



UNIVERSIDAD
POLITÉCNICA
DE MADRID

PROCESO DE
COORDINACIÓN DE LAS
ENSEÑANZAS PR/CL/001



E.T.S. de Ingenieros de
Telecomunicacion

ANX-PR/CL/001-01

GUÍA DE APRENDIZAJE

ASIGNATURA

95000514 - Programación Para Big Data

PLAN DE ESTUDIOS

09ID - Grado En Ingeniería Y Sistemas De Datos

CURSO ACADÉMICO Y SEMESTRE

2021/22 - 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.....	4
6. Cronograma.....	6
7. Actividades y criterios de evaluación.....	9
8. Recursos didácticos.....	13
9. Otra información.....	14

1. Datos descriptivos

1.1. Datos de la asignatura

Nombre de la asignatura	95000514 - Programación para Big Data
No de créditos	6 ECTS
Carácter	Obligatoria
Curso	Segundo curso
Semestre	Tercer semestre
Período de impartición	Septiembre-Enero
Idioma de impartición	Castellano
Titulación	09ID - Grado en Ingeniería y Sistemas de Datos
Centro responsable de la titulación	09 - Escuela Técnica Superior De Ingenieros De Telecomunicacion
Curso académico	2021-22

2. Profesorado

2.1. Profesorado implicado en la docencia

Nombre	Despacho	Correo electrónico	Horario de tutorías *
Juan Fernando Sanchez Rada (Coordinador/a)	B207	jf.sanchez@upm.es	Sin horario. Contactar previamente por e-mail.
Oscar Araque Iborra	C212	o.araque@upm.es	Sin horario. Contactar previamente por e-mail.

* Las horas de tutoría son orientativas y pueden sufrir modificaciones. Se deberá confirmar los horarios de tutorías

con el profesorado.

2.2. Personal investigador en formación o similar

Nombre	Correo electrónico	Profesor responsable
Tardio Rubio, Jorge	j.tardio@upm.es	Sanchez Rada, Juan Fernando

3. Conocimientos previos recomendados

3.1. Asignaturas previas que se recomienda haber cursado

- Bases De Datos Relacionales Y Datos Estructurados
- Programación

3.2. Otros conocimientos previos recomendados para cursar la asignatura

El plan de estudios Grado en Ingeniería y Sistemas de Datos no tiene definidos otros conocimientos previos para esta asignatura.

4. Competencias y resultados de aprendizaje

4.1. Competencias

CB01 - 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

CB02 - 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

CB03 - 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

CB04 - Que los estudiantes puedan transmitir información, ideas, problemas y soluciones a un público tanto

especializado como no especializado

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

CE07 - Que los estudiantes sepan desplegar, configurar y utilizar infraestructuras de computación conectadas de altas prestaciones para el almacenamiento y tratamiento de datos, en el ámbito de la ingeniería de telecomunicación, tanto en la nube como en sistemas locales y centros de procesado de datos.

CE09 - Que los estudiantes tengan la capacidad de aplicar las características, funcionalidades y estructura de Internet y las redes de ordenadores a la construcción de infraestructuras e integración de aplicaciones telemáticas y servicios.

CE10 - Que los estudiantes tengan la capacidad de construir sistemas, aplicaciones y servicios telemáticos, interconectados y multiplataforma a partir de su comprensión de la arquitectura web.

CE17 - Que los estudiantes tengan la capacidad de utilizar los fundamentos de la programación, sistemas operativos, bases de datos, tecnología web y las redes y servicios de telecomunicación en proyectos de ingeniería de datos y sistemas.

CG01 - Tener capacidad de trabajar en entornos internacionales y multidisciplinares, haciendo uso de la lengua inglesa en forma oral y escrita.

CG02 - Desarrollar la capacidad de trabajar en equipo empleando metodologías ágiles para diseñar soluciones eficientes, fiables y robustas.

CG03 - Ser capaz de explicar de forma oral o escrita las soluciones planteadas para la resolución de un problema.

CG04 - Saber identificar y utilizar las herramientas de las Tecnologías de la Información y de las Comunicaciones más adecuadas para plantear y construir soluciones a problemas

CG05 - Tener la capacidad de concebir y proponer soluciones creativas aplicando los métodos científico y de ingeniería para la definición y resolución de problemas formalizando los objetivos buscados y considerando los recursos disponibles.

CG09 - Desarrollar la capacidad de aprendizaje a lo largo de la vida (lifelong learning) para adaptarse a un sector tecnológico en continua evolución.

4.2. Resultados del aprendizaje

RA057 - Comprender los retos y problemas de un análisis de Big Data.

RA058 - Comprender y evaluar la complejidad de un programa de procesado de Big Data, así como tomar decisiones para su adecuada implementación.

RA061 - Diseñar e implementar una solución completa a un caso de uso.

RA059 - Conocer y utilizar las bibliotecas de más amplio uso en el ámbito de Big Data en el lenguaje de programación seleccionado.

RA060 - Saber filtrar y limpiar datos usando bibliotecas y facilidades del lenguaje de programación más adecuado.

5. Descripción de la asignatura y temario

5.1. Descripción de la asignatura

La ingeniería de datos requiere programar aplicaciones que manejen eficientemente grandes conjuntos de datos con el objetivo de extraer información útil para la toma de decisiones y la mejor comprensión de los escenarios de actividad.

El modelo de trabajo se puede organizar en varias fases que trabajan con los datos, desde su obtención inicial hasta la generación de los resultados de interés.

En primer lugar se requieren sistemas automatizados que sean capaces de extraer datos de fuentes diversas, tanto por su formato, como por la velocidad de generación y la técnica de obtención de los datos. Habitualmente es necesario realizar un primer procesamiento de los datos en bruto para filtrarlos, limpiarlos y validarlos. Estos datos han de quedar almacenados de manera que se pueda acceder a ellos para ejecutar diversos tipos de análisis, así como para poder gestionar la adición de más información y en general el mantenimiento de los conjuntos de datos.

Una vez se dispone de un conjunto de datos adecuadamente preparado, es posible procesarlos para obtener información útil. Se pueden aplicar técnicas de carácter estadístico así como técnicas de aprendizaje automático para obtener generalizaciones y predicciones.

Para que los resultados obtenidos sean de la máxima utilidad para las personas que han de profundizar en su conocimiento y tomar decisiones, es fundamental poder presentarlos mediante herramientas que permitan resaltar

los aspectos clave y explorar sus facetas de interés.

El ingeniero de datos además ha de valorar el coste en recursos que tienen las diversas técnicas aplicables, con el fin de que los resultados se obtengan de la manera más pronta y eficiente posible.

En esta asignatura se tratarán los fundamentos de la programación de aplicaciones para la captura, preprocesamiento, almacenamiento, análisis y visualización de datos. Para cada una de estas áreas se estudiarán los fundamentos teóricos y se aprenderá a trabajar con diversas herramientas de uso habitual en el entorno de Python.

A lo largo del curso los estudiantes irán desarrollando, a nivel práctico, un caso de uso realista en el que tendrán que aplicar las diversas técnicas que se tratan en las clases.

5.2. Temario de la asignatura

1. Introducción. Sistemas de Big Data y el ciclo de vida del dato: captura, preprocesado, procesado, almacenamiento y presentación. La problemática de Big Data para procesamiento y almacenamiento
2. Entorno de programación de procesamiento de datos
3. Estructuras de datos para análisis de datos
4. Captura de datos de ficheros, bases de datos y web
5. Preprocesado, filtrado y limpieza de datos
6. Análisis de datos
7. Almacenamiento de datos
8. Aprendizaje de datos
9. Visualización de datos
10. Fundamentos de complejidad algorítmica y de evaluación de complejidad algorítmica

6. Cronograma

6.1. Cronograma de la asignatura *

Sem	Actividad presencial en aula	Actividad presencial en laboratorio	Tele-enseñanza	Actividades de evaluación
1	Introducción Duración: 01:00 LM: Actividad del tipo Lección Magistral Entorno de programación de procesado de datos Duración: 01:00 LM: Actividad del tipo Lección Magistral	Entorno de programación de procesado de datos Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio		
2	Estructuras de datos para análisis de datos Duración: 02:00 LM: Actividad del tipo Lección Magistral Estructuras de datos para análisis de datos Duración: 02:00 PR: Actividad del tipo Clase de Problemas			
3	Captura de datos Duración: 02:00 LM: Actividad del tipo Lección Magistral Captura de datos Duración: 02:00 PR: Actividad del tipo Clase de Problemas			
4	Preprocesado, filtrado y limpieza de datos Duración: 02:00 LM: Actividad del tipo Lección Magistral	Prácticas Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio		
5	Preprocesado, filtrado y limpieza de datos Duración: 02:00 LM: Actividad del tipo Lección Magistral Preprocesado, filtrado y limpieza de datos Duración: 02:00 PR: Actividad del tipo Clase de Problemas			
6	Técnicas de análisis de datos Duración: 02:00 LM: Actividad del tipo Lección Magistral	Prácticas Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio		
7	Técnicas de análisis de datos Duración: 02:00 LM: Actividad del tipo Lección Magistral Técnicas de análisis de datos Duración: 02:00 PR: Actividad del tipo Clase de Problemas			

8	<p>Técnicas de análisis de datos Duración: 01:00 LM: Actividad del tipo Lección Magistral</p> <p>Técnicas de análisis de datos Duración: 01:00 PR: Actividad del tipo Clase de Problemas</p>	<p>Prácticas Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio</p>		
9	<p>Técnicas de almacenamiento de datos Duración: 02:00 LM: Actividad del tipo Lección Magistral</p> <p>Técnicas de almacenamiento de datos Duración: 02:00 PR: Actividad del tipo Clase de Problemas</p>			<p>Parcial 1 EX: Técnica del tipo Examen Escrito Evaluación continua Presencial Duración: 01:00</p> <p>Entrega de prácticas TG: Técnica del tipo Trabajo en Grupo Evaluación continua No presencial Duración: 00:00</p>
10	<p>Visualización de datos Duración: 02:00 LM: Actividad del tipo Lección Magistral</p> <p>Visualización de datos Duración: 02:00 PR: Actividad del tipo Clase de Problemas</p>			
11	<p>Visualización de datos Duración: 01:00 LM: Actividad del tipo Lección Magistral</p> <p>Visualización de datos Duración: 01:00 PR: Actividad del tipo Clase de Problemas</p>	<p>Prácticas Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio</p>		
12	<p>Introducción al aprendizaje automático Duración: 02:00 LM: Actividad del tipo Lección Magistral</p>	<p>Prácticas Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio</p>		
13	<p>Fundamentos de complejidad algorítmica Duración: 04:00 LM: Actividad del tipo Lección Magistral</p>			
14		<p>Prácticas Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio</p>		<p>Presentaciones de las prácticas PG: Técnica del tipo Presentación en Grupo Evaluación continua Presencial Duración: 02:00</p> <p>Entrega de prácticas 2 TG: Técnica del tipo Trabajo en Grupo Evaluación continua No presencial Duración: 00:00</p>
15				
16				
17				<p>Parcial 2 EX: Técnica del tipo Examen Escrito Evaluación continua Presencial Duración: 01:00</p> <p>Examen final teoría EX: Técnica del tipo Examen Escrito Evaluación sólo prueba final</p>

				Presencial Duración: 03:00 Presentación y examen de prácticas EP: Técnica del tipo Examen de Prácticas Evaluación sólo prueba final Presencial Duración: 02:00
--	--	--	--	---

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.

7. Actividades y criterios de evaluación

7.1. Actividades de evaluación de la asignatura

7.1.1. Evaluación continua

Sem.	Descripción	Modalidad	Tipo	Duración	Peso en la nota	Nota mínima	Competencias evaluadas
9	Parcial 1	EX: Técnica del tipo Examen Escrito	Presencial	01:00	20%	4 / 10	CE17 CG04 CB05
9	Entrega de prácticas	TG: Técnica del tipo Trabajo en Grupo	No Presencial	00:00	20%	4 / 10	CB02 CB01 CG05 CG09 CE09 CE10 CE17 CG04 CG01 CG02 CB03 CB04 CB05 CG03
14	Presentaciones de las prácticas	PG: Técnica del tipo Presentación en Grupo	Presencial	02:00	20%	4 / 10	CB02 CB01 CG05 CG09 CE09 CE10 CE17 CG04 CG01 CG02 CE07 CB03 CB04 CB05 CG03

14	Entrega de prácticas 2	TG: Técnica del tipo Trabajo en Grupo	No Presencial	00:00	20%	0 / 10	
17	Parcial 2	EX: Técnica del tipo Examen Escrito	Presencial	01:00	20%	4 / 10	CB02 CB01 CE09 CE17 CG04 CB04 CB05

7.1.2. Evaluación sólo prueba final

Sem	Descripción	Modalidad	Tipo	Duración	Peso en la nota	Nota mínima	Competencias evaluadas
17	Examen final teoría	EX: Técnica del tipo Examen Escrito	Presencial	03:00	40%	4 / 10	CG05 CG09 CE09 CE10 CE17 CG04 CE07 CB03 CB04 CB05 CG03 CB02 CB01
17	Presentación y examen de prácticas	EP: Técnica del tipo Examen de Prácticas	Presencial	02:00	60%	4 / 10	CB02 CB01 CE09 CE10 CE17 CG04 CG01 CE07 CB03 CB05 CG03

7.1.3. Evaluación convocatoria extraordinaria

Descripción	Modalidad	Tipo	Duración	Peso en la nota	Nota mínima	Competencias evaluadas
Examen final extraordinario de teoría	EX: Técnica del tipo Examen Escrito	Presencial	03:00	40%	4 / 10	CB02 CB01 CG05 CG09 CE09 CE10 CE17 CG04 CE07 CB03 CB04 CB05 CG03
Presentación y examen de prácticas extraordinario	EP: Técnica del tipo Examen de Prácticas	Presencial	02:00	60%	4 / 10	CB02 CB01 CG05 CG09 CE09 CE10 CE17 CG04 CG01 CE07 CB03 CB04 CB05 CG03

7.2. Criterios de evaluación

Se plantean dos tipos de evaluación:

Evaluación continua

La evaluación continua constará de:

- dos exámenes tipo test de los temas 1-5 (20% nota) y 6-10 (20% nota).
- prácticas de los temas 1-5 (20% nota) y 6-10 (20% nota), que se desarrollarán en grupos, y que serán defendidas en una presentación final (20% nota).

Todas las notas deberán ser mayores o iguales a 4.

Evaluación por prueba final

Un examen tipo test de la asignatura (40%), y una memoria y un examen de las prácticas de la asignatura (60%). Ambas notas deberán ser mayores o iguales a 4.

La convocatoria **extraordinaria** también consistirá en un examen tipo test de la asignatura (40%), y una memoria y un examen de las prácticas de la asignatura (60%). Ambas notas deberán ser mayores o iguales a 4.

8. Recursos didácticos

8.1. Recursos didácticos de la asignatura

Nombre	Tipo	Observaciones
Practical Data Analysis Using Jupyter Notebook: Learn how to speak the language of data by extracting useful and actionable insights using Python	Bibliografía	Practical Data Analysis Using Jupyter Notebook: Learn how to speak the language of data by extracting useful and actionable insights using Python, M. Wintjen and A. Vlahutin, Packt, 2020
Data Science Projects with Python: A case study approach to successful data science projects using Python, pandas, and scikit-learn, S. Klosterman, Packt, 2019	Bibliografía	
Python for Data Analysis: Data Wrangling with Pandas, NumPy, and Ipython, W. McKinney, O'Reilly, 2017.	Bibliografía	
Pandas for Everyone: Python Data Analysis (Addison-Wesley Data & Analytics Series), 2017.	Bibliografía	
Introduction to Data Science. A Python Approach to Concepts, Techniques and Applications. Laura Igual, Santi Seguí. Springer, 2017	Bibliografía	

9. Otra información

9.1. Otra información sobre la asignatura

La asignatura de Programación para Big Data, en el contexto del Grado en Ingeniería de Sistemas de Datos, permite a los alumnos aplicar conceptos de programación y de ingeniería adquiridos previamente, de forma que puedan analizar grandes volúmenes de datos, y generar nueva información. De esta forma, esta asignatura está en condiciones de contribuir a los Objetivos de Desarrollo Sostenible (ODS) 4 y 9 de Naciones Unidas, tratando de aumentar, en lo posible, el número de personas con las competencias profesionales y técnicas necesarias para acceder al empleo y al emprendimiento (Meta 4.4), a garantizar que el alumnado adquiriera los conocimientos teóricos y prácticos necesarios para promover desarrollos sostenibles (Meta 4.7) y promover la preparación de profesionales capaces de desarrollar infraestructuras fiables, sostenibles, resilientes y de calidad (Meta 9.1).

En resumen, la asignatura se relaciona con los ODS 4 y 9:

- Subobjetivo 4.4: Aumentar considerablemente el número de jóvenes y adultos que tienen las competencias profesionales y técnicas necesarias para acceder al empleo y al emprendimiento.
- Subobjetivo 9.1: Desarrollar infraestructuras fiables, sostenibles, resilientes y de calidad