

Guía docente de Asignatura – Máster en minería de datos e inteligencia de negocios

Datos generales de la asignatura

Asignatura:	Gestión y Explotación de Almacenes de Datos - 606542
Curso académico:	2023-24
Carácter:	Obligatoria
Curso:	Primero
Semestre:	2

Créditos ECTS

Presenciales:	4,5
No presenciales:	1,5
Total	6,0

Actividades docentes

Clases teóricas:	50%
Clases prácticas:	50%
Total:	100%
Departamentos responsables:	Sistemas Informáticos y Computación
Profesores:	Gabriel Marín Díaz y José Javier Galán Hernández
Profesor Coordinador:	Gabriel Marín Díaz

Datos específicos de la asignatura

Breve descriptor:	Estudio de técnicas para la obtención de información (análisis y minería de datos) a partir de distintas fuentes de datos: bases de datos relacionales, orientadas a documento, páginas web, escucha en redes sociales ... y aplicar estas técnicas a diferentes campos científicos y empresariales.
Requisitos:	Es importante haber cursado alguna asignatura de bases de datos que incluya el estudio del modelo relacional y el uso práctico del lenguaje SQL.

Competencias

Generales:	La necesidad de recopilar información para su análisis está presente en nuestras vidas... El problema fundamental en los procesos de transformación digital empresariales se basa en la integridad del dato y en la capacidad de transformar el dato en información. Conocer las herramientas y lenguaje que se utiliza habitualmente para extraer, almacenar y procesar información es uno de los objetivos principales de esta asignatura. En este curso se estudian los almacenes de datos desde el punto de vista de su aplicación en técnicas básicas tanto de análisis como de minería de datos (v.g. asociación, predicción o clustering), junto con sus características, limitaciones y posibles ámbitos de implantación (por ejemplo, el ámbito científico, el empresarial o el financiero). Para ello, se utiliza el lenguaje de programación Python, bases de datos relacionales como MySQL y orientadas al documento como MongoDB.
Transversales:	
Específicas:	Identificar los elementos presentes en un problema real. Conocer en profundidad los sistemas de gestión de bases de datos relacionales y orientados a documentos. Entender las distintas fuentes de información de una compañía y cómo podemos integrar toda esta en almacenes de datos. Realizar un análisis de la WEB e integrar dicha información con la generada a través de los sistemas tradicionales para ayudar en la toma de decisiones. Conocer en profundidad el lenguaje Python como herramienta de trabajo en modelos analíticos y predictivos.

	Conceptualizar, diseñar e implementar soluciones que permitan adaptar a casos reales soluciones basadas en el aprendizaje automático.
Objetivos	- Conocer los fundamentos del lenguaje SQL. - Conocer los fundamentos del lenguaje Python. - Aplicar Python en el conocimiento de los fundamentos y técnicas de Web Scraping. - Aplicar Python en el conocimiento de los fundamentos y técnicas de escucha en Redes Sociales. - Conocer los fundamentos de MongoDB como herramienta de gestión de base de datos orientada a documentos. - Conocer los fundamentos del Big Data (Machine & Deep Learning). - Conocer los fundamentos del procesamiento distribuido.
Contenidos	- Repaso de bases de datos, lenguaje SQL. - Introducción al lenguaje Python. * Acceso a bases de datos. * Recopilación de datos en la Web (Web Scraping: Selenium, BeautifulSoup). * Acceso a datos en redes sociales. * Recolección mediante APIs. - Introducción a MongoDB con Python. - Esquema de programación Map Reduce. - Introducción a Big Data (Machine Learning & Deep Learning). - Introducción a los Sistemas Distribuidos. - Interpretabilidad de algoritmos de caja negra.
Evaluación	La evaluación en periodo ordinario consistirá en la resolución y exposición en clase de tres trabajos propuestos por el profesor durante el curso: - Ejercicio práctico 1. Bases de Datos con Python. El peso de esta actividad corresponde con el 30% del valor de la nota final. - Ejercicio práctico 2. Web Scraping con Python. El peso de esta actividad corresponde con el 30% del valor de la nota final. - Ejercicio práctico 3. Machine & Deep Learning con Python. El peso de esta actividad corresponde con el 30% del valor de la nota final. Además, la asistencia y participación del estudiante en el aula (exposición de trabajos) tendrá un peso de un 10% en la nota final. En evaluación extraordinaria, se requerirá la entrega de los tres trabajos de evaluación ordinaria, aplicándose los mismos pesos que en evaluación ordinaria. Si algún estudiante no ha podido entregar alguno de los trabajos en la convocatoria ordinaria, podrá hacerlo en extraordinaria. En evaluación extraordinaria los trabajos serán entregados a través del campus virtual y no expuestos en clase, por lo que el 10% de asistencia y participación se puntuará como 0 (perdiendo esa parte de la calificación).
Bibliografía	* Parker, James R. Python : An introduction to programming. Dulles (Virginia, EEUU) : Mercury Learning and Information, cop. 2017 * Russell, Matthew A. Mining the social web : [data mining Facebook, Twitter, LinkedIn, Google+, GitHub, and more] / Matthew A. Russell. Sebastopol (California) : O'Really Media, 2013 * Documentación Hadoop: http://hadoop.apache.org/docs/current/ * Documentación MrJob: https://pythonhosted.org/mrjob/ * Karau H., Konwinski A., Wendell P., and Zaharia M. Learning Spark. O'Reilly. 2015 * Documentación Spark: http://spark.apache.org/docs/latest/ * Python paso a paso. Angel Hinojosa Gutierrez * Interpretabilidad Machine Learning: https://christophm.github.io/interpretable-ml-book/ - Christoph Molnar