

FICHA TÉCNICA DE LA ASIGNATURA

Datos de la asignatura	
Nombre completo	Adquisición de Datos
Código	DTC-IMAT-211
Título	Grado en Ingeniería Matemática e Inteligencia Artificial
Impartido en	Grado en Ingeniería Matemática e Inteligencia Artificial [Segundo Curso]
Créditos	4,5 ECTS
Carácter	Obligatoria (Grado)
Departamento / Área	Departamento de Telemática y Computación

Datos del profesorado	
Profesor	
Nombre	Federico Muñoz Babiano
Departamento / Área	Departamento de Telemática y Computación
Correo electrónico	fmuoz@icai.comillas.edu
Profesor	
Nombre	Patricia Domínguez González
Departamento / Área	Departamento de Telemática y Computación
Correo electrónico	pdominguez@icai.comillas.edu

DATOS ESPECÍFICOS DE LA ASIGNATURA

Contextualización de la asignatura
Aportación al perfil profesional de la titulación
<p>Esta asignatura forma parte del bloque de asignaturas metodologías dentro del área de "Computación" junto con Visualización y Bases de datos. En este curso, conocerás los fundamentos del trabajo con datos, entendiendo la obtención de las diferentes fuentes, la limpieza y organización de los datos y su posterior uso, todo ello a través de bloques prácticos con diversos proyectos orientados a afianzar los conocimientos y poder extrapolarlos al día a día de la empresa.</p>
Prerequisitos
<p>Programación en python.</p>

Competencias - Objetivos
Competencias
GENERALES



CG04	Conocimientos básicos sobre el uso y programación de los ordenadores, sistemas operativos, bases de datos y programas informáticos con aplicación en ingeniería.
CG05	Conocimiento de la estructura, organización, funcionamiento e interconexión de los sistemas informáticos, los fundamentos de su programación, y su aplicación para la resolución de problemas propios de la ingeniería
CG07	Capacidad para integrarse en equipos de trabajo y colaborar de forma activa con otras personas, áreas y organizaciones en la consecución de los objetivos ligados a las actividades de extracción de valor de los datos e inteligencia artificial.
ESPECÍFICAS	
CE14	Dominio de los conceptos y técnicas más utilizadas de adquisición y transformación de la información localizada en local o en remoto en el ámbito del análisis de datos y la inteligencia artificial

Resultados de Aprendizaje

RA1	Conocer los procedimientos y fases para implementar correctamente un sistema de adquisición de datos
RA2	Implementar soluciones de ingesta de datos de cualquier fuente y naturaleza para su análisis posterior
RA3	Extraer, procesar y transformar los datos con el objetivo de poseer un dato de calidad: limpio, homogéneo y estandarizado
RA4	Extraer de forma programática la información deseada de los formatos de ficheros de datos más utilizados en la actualidad (TXT, CSV, XML, JSON, ..)
RA5	Conocer el sistema de etiquetado del lenguaje HTML con el objetivo de identificar la información relevante dentro de las páginas web
RA6	Dominar el proceso de automatización y extracción de información de páginas web

BLOQUES TEMÁTICOS Y CONTENIDOS

Contenidos – Bloques Temáticos

Bloque 0 Metodologías y conceptos básicos de trabajo en un equipo de desarrollo

Sesión 1: Principales conceptos de Scrum, Kanban, TDD, Git

Sesión 2: Ejercicios de Git

Bloque 1 Extracción y transformación de datos

Sesión 3: Conceptos para datawarehouse, datalake, ETL, ELTs ...

Sesión 4: Construir una ETL

Sesión 5: Conceptos básicos de expresiones regulares

Sesión 6: Orquestación de pipelines de datos

Sesión 7: Práctica

Bloque 2 Extracción y transformación de datos

Sesión 8: Conceptos de gestión de datos

Sesión 9: Limpieza y calidad de los datos

Sesión 10: Gobierno de los datos
 Sesión 11: Herramientas de calidad de datos
 Sesión 12: Conceptos de series temporales
 Bloque 3: Organización de los datos
 Sesión 13: Diferentes tipos de archivos
 Sesión 14: Concepto de CRUD
 Sesión 15: Práctica de CRUD
 Bloque 4: Archivos de almacenamiento
 Sesión 16: Equilibrio entre la cantidad de datos y el almacenamiento
 Sesión 17: Compresión y descompresión de datos
 Sesión 18: Exportación e importación de datos
 Sesión 19: Trabajo con archivos estructurados
 Bloque 5: La web como fuente de información
 Sesión 20: API
 Sesión 21: WebSocket
 Sesión 22: Conceptos de HTML
 Sesión 23: Web scrapping BS4
 Sesión 24: Web Scrapping Selenium
 Sesión 25: Social Scrapping

METODOLOGÍA DOCENTE

Aspectos metodológicos generales de la asignatura

Metodología Presencial: Actividades

Las actividades formativas serán:

- **Clases magistrales expositivas y participativas:**
 - El profesor combinará exposición de los contenidos teóricos y con ejemplos prácticos, tanto matemáticos como de programación.
 - El alumno dispondrá de algunos ejemplos prácticos de código, generado dentro y fuera del aula por el profesor.
- **Ejercicios prácticos y resolución de problemas:**
 - El alumno resolverá problemas planteados por el profesor de forma presencial durante la segunda hora de clase, fomentando las dinámicas de trabajo cooperativo.
- **Sesiones prácticas con uso de software:**
 - Se dedicarán las sesiones prácticas a resolver dudas de la práctica semanal y a que los alumnos finalicen su práctica.
 - Las prácticas se graduarán por dificultad que los alumnos irán implementando conforme vayan finalizando cada hito.
- **Actividades de evaluación continua del rendimiento:** se realizarán pruebas, desarrollarán prácticas complementarias a las semanales y retos gamificados.

CG04, CG05, CG07, CE14

Metodología No presencial: Actividades



Las actividades formativas serán:

- **Ejercicios prácticos y resolución de problemas:**
 - El alumno dispondrá de problemas concretos enfocados a asimilar los conceptos explicados teóricos en la sesión anterior de teoría para desarrollar de forma no presencial.
 - La solución de estos problemas será subida a la plataforma la semana siguiente, antes de empezar el nuevo tema.
- **Sesiones prácticas con uso de software:**
 - Una vez liberada la práctica semanal después de la sesión de teoría correspondiente, el alumno trabajará sobre ella de forma no presencial. El alumno deberá llegar a la sesión presencial de prácticas con los objetivos propuestos en el enunciado al 80%
 - En el aula se extenderá el enunciado de manera incremental y se cubrirán los hitos planteados de forma progresiva.
- **Estudio personal:** el objetivo principal del trabajo no presencial es llegar a entender y comprender los conceptos teóricos de la asignatura, así como ser capaz de poner en práctica estos conocimientos para resolver los diferentes tipos de problemas. Después de cada explicación teórica el profesor subirá a la web todos los códigos desarrollados y el alumno deberá revisarlos y plantearse cuestiones "Whatif" para asimilar mejor los conceptos teóricos.

CG04, CG05, CG07, CE14

RESUMEN HORAS DE TRABAJO DEL ALUMNO

HORAS PRESENCIALES

HORAS NO PRESENCIALES

CRÉDITOS ECTS: 4,5 (0 horas)

EVALUACIÓN Y CRITERIOS DE CALIFICACIÓN

Actividades de evaluación	Criterios de evaluación	Peso
Exámenes: <ul style="list-style-type: none"> • Prueba Intersemestral. • Examen Final. 	<ul style="list-style-type: none"> • Prueba Intersemestral (15%): comprensión de los conceptos fundamentales de la informática y la programación. • Examen Final (35%): se evaluará el pensamiento computacional y abstracto para la resolución de problemas mediante la programación. 	50 %
Casos prácticos <ul style="list-style-type: none"> • Practicas de cada bloque 	<ul style="list-style-type: none"> • Bloque 1, Extracción y transformación de datos (8%): Conceptos datawarehouse, datalake, ETL, ELT.... • Bloque 2. Extracción y transformación de datos (8%): Conceptos de gestión de datos, calidad, limpieza, Gobierno de los datos... • Bloque 3. Organización de los datos (8%). Diferentes tipologías de archivos no enriquecidos, concepto CRUD ... • Bloque 4. Archivos de almacenamiento(8%). Equilibrio cantidad de datos vs almacenamiento, compresión y 	40 %



	descompresión, archivos estructurados <ul style="list-style-type: none">• Bloque 5. La web como fuente de información(8%). API, WebSocket, Web scrapping.....	
Sesiones practicas <ul style="list-style-type: none">• La actitud, participación y realización de los retos planteados en sesiones colaborativas e individuales• Trabajo en equipo y desarrollo de este	<ul style="list-style-type: none">• Metodología colaborativa y buenas practicas (5%): Trabajos en equipo, siguiendo el uso de buenas practicas colaborativas• Trabajo en clase (5%): Participación y realización de los retos planteados en las sesiones	10 %

Calificaciones

La calificación final en **convocatoria ordinaria** y **extraordinaria** de la asignatura dependerá de la evaluación de las siguientes actividades:

- **Nota Final** = 15% Prueba_Intersemestral + 35% Examen_Final + 10% Práctica trabajo colaborativo y práctico de clase + 40% Casos prácticos
- La inasistencia al 15% o más de las horas presenciales de esta asignatura puede tener como consecuencia la imposibilidad de presentarse a las convocatorias ordinaria y extraordinaria.
- Los casos prácticos serán **INDIVIDUALES**.
- El examen **extraordinario** agrupa las notas de Inter + Final.

BIBLIOGRAFÍA Y RECURSOS

Bibliografía Básica

Presentaciones y códigos proporcionados por los profesores de la asignatura.

En cumplimiento de la normativa vigente en materia de **protección de datos de carácter personal**, le informamos y recordamos que puede consultar los aspectos relativos a privacidad y protección de datos que ha aceptado en su matrícula entrando en esta web y pulsando "descargar"

<https://servicios.upcomillas.es/sedelectronica/inicio.aspx?csv=02E4557CAA66F4A81663AD10CED66792>