

**FICHA IDENTIFICATIVA****Datos de la Asignatura**

Código	36516
Nombre	Datos no Estructurados
Ciclo	Grado
Créditos ECTS	6.0
Curso académico	2023 - 2024

Titulación(es)

Titulación	Centro	Curso	Periodo
1332 - Grado en Inteligencia y Analítica de Negocios/BIA	Facultad de Economía	4	Primer cuatrimestre

Materias

Titulación	Materia	Caracter
1332 - Grado en Inteligencia y Analítica de Negocios/BIA	24 - Herramientas y Técnicas de Análisis de Datos	Obligatoria

Coordinación

Nombre	Departamento
MONTORO PONS, JUAN DE DIOS	110 - Economía Aplicada
MORILLAS JURADO, FRANCISCO GABRIEL	110 - Economía Aplicada

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.

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.
Estadística Básica e Inferencia, así como Aprendizaje Automático (machine learning).

COMPETENCIAS

1332 - Grado en Inteligencia y Analítica de Negocios/BIA

- Que los estudiantes puedan transmitir información, ideas, problemas y soluciones a un público tanto especializado como no especializado.
- Capacidad para analizar y buscar información proveniente de fuentes diversas.
- Capacidad de aprendizaje autónomo.
- Capacidad para utilizar las TIC, tanto en el ámbito de estudio como en el desarrollo profesional.
- 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.
- Aplicar métodos y técnicas de análisis, síntesis y representación gráfica mediante programas informáticos.
- Reorganizar y reestructurar variables y bases de datos.
- Aplicar minería de datos mediante programas informáticos.
- Aplicar, utilizando software, técnicas de aprendizaje automático (machine learning) supervisado.
- Aplicar, utilizando software, técnicas de aprendizaje automático (machine learning) no supervisado y semisupervisado.
- Manejar aplicaciones de descarga, capturar y manipular valores de datos no estructurados.



RESULTADOS DE APRENDIZAJE

- Sintetizar información.
- Saber diseñar e implementar una investigación y su aplicación al entorno económico.
- Aplicar las técnicas de aprendizaje automático a problemas de negocio.
- Conocer los diferentes tipos de bases de datos existentes y los tipos de datos que utilizan.
- Aprender los conceptos de la Teoría de Grafos.
- Conocer cómo funciona un sistema de recomendación.
- Aprender a procesar datos no numéricos mediante programas informáticos.
- Aprender a realizar diccionarios con los que realizar procesamiento automático de lenguaje natural para su explotación de negocio.
- Saber implementar una investigación y su aplicación al entorno económico con datos textuales.
- Manejar herramientas cuantitativas avanzadas y su aplicación al entorno de negocios.
- Aprender a manejar APIs de descarga de datos en redes sociales.
- Aprender a reconocer patrones en datos en formato texto
- Aprender procesos para generar recomendaciones de compra.

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

- a. Técnicas de aprendizaje supervisado y no supervisado
- b. Modelización probabilística de temas (topic modeling)
- c. Clasificación de textos
- d. Modelización predictiva
- e. Técnicas avanzadas para el análisis de texto

7. Sistemas de Recomendación

- a. Tipos de sistemas de recomendación
- b. Preferencias implícitas y explícitas
- c. Construcción de un sistema de recomendación
- d. Evaluación de un sistema de recomendación
- e. Aplicaciones

8. Redes y Redes Sociales

- a. Definición y tipologías
- b. Modelos estadísticos de redes sociales
- c. Modelos de grafos aleatorios
- d. Redes sociales y segmentación
- e. Redes neuronales aplicadas



VOLUMEN DE TRABAJO

ACTIVIDAD	Horas	% Presencial
Prácticas en aula informática	45,00	100
Clases de teoría	15,00	100
Elaboración de trabajos en grupo	10,00	0
Elaboración de trabajos individuales	10,00	0
Preparación de actividades de evaluación	10,00	0
Preparación de clases de teoría	20,00	0
Preparación de clases prácticas y de problemas	30,00	0
Resolución de casos prácticos	10,00	0
TOTAL	150,00	

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

Evaluación consistente en una o varias pruebas que considerarán tanto cuestiones teóricas como prácticas. (50%)

Evaluación de las actividades prácticas desarrolladas por el/la alumno/a durante el curso, a partir de la elaboración de trabajos/memorias y/o exposiciones orales, con defensa de las posiciones desarrolladas por el/la alumno/a. (30%)

Evaluación continua de cada alumno, basada en la participación y grado de implicación del/de la alumno/a en el proceso de enseñanza-aprendizaje, teniendo en cuenta la asistencia regular a las actividades presenciales previstas y la resolución de cuestiones y problemas propuestos periódicamente. (20%)

REFERENCIAS



Básicas

- 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.; Meiner, 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. OReilly Media, CA.