

FICHA TÉCNICA DE LA ASIGNATURA

Datos de la asignatura	
Nombre completo	Adquisición y Transformación
Código	DTC-MBD-517
Título	N/A
Impartido en	Máster Universitario en Big Data [Primer Curso]
Créditos	3,0 ECTS
Carácter	Obligatoria
Departamento / Área	Departamento de Telemática y Computación

Datos del profesorado	
Profesor	
Nombre	Sofía Sánchez Urbano
Correo electrónico	ssurbano@icai.comillas.edu
Profesor	
Nombre	Alberto José López Espinosa
Departamento / Área	Departamento de Telemática y Computación
Correo electrónico	ajlespinosa@icai.comillas.edu

DATOS ESPECÍFICOS DE LA ASIGNATURA

Contextualización de la asignatura	
Aportación al perfil profesional de la titulación	
Capacidad para recoger, limpiar y estructurar grandes cantidades de datos a través de información publicada en internet.	
Recopilación de información para la creación de bases de datos.	
Prerrequisitos	
Nociones básicas de Python.	

Competencias - Objetivos	
Competencias	
Conocimientos o contenidos	
CO1	Entender los fundamentos de la analítica de datos y su aplicación en diversas áreas de la inteligencia artificial, destacando la integración en soluciones complejas y multidisciplinares para el análisis avanzado de datos masivos atendiendo a la diversidad de problemas específicos de cada área.



CO2	Comprender las técnicas de procesados de datos, las arquitecturas y herramientas más habituales y apropiadas para condiciones y requisitos de casos específicos.
Competencias	
CP1	Integrar las arquitecturas, técnicas de inteligencia artificial, análisis avanzado de datos y de visualización y de cumplimiento legal para ofrecer la solución global óptima.
CP4	Implementar las técnicas de procesamiento de datos y usar las herramientas más habituales y apropiadas a las condiciones y requisitos de casos específicos.
CP7	Aplicar conocimientos avanzados en Big Data y analítica de datos para desarrollar soluciones innovadoras en proyectos y en investigación, aportando y evaluando soluciones óptimas para el procesamiento y análisis de datos a gran escala.
Habilidades o destrezas	
HA1	Comunicar de manera oral y escrita con rigor técnico, claridad expositiva y coherencia argumentativa a todo tipo de interlocutores, técnicos y no técnicos.
HA2	Trabajar en equipos de carácter pluridisciplinar y/o internacional y organizar y liderar adecuadamente las dinámicas de grupo.
HA3	Desarrollar las habilidades interpersonales que requieren los entornos profesionales actuales (empatía, tolerancia, respeto, capacidad para aunar intereses contrapuestos).
HA4	Gestionar, organizar y planificar adecuadamente el trabajo y el tiempo, cumpliendo objetivos y estándares de calidad.
HA5	Mantener una formación y aprendizaje continuo y adaptación a los cambios tecnológicos y científicos.

BLOQUES TEMÁTICOS Y CONTENIDOS

Contenidos – Bloques Temáticos

Bloque 01. Extracción y transformación de datos. Procesos ETL. Gramática y expresiones regulares. CSV + JSON.

Bloque 02. API REST. Peticiones HTTP. Parseo de respuestas con Postman y utilizando Python. Interceptar peticiones XHR.

Bloque 03. La web como fuente de información: HTML y CSS. Proyecto de webscraping. Aspectos legales webscraping. Parseo de HTML con BeautifulSoup4 y Selenium. Interactuando con JS.

Bloque 04. Limpieza y calidad de los datos extraídos. Integridad y normalización. Problemas de codificación. Extracción de datos de PDF. OCR.

Teoría:

Clase 01. APIS con Postman.

Clase 02. Python. Requests.

Clase 03. Python. Estructura de datos.

Clase 04. Peticiones con cabeceras.



Clase 05. Formatos de respuesta.

Clase 06. Web scrapping con Bs4.

Clase 07. Estructura de proyecto. Spider + Fetcher.

Clase 08. Estructura de proyecto. Crawler + Wrangler.

Clase 09. Logueo de errores.

Clase 10. Web scraping con Selenium básico.

Clase 11. Web scraping con selenium avanzado.

Clase 12. Limpieza de datos.

Clase 13. Repaso.

Clase 14. Carrera profesional.

METODOLOGÍA DOCENTE

Aspectos metodológicos generales de la asignatura

Metodología Presencial: Actividades

Teoría:

Presentación de los conceptos básicos por parte del profesor y recomendaciones prácticas para llevarlos a cabo.

Se valorará la participación de los alumnos para generar entornos de debate en clase.

Práctica.

Introducción a la práctica asociada a cada una de las clases teóricas y seguimiento de la misma en clase.

Se valorará la cooperación entre los alumnos para la resolución de problemas.

CO1, CO2, CP1, CP4,
CP7, HA1, HA2, HA3,
HA4, HA5

Metodología No presencial: Actividades

Práctica.

Resolución de las prácticas presentadas en clase.

CO1, CO2, CP1, CP4,
CP7, HA1, HA2, HA3,
HA4, HA5

RESUMEN HORAS DE TRABAJO DEL ALUMNO

HORAS PRESENCIALES

Clases magistrales expositivas y participativas:
Exposición de contenidos fundamentales por parte del profesor impulsando la reflexión y participación del alumno.

13.00

Ejercicios prácticos y resolución de problemas: Sesiones prácticas con uso de software:
Actividad formativa con ordenador que, bajo la guía del profesor-tutor, fomenta el aprendizaje autónomo y/o cooperativo del alumno mediante la ejecución de programas para la consecución de los objetivos marcados

13.00



HORAS NO PRESENCIALES		
Estudio personal: Reflexión y análisis individual de los contenidos teóricos y prácticos de las materias y/o asignaturas del Master	Trabajos: Los alumnos tendrán que hacer trabajos breves (individuales y/o en grupo), por indicación del profesor	Proyectos: Los alumnos tendrán que hacer trabajos de tamaño medio o grande (individuales y/o en grupo), por indicación del profesor
12.00	10.00	42.00
CRÉDITOS ECTS: 3,0 (90,00 horas)		

EVALUACIÓN Y CRITERIOS DE CALIFICACIÓN

Actividades de evaluación	Criterios de evaluación	Peso
Examen tipo test.	Selección de respuestas correctas entre opciones dadas. Se busca entender conceptos teóricos vistos en clase.	30 %
Prácticas individuales.	Creación de scripts y peticiones individuales. Se valorará la limpieza del código, su eficacia y el nivel de documentación.	20 %
Prácticas compuestas.	Creación de proyectos y flujos de ejecución. Se valorará la limpieza del código, su eficacia y el nivel de documentación.	50 %

Calificaciones

Para pasar el curso la puntuación del examen debe ser mayor que 5 sobre 10.

Para pasar el curso la puntuación de la media de las prácticas debe ser mayor que 5 sobre 10.

La asistencia al 85% de las clases es obligatoria para poder hacer el examen según el Artículo 93 del Reglamento General de la Universidad de Comillas.

PLAN DE TRABAJO Y CRONOGRAMA

Actividades	Fecha de realización	Fecha de entrega
Clase 01. APIS con Postman.	semana 01	semana 03
Clase 02. Python. Requests.		



Clase 03. Python. Estructura de datos. Clase 04. Peticiones con cabeceras.	semana 02	semana 04
Clase 05. Formatos de respuesta. Clase 06. Web scrapping con Bs4.	semana 03	semana 05
Clase 07. Estructura de proyecto. Spider + Fetcher. Clase 08. Estructura de proyecto. Crawler + Wrangler.	semana 04	semana 06
Clase 09. Logueo de errores. Clase 10. Web scraping con Selenium básico.	semana 05	semana 07
Clase 11. Web scraping con selenium avanzado. Clase 12. Limpieza de datos.	semana 06	semana 08
Clase 13. Repaso. Clase 14. Carrera profesional.	semana 07	semana 09

BIBLIOGRAFÍA Y RECURSOS

Bibliografía Básica

Presentaciones teóricas.

Documentación oficial de librerías open source: requests, bs4, selenium & chromedriver.

Bibliografía Complementaria

Web scraping with Python, Ryan Mitchel, O'reilly. ISBN: 1491985577

Python Crash Course, Eric Matthes. ISBN: 1593279280