



UNIVERSIDAD  
POLITÉCNICA  
DE MADRID

PROCESO DE  
COORDINACIÓN DE LAS  
ENSEÑANZAS PR/CL/001



E.T.S. de Ingeniería y Sistemas  
de Telecomunicación

# ANX-PR/CL/001-01

## GUÍA DE APRENDIZAJE

### ASIGNATURA

**595000514 - Programación Para Big Data**

### PLAN DE ESTUDIOS

**59ID - Grado En Ingeniería Y Sistemas De Datos**

### CURSO ACADÉMICO Y SEMESTRE

**2022/23 - 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.....	14
9. Otra información.....	15

## 1. Datos descriptivos

### 1.1. Datos de la asignatura

<b>Nombre de la asignatura</b>	595000514 - Programación para Big Data
<b>No de créditos</b>	6 ECTS
<b>Carácter</b>	Obligatoria
<b>Curso</b>	Segundo curso
<b>Semestre</b>	Tercer semestre
<b>Período de impartición</b>	Septiembre-Enero
<b>Idioma de impartición</b>	Castellano
<b>Titulación</b>	59ID - Grado en Ingeniería y Sistemas de Datos
<b>Centro responsable de la titulación</b>	59 - Escuela Técnica Superior De Ingeniería Y Sistemas De Telecomunicación
<b>Curso académico</b>	2022-23

## 2. Profesorado

### 2.1. Profesorado implicado en la docencia

<b>Nombre</b>	<b>Despacho</b>	<b>Correo electrónico</b>	<b>Horario de tutorías *</b>
Fco. Javier Ramirez Ledesma	A4410	javier.ledesma@upm.es	Sin horario. Se publicarán al inicio del curso. Ver en el Moodle de la asignatura
Javier Martin Rueda (Coordinador/a)	A4410	javier.martin@upm.es	Sin horario. Se publicarán al inicio del curso. Ver en el Moodle de la asignatura

---

\* 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

- Programación
- Bases De Datos Relacionales Y Datos Estructurados

#### 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

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

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.

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

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

## 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 en aula	Actividad en laboratorio	Tele-enseñanza	Actividades de evaluación
1	<b>Introducción</b> Duración: 01:00 LM: Actividad del tipo Lección Magistral  <b>Estructuras de datos para análisis de datos</b> Duración: 01:00 LM: Actividad del tipo Lección Magistral	<b>Entorno de programación de procesado de datos.</b> Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio		
2	<b>Estructuras de datos para análisis de datos</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral  <b>Preprocesado, filtrado y limpieza de datos</b> Duración: 02:00 PR: Actividad del tipo Clase de Problemas			
3	<b>Captura de datos</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral  <b>Captura de datos</b> Duración: 02:00 PR: Actividad del tipo Clase de Problemas			
4	<b>Preprocesado, filtrado y limpieza de datos</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral	<b>Prácticas</b> Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio		
5	<b>Visualización de datos</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral  <b>Preprocesado, filtrado y limpieza de datos</b> Duración: 02:00 PR: Actividad del tipo Clase de Problemas			
6	<b>Técnicas de análisis de datos</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral	<b>Prácticas</b> Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio		
7	<b>Técnicas de análisis de datos</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral  <b>Visualización de datos</b> Duración: 02:00 PR: Actividad del tipo Clase de Problemas			<b>PARCIAL 1</b> EX: Técnica del tipo Examen Escrito Evaluación continua Presencial Duración: 02:00



8	<b>Visualización de datos</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral	<b>Prácticas</b> Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio		<b>ENTREGA DE PRÁCTICAS 1</b> EP: Técnica del tipo Examen de Prácticas Evaluación continua No presencial Duración: 00:00
9	<b>Técnicas de análisis de datos</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral <b>Técnicas de almacenamiento de datos</b> Duración: 02:00 PR: Actividad del tipo Clase de Problemas			
10	<b>Técnicas de almacenamiento de datos</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral <b>Introducción al aprendizaje automático</b> Duración: 02:00 PR: Actividad del tipo Clase de Problemas			
11	<b>Introducción al aprendizaje automático</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral	<b>Prácticas</b> Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio		
12	<b>Introducción al aprendizaje automático</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral	<b>Prácticas</b> Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio		
13	<b>Fundamentos de complejidad algorítmica</b> Duración: 02:00 LM: Actividad del tipo Lección Magistral <b>Fundamentos de complejidad algorítmica</b> Duración: 02:00 PR: Actividad del tipo Clase de Problemas			
14		<b>Prácticas</b> Duración: 02:00 PL: Actividad del tipo Prácticas de Laboratorio		<b>PRESENTACIONES</b> PG: Técnica del tipo Presentación en Grupo Evaluación continua Presencial Duración: 02:00 <b>ENTREGA DE PRÁCTICAS 2</b> EP: Técnica del tipo Examen de Prácticas Evaluación continua No presencial Duración: 00:00
15				
16				
17				<b>PARCIAL 2</b> EX: Técnica del tipo Examen Escrito Evaluación continua Presencial Duración: 02:00 <b>EXAMEN DE TEORÍA</b> EX: Técnica del tipo Examen Escrito Evaluación sólo prueba final No presencial Duración: 03:00 <b>PRESENTACIÓN Y EXAMEN DE PRÁCTICAS</b>

				EP: Técnica del tipo Examen de Prácticas Evaluación sólo prueba final No 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 (progresiva)

Sem.	Descripción	Modalidad	Tipo	Duración	Peso en la nota	Nota mínima	Competencias evaluadas
7	PARCIAL 1	EX: Técnica del tipo Examen Escrito	Presencial	02:00	20%	0 / 10	CG01 CG09 CB01 CG05 CB02 CB05 CE07 CE09 CE17 CG03 CG04 CB03 CE10
8	ENTREGA DE PRÁCTICAS 1	EP: Técnica del tipo Examen de Prácticas	No Presencial	00:00	20%	0 / 10	CG01 CG09 CB01 CG05 CB02 CB05 CE07 CE09 CE17 CG02 CG03 CG04 CB03 CE10
14	PRESENTACIONES	PG: Técnica del tipo Presentación en Grupo	Presencial	02:00	20%	0 / 10	CG01 CG09 CB01 CG05 CB02 CB04 CB05 CE07 CE09 CE17 CG02 CG03 CG04

							CB03 CE10
14	ENTREGA DE PRÁCTICAS 2	EP: Técnica del tipo Examen de Prácticas	No Presencial	00:00	20%	0 / 10	CG01 CG09 CB01 CG05 CB02 CB05 CE07 CE09 CE17 CG02 CG03 CG04 CB03 CE10
17	PARCIAL 2	EX: Técnica del tipo Examen Escrito	Presencial	02:00	20%	0 / 10	CG01 CG09 CB01 CG05 CB02 CB05 CE07 CE09 CE17 CG03 CG04 CB03 CE10

### 7.1.2. Prueba evaluación global

Sem	Descripción	Modalidad	Tipo	Duración	Peso en la nota	Nota mínima	Competencias evaluadas
17	EXAMEN DE TEORÍA	EX: Técnica del tipo Examen Escrito	No Presencial	03:00	40%	0 / 10	CG01 CG09 CB01 CG05 CB02 CB05 CE07 CE09 CE17 CG03 CG04 CB03 CE10
		EP: Técnica					CG01 CG09 CB01 CG05 CB02 CB04 CB05

17	PRESENTACIÓN Y EXAMEN DE PRÁCTICAS	del tipo Examen de Prácticas	No Presencial	02:00	60%	0 / 10	CE07 CE09 CE17 CG02 CG03 CG04 CB03 CE10
----	------------------------------------	------------------------------	---------------	-------	-----	--------	--------------------------------------------------------------

### 7.1.3. Evaluación convocatoria extraordinaria

Descripción	Modalidad	Tipo	Duración	Peso en la nota	Nota mínima	Competencias evaluadas
EXAMEN DE TEORÍA	EX: Técnica del tipo Examen Escrito	Presencial	03:00	40%	0 / 10	CG01 CG09 CB01 CG05 CB02 CB05 CE07 CE09 CE17 CG03 CG04 CB03 CE10
PRESENTACIÓN Y EXAMEN DE PRÁCTICAS	EP: Técnica del tipo Examen de Prácticas	Presencial	02:00	60%	0 / 10	CG01 CG09 CB01 CG05 CB02 CB04 CB05 CE07 CE09 CE17 CG02 CG03 CG04 CB03 CE10

## 7.2. Criterios de evaluación

### Convocatoria ordinaria

Se usará un sistema de evaluación progresiva que constará de:

- \* dos exámenes escritos, que pueden contener preguntas abiertas, de tipo test, problemas y supuestos de carácter teórico-práctico, de los temas 1-5 (20% nota) y 6-10 (20% nota).
- \* dos entregas de prácticas, una hacia la mitad del curso (20% nota) y otra al final (20% nota)
- \* una presentación a los profesores y al conjunto de los alumnos sobre la práctica realizada, en la que podrán formularse preguntas (20% de la nota)

Los exámenes serán individuales. Las prácticas y la presentación se realizarán en grupos.

### Pruebas de evaluación globales

En el mes de enero el estudiante podrá presentarse a un conjunto de pruebas de evaluación globales que constarán de:

- \* Un examen escrito que evaluará la totalidad de la asignatura y que será coincidente con el examen del segundo parcial de la evaluación progresiva (40% nota). Este examen puede contener preguntas abiertas, de tipo test, problemas y supuestos de carácter teórico-práctico.
- \* Una entrega de las prácticas realizadas a lo largo del curso, que deberá entregar unos días antes del examen global (se publicará la fecha límite con antelación suficiente) (40% nota).
- \* Una presentación oral de las prácticas ante los profesores, en la que podrán formularse preguntas (20% nota). La presentación podría realizarse en otra fecha si lo aconsejan motivos organizativos, y siempre de acuerdo con la disponibilidad del estudiante y de los profesores.

Estas pruebas serán individuales, aunque los profesores podrían admitir entregas y presentaciones de prácticas en grupo si lo consideran oportuno, y solo tras consentimiento expreso previo a los estudiantes.

### Convocatoria extraordinaria

Constará de:

- \* Un examen escrito que evaluará la totalidad de la asignatura (40% nota). Este examen puede contener preguntas abiertas, de tipo test, problemas y supuestos de carácter teórico-práctico.
- \* Una entrega de las prácticas realizadas a lo largo del curso, que deberá entregar unos días antes del examen (se publicará la fecha límite con antelación suficiente) (40% nota).
- \* Una presentación oral de las prácticas ante los profesores, en la que podrán formularse preguntas (20% nota).

Estas pruebas serán individuales, aunque los profesores podrían admitir entregas y presentaciones de prácticas en grupo si lo consideran oportuno, y solo tras consentimiento expreso previo a los estudiantes.

### **Otras consideraciones**

No existe liberación de bloques de evaluación, ni de cara a las pruebas globales de la convocatoria ordinaria, ni para la convocatoria extraordinaria, ni para cursos posteriores.

No se requiere nota mínima en ninguna de las actividades de evaluación.

## 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, M. Wintjen and A. Vlahutin, Packt, 2020	Bibliografía	
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	
Entorno de programación: Visual Studio Code, pyCharm, git, Jupyter Notebook, Anaconda. Bibliotecas numpy, pandas, matplotlib, scikit-learn, seaborn, streamlit, plotly, dash, keras, tensorflow, cuda toolkit, cuDNN...	Equipamiento	Se proporcionará preinstalado en los laboratorios de la escuela. También se darán indicaciones para que los estudiantes puedan instalarlo en sus ordenadores.



## 9. Otra información

---

### 9.1. Otra información sobre la asignatura

#### Cronograma de la presente guía

El cronograma de actividades presente en esta guía debe entenderse como planificación del orden y duración de los temas y actividades, pero no tiene en cuenta circunstancias como días festivos, comienzo de las clases a mitad de la semana, días con actividad lectiva cambiada para igualar el número de lectivos de cada día de la semana, etc. En el sitio Moodle de la asignatura se publicará un calendario con la planificación de clases concreta para el curso académico.

#### Objetivos de Desarrollo Sostenible

Se fomentará el incremento de la concienciación de los estudiantes acerca de los ODS 5 (Igualdad de género), ODS 7 (Energía asequible y no contaminante) y ODS 13 (Acción por el clima) por medio del uso en las actividades formativas de conjuntos de datos que permitan resaltar especialmente los ODS mencionados.