

FUNDAMENTOS EN CIENCIA DE DATOS E INTELIGENCIA ARTIFICIAL: PREPROCESAMIENTO Y ANÁLISIS EXPLORATORIO DE DATOS (EDA)

1.- Datos de la Asignatura

Código	306700	Plan	M207	ECTS	6
Carácter	Básica	Curso	Máster	Periodicidad	1º Semestre
Idioma de impartición asignatura		Castellano			
Área	Arquitectura y Tecnología de Computadores				
Departamento	Informática y Automática				
Plataforma virtual	Campus Virtual de la Universidad de Salamanca				

1.1.- Datos del profesorado

Profesor Coordinador	Miguel Ángel Conde González	Grupo / s	1
Departamento	Informática y Automática		
Área	Arquitectura y Tecnología de Computadores		
Centro	Escuela Politécnica Superior de Zamora		
Despacho	220 - Administrativo		
Horario de tutorías	https://politecnica zamora.usal.es/wp-content/uploads/sites/78/2025/01/Tutorias-V3D.pdf		
URL Web	https://produccioncientifica.usal.es/investigadores/1225833/detalle		
E-mail	mconde@usal.es	Teléfono	923 29 45 00 ext. 3746

1.1.- Datos del profesorado

Profesor Coordinador	María Concepción Vega Hernández	Grupo / s	1
Departamento	Estadística		
Área	Estadística e Investigación Operativa		
Centro	E. Politécnica Superior de Zamora		
Despacho	208 - Edificio Administrativo		
Horario de tutorías	A convenir con el estudiante		
URL Web	https://produccioncientifica.usal.es/investigadores/57779/detalle		
E-mail	mvegahdz@usal.es	Teléfono	923 294400 Ext. 6980

2.- Recomendaciones previas

No es necesaria ninguna recomendación previa.

3.- Objetivos de la asignatura

La asignatura tiene como objetivo proporcionar al estudiante los conocimientos y competencias básicas necesarias para trabajar con datos de forma rigurosa y eficiente en contextos de ingeniería y aplicaciones de inteligencia artificial.

A lo largo de la asignatura, el alumnado aprenderá a comprender la naturaleza de los datos, identificar problemas de calidad y aplicar técnicas de preparación y transformación que permitan obtener conjuntos de datos adecuados para posteriores procesos analíticos y de modelado. Asimismo, se introducen los fundamentos del análisis exploratorio de datos (EDA) como herramienta para descubrir patrones, relaciones, tendencias y anomalías mediante técnicas estadísticas básicas y visualización de datos.

La materia combina fundamentos teóricos con una orientación práctica basada en el uso de herramientas ampliamente utilizadas en ciencia de datos, permitiendo al estudiante desarrollar habilidades para manipular, analizar y representar información de manera efectiva. La asignatura constituye una base esencial para el resto de materias del máster, ya que proporciona las competencias necesarias para abordar de forma adecuada el tratamiento y comprensión de los datos en proyectos de ciencia de datos e inteligencia artificial aplicada a la ingeniería.

4.- Competencias a adquirir / Resultados de aprendizaje

Competencias <i>Complete esta columna si su titulación no ha sido adaptada al RD822/2021</i>	Resultados de aprendizaje <i>Complete esta columna si su titulación ha sido adaptada al RD822/2021</i>
4.1: Competencias Básicas:	4.1: Conocimientos: C1. Explicar y fundamentar los principios estadísticos y probabilísticos aplicables al tratamiento e interpretación de datos en contextos de ingeniería complejos. C2. Identificar y describir los lenguajes y librerías de programación avanzada utilizados en el desarrollo de soluciones algorítmicas para ciencia de datos e IA tales como (Python, R, SQL, etc.). C8. Identificar y comparar herramientas de visualización de datos e inteligencia de negocio aplicables al análisis y diagnóstico de datos complejos. C9. Evaluar los riesgos éticos, legales y sociales asociados al uso de IA y datos, aplicando criterios de responsabilidad, transparencia y seguridad en el diseño de soluciones tecnológicas.
4.2: Competencias Específicas:	4.2: Habilidades: H1. Programar y depurar algoritmos de procesamiento y análisis de datos en entornos técnicos utilizando herramientas como Python, R, SQL y plataformas de Big Data. H4. Construir dashboards interactivos y visualizaciones avanzadas que faciliten la toma de decisiones basada en datos para diferentes perfiles técnicos y de gestión, en procesos de ingeniería. H6. Seleccionar e implementar arquitecturas de almacenamiento y procesamiento distribuido, incluyendo tecnologías como Hadoop, Spark, y bases de datos NoSQL.
4.3: Competencias Transversales:	4.3: Competencias: K1. Integrar conocimientos multidisciplinares de datos, inteligencia artificial e ingeniería para resolver problemas técnicos complejos en entornos nuevos o emergentes.

	K2. Analizar y sintetizar información técnica y científica, interpretando datos complejos con pensamiento crítico y rigor metodológico.
--	---

5.- Contenidos (temario)

Tema 1. Introducción a la Ciencia de Datos y la IA aplicada a la ingeniería

Tema 2. Manipulación y estructuras de datos con NumPy y Pandas

Tema 3. Calidad y preprocesamiento de datos

Tema 4. Análisis estadístico descriptivo

Tema 5. Análisis Exploratorio de Datos (EDA)

Tema 6. Visualización de datos

Tema 7. Introducción al uso de Scikit-learn para preprocesamiento

Tema 8. Casos de aplicación en ingeniería

Programa de Prácticas

Las actividades prácticas de la asignatura estarán orientadas a reforzar y aplicar los conceptos fundamentales relacionados con el preprocesamiento, análisis y visualización de datos en contextos de ciencia de datos e inteligencia artificial aplicada a la ingeniería. A través de ejercicios guiados y casos prácticos, el estudiantado trabajará con conjuntos de datos reales o simulados, aprendiendo a realizar tareas de limpieza, transformación, análisis estadístico básico y análisis exploratorio de datos (EDA). Asimismo, se introducirán herramientas y bibliotecas ampliamente utilizadas en el ámbito profesional y académico, como NumPy, Pandas, Matplotlib y Scikit-learn, que permitirán desarrollar competencias prácticas para la manipulación, comprensión y preparación de datos de cara a su posterior utilización en técnicas de aprendizaje automático e inteligencia artificial.

6.- Metodologías docentes

La asignatura se desarrolla con el objetivo de que el estudiantado adquiera conocimientos teóricos y prácticos fundamentales en el ámbito del preprocesamiento y análisis exploratorio de datos aplicados a la ciencia de datos y la inteligencia artificial en ingeniería, mediante una metodología orientada al aprendizaje progresivo y a la aplicación práctica continua de los conceptos trabajados. Dado su carácter de asignatura básica dentro del máster, se favorecerá la adquisición de competencias instrumentales relacionadas con la manipulación, limpieza, análisis y visualización de datos, permitiendo al alumnado afrontar con garantías asignaturas posteriores más avanzadas del ámbito de la inteligencia artificial y el aprendizaje automático. Para ello, se combinarán contenidos conceptuales con ejercicios prácticos, actividades guiadas y casos aplicados utilizando herramientas y bibliotecas ampliamente empleadas en ciencia de datos.

Interacción con el estudiante

Se fomentará la interacción del estudiantado a través de diferentes mecanismos propios de la enseñanza online:

- **Clases sincrónicas:** dado el carácter teórico-práctico de la asignatura, el profesor estará presente en clases sincrónicas para resolver las posibles dudas de los alumnos y guiarlos en la realización de los ejercicios.
- **Materiales docentes y contenidos multimedia:** el alumnado dispondrá de recursos teóricos, ejemplos prácticos, videotutoriales y cuadernos interactivos que facilitarán el aprendizaje autónomo y progresivo de los contenidos de la asignatura.
- **Tutorías y seguimiento docente:** el profesorado ofrecerá tutorías virtuales sincrónicas o asincrónicas para resolver dudas relacionadas con los contenidos, ejercicios prácticos y herramientas de programación empleadas.

MODELO ÚNICO de guía docente de asignaturas de Grado y Máster Universitario

- **Campus virtual:** se utilizará la plataforma de enseñanza virtual del máster como espacio principal de interacción, donde estarán disponibles materiales docentes, actividades, prácticas, foros de discusión y canales de comunicación entre profesorado y estudiantado. Asimismo, las entregas de ejercicios y actividades de evaluación se realizarán a través de esta plataforma.

6.1.- Distribución de metodologías docentes

		Horas dirigidas por el profesor		Horas de trabajo autónomo	HORAS TOTALES
		Horas presenciales.	Horas no presenciales.		
Sesiones magistrales			12	24	36
Prácticas	En aula				
	En el laboratorio				
	En entorno virtual	12		24	36
	De campo				
	De visualización (visu)				
Seminarios			6	4	10
Exposiciones y debates					
Tutorías			6		6
Actividades de seguimiento online			10	20	30
Preparación de trabajos y ejercicios			6	10	16
Estudio autónomo de materiales				10	10
Exámenes/pruebas de evaluación			6		6
TOTAL		12	46	92	150

7.- Recursos, bibliografía, referencias electrónicas o de otro tipo

Bibliografía básica

- Russell, S. J., & Norvig, P. (2022). Inteligencia artificial: Un enfoque moderno (4.^a ed.). Pearson.
- Jafari, R. (2022). Hands-On Data Preprocessing in Python: Learn how to Effectively Prepare Data for Successful Data Analytics. Alemania: Packt Publishing.
- McKinney, W. (2022). Python for Data Analysis: Data Wrangling with Pandas, NumPy, and Jupyter (3rd ed.). O'Reilly Media.
- VanderPlas, J. (2016). Python Data Science Handbook: Essential Tools for Working with Data. O'Reilly Media.
- Bruce, P., Bruce, A., & Gedeck, P. (2020). Practical Statistics for Data Scientists: 50+ Essential Concepts Using R and Python (2nd ed.). O'Reilly Media.
- Grus, J. (2019). Data Science from Scratch: First Principles with Python (2nd ed.). O'Reilly Media.
- Downey, A. (2014). Think Stats: Exploratory Data Analysis (2nd ed.). O'Reilly Media.

Bibliografía complementaria

- Géron, A. (2020). Aprendizaje automático práctico con Scikit-Learn, Keras y TensorFlow (2.^a ed.). O'Reilly Media.
- Raschka, S., & Mirjalili, V. (2019). Python machine learning: Machine learning y deep learning con Python, scikit-learn y TensorFlow (3.^a ed.). Packt Publishing.

MODELO ÚNICO de guía docente de asignaturas de Grado y Máster Universitario

- Kelleher, J. D., Mac Namee, B., & D'Arcy, A. (2020). Fundamentals of machine learning for predictive data analytics: Algorithms, worked examples, and case studies (2nd ed.). MIT Press.
- Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). Data mining: Practical machine learning tools and techniques (4th ed.). Morgan Kaufmann.
- Chollet, F. (2021). Deep learning with Python (2nd ed.). Manning Publications.

8.- Evaluación

8.1: Criterios de evaluación:

Se utilizará el sistema de calificaciones vigente (RD 1125/2003) artículo 5º. Los resultados obtenidos por el alumno en cada una de las materias del plan de estudios se calificarán en función de la siguiente escala numérica de 0 a 10, con expresión de un decimal, a la que podrá añadirse su correspondiente calificación cualitativa: 0 - 4,9: Suspenso (SS), 5,0 - 6,9: Aprobado (AP), 7,0 - 8,9: Notable (NT), 9,0 - 10: Sobresaliente (SB). La mención de Matrícula de Honor podrá ser otorgada a alumnos que hayan obtenido una calificación igual o superior a 9,0. Su número no podrá exceder del 5% de los alumnos matriculados en una asignatura en el correspondiente curso académico, salvo que el número de alumnos matriculados sea inferior a 20, en cuyo caso se podrá conceder una sola Matrícula de Honor.

Se tendrá en cuenta el Reglamento de Evaluación de la Universidad de Salamanca.

8.2: Sistemas de evaluación:

Evaluación continua: seguimiento de la evolución en clase del alumno, participación en clase, prácticas y trabajos realizados (incluyendo defensa de estos).
Exámenes teórico-prácticos.

8.3: Consideraciones generales y recomendaciones para la evaluación y la recuperación:

Primera Convocatoria		
Prácticas	Proyecto	Continua
40%	40%	20%
Segunda Convocatoria		
Prácticas	Proyecto	
50%	50%	

Recomendaciones evaluación:

La participación del alumno unido al trabajo continuo permite superar sin dificultad la asignatura.

Recomendaciones recuperación:

A cada alumno se le indicará, de forma individualizada, qué partes de la asignatura debe reforzar para poder superarla.

9.- Organización docente semanal