



FICHA IDENTIFICATIVA

DATOS DE LA ASIGNATURA

Código: 36516

Nombre: Datos no Estructurados

Ciclo: Grado

Créditos ECTS: 6

Curso académico: 2025-26

TITULACIONES

Titulación	Centro	Curso	Periodo
1332 - Grado en Inteligencia y Analítica de Negocios/BIA	Facultat d'Economia	4	Sin determinar, Primer cuatrimestre

MATERIAS

Titulación	Materia	Carácter
1332 - Grado en Inteligencia y Analítica de Negocios/BIA	Herramientas y Técnicas de Análisis de Datos	OBLIGATORIA

COORDINACIÓN

MONTALVA CANDEL IGNACIO

MONTALVA CANDEL IGNACIO

MONTALVA CANDEL IGNACIO

RESUMEN

Con el desarrollo de la Web 2.0 y las Redes Sociales llegaron los datos masivos. La mayoría de estos datos sin una estructura inicial prefijada. Nos referimos a documentos, fotos, videos, correos electrónicos, tweets, etc. La mayor parte de estos datos los generan los propios usuarios. Todos los días se envían correos electrónicos, se publican tweets o se suben fotos o videos a las diferentes redes sociales. Todos estos datos facilitan mucha información sobre las personas que la están generado. Sin embargo, al ser datos No estructurados se deberán transformar la información que generan para poder analizarla y convertirla en conocimiento.

Un ejemplo, es la información de una empresa en la red social Twitter. Se pueden recopilar todos los tweets que mencionen a una empresa, y mediante una transformación en una base de datos estructurada se podrá hacer análisis de sentimiento y saber qué opina las personas de esta empresa. También se pueden descargar los tweets de la competencia para saber la opinión sobre esas empresas y poder modificar la estrategia de la empresa.



Los datos No estructurados precisarán de bases de datos no relacionales, como pueden ser MongoDB (basada en documentos), Neo4j (basada en grafos), etc. Las nuevas bases de datos permitirán realizar un análisis diferente al que se realiza con las bases de datos relacionales.

El objetivo de la asignatura será que los alumnos conozcan y puedan trabajar con estos datos no estructurados y realizar análisis de estos para tener un mejor conocimiento de la empresa y de la competencia.

nto de la empresa y de la competencia.

CONOCIMIENTOS PREVIOS

RELACIÓN CON OTRAS ASIGNATURAS DE LA MISMA TITULACIÓN

No se han especificado restricciones de matrícula con otras asignaturas del plan de estudios.

OTROS TIPOS DE REQUISITOS

Es recomendable tener conocimientos del software R y RStudio.
Así mismo, conocimientos de Estadística Básica e Inferencia, así como Aprendizaje Automático (machine learning).

COMPETENCIAS / RESULTADOS DE APRENDIZAJE

-

Aplicar, utilizando software, técnicas de aprendizaje automático (machine learning) no supervisado y semisupervisado.

Aplicar, utilizando software, técnicas de aprendizaje automático (maching learning) supervisado.

Aplicar métodos y técnicas de análisis, síntesis y representación gráfica mediante programas informáticos.

Aplicar minería de datos mediante programas informáticos.

Capacidad de aprendizaje autónomo.

Capacidad para analizar y buscar información proveniente de fuentes diversas.

Capacidad para definir, resolver y exponer de forma sistémica problemas complejos.

Capacidad para trabajar en equipo, con el compromiso por la calidad, la ética, la igualdad entre personas y la responsabilidad social.

Capacidad para utilizar las TIC, tanto en el ámbito de estudio como en el desarrollo profesional.

Manejar aplicaciones de descarga, capturar y manipular valores de datos no estructurados.

Que los estudiantes puedan transmitir información, ideas, problemas y soluciones a un público tanto especializado como no especializado.



Reorganizar y reestructurar variables y bases de datos.

DESCRIPCIÓN DE CONTENIDOS

1. Introducción

- a. Datos, Información y Aprovechamiento
- b. Datos estructurados vs no estructurados
- c. La ilusión del análisis de datos no estructurados
- d. Estándares de marcado: XML, JSON,...

2. Captura de Datos en la Web I

- a. Captura de datos de sitios web: captura directa
- b. Una aplicación a la obtención de un índice: IPC
- c. Captura de datos mediante APIs
- d. Buenas prácticas

3. Web Scraping: Captura de Datos en la Web II

- a. Estructura de una página web: HTML
- b. HTML dinámico: JavaScript
- c. Hojas de Estilo: CSS, W3-CSS, Bootstrap(CSS)
- d. Captura completa de un sitio web
- e. Captura específica de datos de la web
- f. Scripts que imitan la intervención humana: Bots

4. Minería de Texto

- a. Tipos y usos de las técnicas de análisis de texto
- b. Enfoques: bag of words (BoW) y natural language processing (NLP)
- c. El flujo de trabajo básico
- d. Preparación y estructuras de datos
- e. Visualizaciones y análisis descriptivo

5. Análisis de Sentimientos

- a. Contenido subjetivo en textos, polaridad y emociones
- b. Enfoque basado en diccionarios vs aprendizaje automático
- c. Diccionarios de sentimientos en Español
- d. Aplicaciones del análisis de sentimientos



6. Aprendizaje Automático y Analítica de Texto

- Técnicas de aprendizaje supervisado y no supervisado
- Modelización probabilística de temas (topic modeling)
- Clasificación de textos
- Modelización predictiva
- Técnicas avanzadas para el análisis de texto

7. Sistemas de Recomendación

- Tipos de sistemas de recomendación
- Preferencias implícitas y explícitas
- Construcción de un sistema de recomendación
- Evaluación de un sistema de recomendación
- Aplicaciones

8. Redes y Redes Sociales

- Definición y tipologías
- Modelos estadísticos de redes sociales
- Modelos de grafos aleatorios
- Redes sociales y segmentación
- Redes neuronales aplicadas

VOLUMEN DE TRABAJO (HORAS)

ACTIVIDADES PRESENCIALES

Actividad	Horas
Teoría	15,00
Aula informática	45,00
Total horas	60,00

ACTIVIDADES NO PRESENCIALES

Actividad	Horas
Asistencia a otras actividades	0,00
Elaboración de trabajos individuales o en grupo	20,00
Estudio y trabajo autónomo	0,00
Preparación de clases	50,00
Preparación de actividades de evaluación	10,00
Resolución de casos prácticos	10,00
Total horas	90,00

METODOLOGÍA DOCENTE



- Clase teórica presencial para presentar los contenidos teóricos esenciales de la materia.
- Clases prácticas presenciales, relativas a la resolución de problemas, estudios de casos, con aplicación de técnicas, presentaciones orales, debates, individualmente y/o en equipo.
- Trabajo autónomo supervisado y basado en la realización de ejercicios, casos prácticos y cuestiones a debatir o experimentos on line, con apoyo tutorial.

EVALUACIÓN

Durante el curso se realizarán varias prácticas y pruebas de evaluación, mediante las cuales los alumnos podrán poner de manifiesto los conceptos adquiridos, tanto teóricos como prácticos. El conjunto de las prácticas y pruebas realizadas supondrá el 60% de la nota final.

Como proyecto final de curso, los alumnos deberán realizar un trabajo en el que tendrán que plasmar los conocimientos adquiridos en la asignatura. Este informe se deberá entregar, como máximo, en la fecha fijada para el examen de primera convocatoria (40% de la nota final).

Si bien las prácticas no son recuperables, los alumnos que no superen la asignatura en primera convocatoria podrán recuperar el informe final, que se deberá entregar, como máximo, en la fecha fijada para el examen de segunda convocatoria.

Para poder aplicar los porcentajes anteriormente mencionados será necesario obtener, al menos, una calificación de 5 en el proyecto final de curso.

BIBLIOGRAFÍA

- Anandarajan, M.; Hill, C. y Nolan, T. (2019) Practical Text Analytics. Maximizing the Value of Text Data. Springer Nature, Switzerland.
- Bali, R. Sarkar, D. y Sharma, T. (2017) Learning Social Media Analytics with R. Pack Publishing, Birmingham, UK.
- Gorakala, S.K. y Usuelli, M. (2015) Building a Recommendation System with R. Pack Publishing, Birmingham, UK.
- Kwartler, T. (2017). Text mining in practice with R. John Wiley & Sons.
- Munzert, S.; Rubba, C.; Meißner, P. y Nyhuis, D. (2015) Automated data collection with R: a practical guide to web scraping and text mining. John Wiley & Sons, UK.
- Rusell, Matthew A. (2013) Mining the Social Web. O'Reilly Media, CA.