GM - Grado En Matematicas

CURSO ACADÉMICO Y SEMESTRE

2024/25 - Primer 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.....	3
6. Cronograma.....	7
7. Actividades y criterios de evaluación.....	11
8. Recursos didácticos.....	14

1. Datos descriptivos

1.1. Datos de la asignatura

Nombre de la asignatura	305000140 - Introducción a la Ciencia de Datos
No de créditos	6 ECTS
Carácter	Optativa
Curso	Cuarto curso
Semestre	Séptimo semestre
Período de impartición	Septiembre-Enero
Idioma de impartición	Castellano
Titulación	30GM - Grado en Matematicas
Centro responsable de la titulación	30 - Escuela Politecnica De Enseñanza Superior
Curso académico	2024-25

2. Profesorado

2.1. Profesorado implicado en la docencia

Nombre	Despacho	Correo electrónico	Horario de tutorías *
Javier De Vicente Buendia (Coordinador/a)	A-315 ETSIAE	fj.devicente@upm.es	L - 10:30 - 12:30 J - 11:00 - 13:00 J - 15:00 - 17:00
Jesus Garicano Mena	A-314 ETSIAE	jesus.garicano.mena@upm.es	Sin horario.

* 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

- Álgebra Lineal
- Estadística
- Cálculo En Una Variable
- Programación
- Cálculo En Varias Variables

3.2. Otros conocimientos previos recomendados para cursar la asignatura

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

4. Competencias y resultados de aprendizaje

4.1. Competencias

CB1 - Que los estudiantes hayan demostrado poseer y comprender conocimientos en un área de estudio que parte de la base de la educación secundaria general, y se suele encontrar a un nivel que, si bien se apoya en libros de texto avanzados, incluye también algunos aspectos que implican conocimientos procedentes de la vanguardia de su campo de estudio

CB2 - Que los estudiantes sepan aplicar sus conocimientos a su trabajo o vocación de una forma profesional y posean las competencias que suelen demostrarse por medio de la elaboración y defensa de argumentos y la resolución de problemas dentro de su área de estudio

CB3 - Que los estudiantes tengan la capacidad de reunir e interpretar datos relevantes (normalmente dentro de su área de estudio) para emitir juicios que incluyan una reflexión sobre temas relevantes de índole social, científica o ética

CB4 - Que los estudiantes puedan transmitir información, ideas, problemas y soluciones a un público tanto especializado como no especializado

CB5 - Que los estudiantes hayan desarrollado aquellas habilidades de aprendizaje necesarias para emprender estudios posteriores con un alto grado de autonomía

CE4 - Abstraer las propiedades estructurales de objetos matemáticos, de la realidad observada o de otros ámbitos distinguiéndolas de aquellas puramente ocasionales.

CE7 - Resolver problemas de Matemáticas, mediante habilidades de cálculo básico y tecnologías de computación, planificando su resolución en función de las herramientas de que se disponga y de las restricciones de tiempo y recursos.

CE8 - Utilizar aplicaciones informáticas de análisis estadístico, cálculo numérico y simbólico, visualización gráfica, optimización u otras para experimentar en Matemáticas, buscar soluciones y resolver modelos matemáticos de sistemas reales.

4.2. Resultados del aprendizaje

RA132 - Realizar e interpretar simulaciones de modelos matemáticos sencillos mediante el lenguaje de programación adquirido en la asignatura de programación.

RA147 - Conocer y manejar las herramientas de visualización de resultados que más se adapten a cada situación.

RA178 - Construir un modelo matemático de un fenómeno real. Identificar todos los componentes de este. Evaluar la calidad del modelo e interpretar los resultados.

RA202 - Analizar e interpretar las posibles relaciones estadísticas entre varias variables

5. Descripción de la asignatura y temario

5.1. Descripción de la asignatura

Introducción a la Ciencia de Datos

La asignatura de Introducción a la Ciencia de Datos tiene como objetivo fundamental dotar a los estudiantes de las herramientas y conocimientos básicos necesarios para adentrarse en esta disciplina.

La Ciencia de Datos se caracteriza por la integración de técnicas provenientes de diversas áreas, como las matemáticas, la estadística y la programación, para extraer información valiosa a partir de datos.

En este sentido, la asignatura comienza con una revisión de los conceptos esenciales de estas tres áreas,

sentando las bases para un aprendizaje sólido.

Análisis de Datos: La base de la Ciencia de Datos

El primer bloque temático se centra en el análisis de datos, una etapa crucial que transforma la información en bruto en datos estructurados y analizables. En este contexto, se introduce el concepto de "Análisis exploratorio de datos", que permite conocer los diferentes tipos de datos y formatos en los que se pueden presentar. A continuación, se exploran diversas herramientas para "limpiar" los datos, identificando y corrigiendo errores, inconsistencias o datos ausentes, y evaluando la adecuación de las muestras. Finalmente, se abordan técnicas de reducción de dimensionalidad, cuyo objetivo es disminuir el número de variables en un conjunto de datos sin perder información relevante.

Descifrando los datos: Técnicas de Ciencia de Datos

Los siguientes bloques se adentran en las técnicas empleadas en la Ciencia de Datos, comenzando con dos enfoques con características comunes pero metodologías distintas: las técnicas de aprendizaje automático y las técnicas de descomposición modal basadas en datos. Ambos enfoques involucran optimización y aprendizaje a partir de datos, aunque difieren en sus objetivos y metodologías.

Las técnicas de aprendizaje automático se enfocan en identificar patrones y relaciones a partir de una muestra de datos para realizar predicciones o tomar decisiones fundamentadas. Estas técnicas generalmente implican entrenar un modelo en un conjunto de datos para capturar los patrones subyacentes. El modelo entrenado puede usarse luego para realizar predicciones o clasificaciones de nuevos datos. Ejemplos de técnicas de aprendizaje automático incluyen algoritmos de aprendizaje supervisado como regresión lineal, regresión logística y máquinas de vectores de soporte, así como algoritmos de aprendizaje no supervisado como agrupamiento de k-means y análisis de componentes principales.

En cuanto a las técnicas de descomposición modal basadas en datos, se realiza un análisis introductorio para comprender estos algoritmos que permiten describir relaciones entre los datos, reducir la dimensionalidad o detectar las estructuras que subyacen en un fenómeno. Se presentan DMD y POD como ejemplos representativos de este tipo de técnicas.

Aplicación de la Ciencia de Datos en el mundo real

El objetivo final de la asignatura es que los estudiantes adquieran las competencias necesarias para aplicar la Ciencia de Datos a la resolución de problemas reales en diversos ámbitos. Para ello, se presentarán estudios de casos y proyectos que abordarán problemas del mundo real en áreas como el procesamiento de señales de sistemas dinámicos, la recomendación y la detección de anomalías o patrones, el desarrollo de modelos predictivos, modelos reducidos, aplicaciones aeronáuticas, médicas, neurociencia, entre otras.

5.2. Temario de la asignatura

1. Introducción a la Ciencia de Datos

1.1. Conceptos de Programación

1.1.1. Librerías Numpy-Scipy

1.1.2. Librería Pandas

1.1.3. Librerías Matplotlib-Seaborn

1.1.4. Librería Sklearn

1.1.5. Programación Orientada a Objetos

1.2. Repaso Conceptos de Estadística

2. Conceptos Básicos en el Análisis de Datos.

2.1. Preproceso de Datos

2.2. Postproceso de Datos

3. Métodos de Aprendizaje Supervisado I

3.1. Regresión Lineal

3.2. Regresión Logística

3.3. Árboles de Decisión y Bosques Aleatorios

3.4. Maquinas de Vector Soporte

4. Métodos de Aprendizaje no Supervisado

4.1. Calsificación Jerárquica

4.2. K-means

4.3. DBSCAN

5. Métodos de Aprendizaje Supervisado II

5.1. Redes Neuronales

6. Métodos de descomposición modal basados en datos

6.1. DMD

6.2. POD

6. Cronograma

6.1. Cronograma de la asignatura *

Sem	Actividad tipo 1	Actividad tipo 2	Tele-enseñanza	Actividades de evaluación
1	<p>Introducción a la asignatura Duración: 01:00 LM: Actividad del tipo Lección Magistral</p> <p>Conceptos básicos de Programación Duración: 02:00 LM: Actividad del tipo Lección Magistral</p> <p>Conceptos básicos de programación Duración: 02:00 PR: Actividad del tipo Clase de Problemas</p>			
2	<p>Repaso Estadística Duración: 01:00 LM: Actividad del tipo Lección Magistral</p> <p>Tipos de Datos Duración: 02:00 LM: Actividad del tipo Lección Magistral</p> <p>Práctica Tipos de Datos Duración: 02:00 PR: Actividad del tipo Clase de Problemas</p>			
3	<p>Introducción al Aprendizaje Automático Duración: 01:00 LM: Actividad del tipo Lección Magistral</p> <p>Práctica Evaluable de Tipos de Datos Duración: 02:00 OT: Otras actividades formativas / Evaluación</p> <p>Regresión Lineal Duración: 02:00 LM: Actividad del tipo Lección Magistral</p>			<p>Práctica Evaluable de Tipos de Datos EP: Técnica del tipo Examen de Prácticas Evaluación Progresiva Presencial Duración: 02:00</p>
4	<p>Regresión polinómica Duración: 01:00 LM: Actividad del tipo Lección Magistral</p> <p>Regularización en Regresión Duración: 02:00 LM: Actividad del tipo Lección Magistral</p> <p>Practica Regresión Duración: 02:00 PR: Actividad del tipo Clase de Problemas</p>			

5	<p>Regresión Logística Duración: 01:00 LM: Actividad del tipo Lección Magistral</p> <p>Regresión Logística Duración: 02:00 LM: Actividad del tipo Lección Magistral</p> <p>Práctica Evaluable Regresión Lineal Duración: 02:00 OT: Otras actividades formativas / Evaluación</p>			<p>Práctica Evaluable de Regresión Lineal EP: Técnica del tipo Examen de Prácticas Evaluación Progresiva Presencial Duración: 02:00</p>
6	<p>Árboles de Decisión Duración: 01:00 LM: Actividad del tipo Lección Magistral</p> <p>Random Forest Duración: 02:00 LM: Actividad del tipo Lección Magistral</p> <p>Práctica Árboles de Decisión y Random Forest Duración: 02:00 PR: Actividad del tipo Clase de Problemas</p>			
7	<p>Máquinas de Vector Soporte Duración: 01:00 LM: Actividad del tipo Lección Magistral</p> <p>Práctica SVM Duración: 02:00 PR: Actividad del tipo Clase de Problemas</p> <p>K-NN Duración: 01:00 LM: Actividad del tipo Lección Magistral</p> <p>Práctica k-NN Duración: 01:00 PR: Actividad del tipo Clase de Problemas</p>			<p>Trabajo en grupo Aprendizaje Automático Supervisado TG: Técnica del tipo Trabajo en Grupo Evaluación Progresiva No presencial Duración: 02:00</p>
8	<p>Aprendizaje no Supervisado Duración: 01:00 LM: Actividad del tipo Lección Magistral</p> <p>HC, K-means y DBSCAN Duración: 02:00 LM: Actividad del tipo Lección Magistral</p> <p>Práctica Aprendizaje no supervisado Duración: 02:00 PR: Actividad del tipo Clase de Problemas</p>			
9	<p>Importancia de las características Duración: 01:00 LM: Actividad del tipo Lección Magistral</p> <p>Reducción de dimensionalidad Duración: 02:00 LM: Actividad del tipo Lección Magistral</p> <p>Práctica Reducción Dimensionalidad Duración: 02:00 PR: Actividad del tipo Clase de Problemas</p>			<p>Trabajo en grupo Aprendizaje Automático no Supervisado TG: Técnica del tipo Trabajo en Grupo Evaluación Progresiva No presencial Duración: 02:00</p>

10	<p>Introducción a las Redes Neuronales Duración: 01:00 LM: Actividad del tipo Lección Magistral</p> <p>Redes Neuronales Duración: 02:00 LM: Actividad del tipo Lección Magistral</p> <p>Práctica Redes Neuronales Duración: 02:00 PR: Actividad del tipo Clase de Problemas</p>			
11	<p>Método Backpropagation Duración: 01:00 LM: Actividad del tipo Lección Magistral</p> <p>Redes Neuronales con Keras Duración: 02:00 LM: Actividad del tipo Lección Magistral</p> <p>Práctica Redes Neuronales Keras Duración: 02:00 PR: Actividad del tipo Clase de Problemas</p>			
12	<p>Introducción Descomposición Modal basada en Datos Duración: 01:00 LM: Actividad del tipo Lección Magistral</p> <p>DMD Duración: 02:00 LM: Actividad del tipo Lección Magistral</p> <p>Práctica DMD Duración: 02:00 PR: Actividad del tipo Clase de Problemas</p>			<p>Trabajo en grupo Redes Neuronales TG: Técnica del tipo Trabajo en Grupo Evaluación Progresiva No presencial Duración: 02:00</p>
13	<p>Repaso SVD Duración: 01:00 LM: Actividad del tipo Lección Magistral</p> <p>POD Duración: 02:00 LM: Actividad del tipo Lección Magistral</p> <p>Prácticas POD Duración: 02:00 PR: Actividad del tipo Clase de Problemas</p>			<p>Trabajo en grupo Descomposición Modal TG: Técnica del tipo Trabajo en Grupo Evaluación Progresiva No presencial Duración: 02:00</p>
14	<p>Presentación Trabajos Grupales Duración: 01:00 OT: Otras actividades formativas / Evaluación</p> <p>Presentación Trabajos Grupales Duración: 02:00 OT: Otras actividades formativas / Evaluación</p>			<p>Presentación Trabajos Grupales PG: Técnica del tipo Presentación en Grupo Evaluación Progresiva Presencial Duración: 03:00</p>
15				

16				
17				Examen Ordinario EX: Técnica del tipo Examen Escrito Evaluación Global Presencial Duración: 03:00

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.

7. Actividades y criterios de evaluación

7.1. Actividades de evaluación de la asignatura

7.1.1. Evaluación (progresiva)

Sem.	Descripción	Modalidad	Tipo	Duración	Peso en la nota	Nota mínima	Competencias evaluadas
3	Práctica Evaluable de Tipos de Datos	EP: Técnica del tipo Examen de Prácticas	Presencial	02:00	10%	5 / 10	CB1 CB3 CE4 CE8
5	Práctica Evaluable de Regresión Lineal	EP: Técnica del tipo Examen de Prácticas	Presencial	02:00	10%	5 / 10	CB1 CB5 CE4 CE7 CE8
7	Trabajo en grupo Aprendizaje Automático Supervisado	TG: Técnica del tipo Trabajo en Grupo	No Presencial	02:00	15%	5 / 10	CB1 CB3 CB4 CB5 CE4 CE7 CE8
9	Trabajo en grupo Aprendizaje Automático no Supervisado	TG: Técnica del tipo Trabajo en Grupo	No Presencial	02:00	15%	5 / 10	CB1 CB3 CB4 CB5 CE4 CE7 CE8
12	Trabajo en grupo Redes Neuronales	TG: Técnica del tipo Trabajo en Grupo	No Presencial	02:00	15%	5 / 10	CB1 CB3 CB4 CB5 CE4 CE7 CE8
13	Trabajo en grupo Descomposición Modal	TG: Técnica del tipo Trabajo en Grupo	No Presencial	02:00	15%	5 / 10	

14	Presentación Trabajos Grupales	PG: Técnica del tipo Presentación en Grupo	Presencial	03:00	20%	5 / 10	CB1 CB2 CB3 CB4 CB5 CE4 CE7 CE8
----	--------------------------------	--	------------	-------	-----	--------	--

7.1.2. Prueba evaluación global

Sem	Descripción	Modalidad	Tipo	Duración	Peso en la nota	Nota mínima	Competencias evaluadas
17	Examen Ordinario	EX: Técnica del tipo Examen Escrito	Presencial	03:00	100%	5 / 10	CB1 CB2 CB3 CB4 CB5 CE4 CE7 CE8

7.1.3. Evaluación convocatoria extraordinaria

Descripción	Modalidad	Tipo	Duración	Peso en la nota	Nota mínima	Competencias evaluadas
Examen Extraordinario	EX: Técnica del tipo Examen Escrito	Presencial	03:00	100%	5 / 10	CB1 CB2 CB3 CB4 CB5 CE4 CE7 CE8

7.2. Criterios de evaluación

Se establecerá un sistema de evaluación progresiva durante el periodo de convocatoria ordinaria.

Esta evaluación progresiva constará de, al menos, 2 pruebas individuales y 4 entregas en grupo.

Las pruebas individuales tendrán el formato de práctica evaluable con entrega de informe y se concentrarán en la parte inicial de la asignatura. Cada una de ellas constituye el 10% de la nota de evaluación progresiva.

Las entregas de grupo se organizarán alrededor de los bloques temáticos comentados en el apartado anterior.

Dependiendo del número de alumnos en la asignatura los grupos se constituirán con un máximo de 4 integrantes.

Para cada una de esas 4 pruebas está prevista la entrega de un informe. Además en las últimas semanas de clase cada grupo hará una exposición oral de no más de 15 minutos sobre alguno de los trabajos entregados. La nota de cada trabajo será el 15% de la evaluación progresiva mientras que la exposición oral será el 20% de la misma.

En este sistema de evaluación progresiva no habrá examen ordinario y la nota obtenida será aquella el resultado de la siguiente fórmula.

$$\text{Nota} = 0.2 * ((I_1 + I_2) / 2) + 0.8 * ((G_1 + G_2 + G_3 + G_4) / 4)$$

Para poder optar a la evaluación progresiva los alumnos deben haber presentado todas las prácticas y trabajos, así como haber participado en la presentación oral.

El aprobado en evaluación progresiva se consigue con una nota igual o superior a 5.0.

Para aquellos estudiantes que no superen la asignatura mediante la evaluación progresiva habrá una prueba final que constituirá el 100% de la nota.

8. Recursos didácticos

8.1. Recursos didácticos de la asignatura

Nombre	Tipo	Observaciones
Moodle de la asignatura	Recursos web	En la página de Moodle de la asignatura se alojarán las presentaciones de los diferentes temas tratados.
Repositorio de código	Recursos web	Se pondrá a disposición de los alumnos un repositorio con los diferentes códigos realizados en clase
An Introduction to Statistical Learning. Daniela Witten, Gareth M. James, Trevor Hastie, Robert Tibshirani. 2013 Springer Science & Business Media	Bibliografía	El libro ofrece una visión general del campo del aprendizaje estadístico. Este libro presenta algunas de las técnicas de modelado y predicción más importantes, junto con aplicaciones relevantes.
Introduction to Computation and Programming Using Python: With Application to Understanding Data. John Guttag. 2nd ed. MIT Press, 2016	Bibliografía	