



FICHA TÉCNICA DE LA ASIGNATURA

Datos de la asignatura

Nombre completo	Análisis de Datos no Estructurados/ Unstructured Data Analysis
Código	E000013734
Nivel	Postgrado Oficial Master
Cuatrimestre	Anual
Créditos	3,0 ECTS
Carácter	Obligatoria
Departamento / Área	Departamento de Métodos Cuantitativos

Datos del profesorado

Profesor

Nombre	Jenny Alexandra Cifuentes Quintero
Departamento / Área	Departamento de Métodos Cuantitativos
Despacho	OD-431
Correo electrónico	jacifuentes@icade.comillas.edu

DATOS ESPECÍFICOS DE LA ASIGNATURA

Contextualización de la asignatura

Aportación al perfil profesional de la titulación

En la actualidad, la información valiosa en el proceso de toma de decisiones empresariales no se limita a los datos estructurados tradicionales. Una gran cantidad de datos relevantes proviene de fuentes no estructuradas como redes sociales, blogs, correos electrónicos y otros textos libres. La asignatura de Análisis de Datos No Estructurados tiene como objetivo principal capacitar a los estudiantes para extraer información significativa de estos datos no estructurados.

Las competencias a desarrollar en esta asignatura permiten a los profesionales abordar una variedad más amplia de problemas empresariales y aprovechar una mayor cantidad de información para tomar decisiones informadas y contribuir a la innovación en diversos sectores. De esta manera, la asignatura proporciona una introducción a las técnicas de análisis de texto y otros datos no estructurados, enfocándose en métodos para extraer, procesar y analizar información relevante. Además, la asignatura se centrará en la aplicación práctica de estas técnicas a datos y problemas reales, lo que permitirá a los alumnos adquirir las habilidades necesarias para enfrentar desafíos empresariales concretos en su futura carrera profesional.

Prerrequisitos

Machine Learning. Fundamentos y Aprendizaje Supervisado /Machine Learning. Principles and Supervised Learning

Machine Learning. Aprendizaje No Supervisado /Machine Learning. Unsupervised Learning

Competencias - Objetivos



Competencias

CO4. Conocer de manera intuitiva los conceptos y el lenguaje de las técnicas y métodos avanzados en Business Analytics, desde el análisis de información no estructurada, pasando por las redes neuronales y métodos de Deep Learning y los nuevos avances en el campo de la Inteligencia Artificial, entendiendo su alcance en la empresa y la sociedad.

Resultados de Aprendizaje

RA1. Comprender las características y diferencias entre datos estructurados y no estructurados, así como su relevancia y aplicaciones en diversos contextos empresariales.

RA2. Saber elegir y aplicar técnicas de limpieza, normalización y transformación de datos no estructurados para preparar los datos de manera efectiva para su análisis.

RA3. Conocer los conceptos básicos del procesamiento de lenguaje natural y las principales técnicas de representación de texto, y saber aplicar estas técnicas para extraer y analizar información relevante a partir de datos no estructurados.

RA4. Saber identificar y aplicar técnicas y herramientas avanzadas para la categorización, agrupamiento y análisis de tendencias en documentos.

RA5. Saber interpretar, evaluar y comunicar los resultados derivados del análisis de datos no estructurados

RA6. Saber aplicar técnicas de análisis de datos no estructurados a conjuntos de datos reales usando software apropiado

BLOQUES TEMÁTICOS Y CONTENIDOS

Contenidos – Bloques Temáticos

- **Introducción al análisis de datos no estructurados.** Definición, tipos, fuentes, preprocesamiento, análisis descriptivo y representación con n-gramas y TF-IDF.
- **Modelos de representación numérica de texto:** Bag of Words, TF-IDF, Embeddings, Word2Vec, GloVe y Doc2Vec.
- **Representaciones contextuales de texto con transformers:** mecanismos de atención, BERT y aplicaciones en clasificación y clustering.
- **Técnicas de Modelado de Tópicos.**
- **Análisis de sentimiento:** enfoques léxicos, aprendizaje automático y modelos basados en transformers.

HORAS PRESENCIALES

Lecciones de Carácter expositivo	Ejercicios y resolución de casos y de problemas
15	5
Otras actividades, seminarios, talleres, simulaciones, dinámicas de grupo, etc	
10	

METODOLOGÍA DOCENTE

Aspectos metodológicos generales de la asignatura

La asignatura está diseñada con una orientación eminentemente práctica y aplicada. Los alumnos deberán prepararse antes de cada clase



revisando los contenidos que se tratarán, lo que permitirá un aprovechamiento óptimo de las sesiones presenciales. Durante estas sesiones, se reforzarán los contenidos teóricos y se explicarán los conceptos fundamentales del análisis de datos no estructurados. Las clases incluirán lecciones expositivas combinadas con ejercicios prácticos y la resolución de casos aplicados. Los estudiantes trabajarán en ejemplos sencillos que les permitirán aplicar los conceptos aprendidos y profundizar en su comprensión.

Las actividades prácticas serán un componente clave de la asignatura, enfocándose en el uso de técnicas de preprocesamiento, representación y análisis de datos no estructurados. Se realizarán prácticas tanto en clase como para llevar a cabo en casa, lo que permitirá a los estudiantes desarrollar y perfeccionar sus habilidades de manera continua. Además, los estudiantes trabajarán en un proyecto práctico que implique el análisis de datos no estructurados aplicados a problemas empresariales reales, facilitando la aplicación de técnicas avanzadas en contextos empresariales y económicos. Las simulaciones y dinámicas de grupo se utilizarán para fomentar la colaboración y el aprendizaje activo.

El uso de herramientas de IA generativa como ChatGPT será bienvenido para tareas de depuración y corrección de código, asistencia con la programación y resolución de dudas. Se fomentará su empleo como un "copiloto", pero se exigirá que el alumno supervise, entienda y pueda explicar todo lo realizado. El contenido generado por estas herramientas debe ser revisado críticamente por el estudiante, ya que no siempre puede ser veraz. Sin embargo, su uso queda prohibido en las actividades de evaluación presenciales individuales.

Metodología Presencial: Actividades

Exposición de los principales conceptos teóricos

Realización de ejemplos de aplicación sencillos

Puesta en común y corrección de casos y problemas realizados por los alumnos

Actividades de evaluación

Actividades como seminarios, talleres, simulaciones o dinámicas de grupo.

Metodología No presencial: Actividades

Estudio personal

Realización de casos y trabajos de aplicación práctica, empleando datos y programación

RESUMEN HORAS DE TRABAJO DEL ALUMNO

HORAS NO PRESENCIALES			
Ejercicios y resolución de casos y de problemas	Estudios individual y/o en grupo, y lectura organizada	Otras Actividades	Trabajos monográficos y de investigación, individuales o colectivos
5	15	15	10



CRÉDITOS ECTS: 3,0 (75 ,00 horas)

EVALUACIÓN Y CRITERIOS DE CALIFICACIÓN

Actividades de evaluación	Criterios de evaluación	Peso
Evaluación continua Individual 25%	Calificación numérica de 0 a 10	25%
Examen Final 30%	Calificación numérica de 0 a 10.	30%
Trabajo de aplicación práctica individual 30%	Uso de datos reales y programación Calificación mediante rúbrica Propuesta 20%, Trabajo final 30%, presentación 50%	30 %
Casos prácticos periódicos en grupo 15%	A realizar en clase-casa Calificación numérica 0-10	15 %

Calificaciones

Para aprobar la asignatura, el trabajo de aplicación práctica debe ser realizado y aprobado con una calificación mínima de 5. La defensa oral del trabajo es obligatoria, ya que en una asignatura de carácter aplicado es necesario demostrar que se han adquirido las habilidades necesarias para aplicar los conceptos teóricos a datos reales, utilizando las técnicas de análisis adecuadas y las habilidades de programación pertinentes.

El trabajo de aplicación práctica tendrá un peso del 30% en la nota final de la asignatura. Esta calificación se dividirá de la siguiente manera: 6% para la propuesta inicial del trabajo, 9% para el entregable escrito, y 15% para la defensa oral. La defensa oral es especialmente importante, ya que permite evaluar la capacidad del estudiante para explicar y justificar el análisis realizado, así como su competencia en la aplicación práctica de los conceptos aprendidos durante el curso.

PLAN DE TRABAJO Y CRONOGRAMA

Actividades	Fecha de realización	Fecha de entrega
-------------	----------------------	------------------



Propuesta de trabajo		Hacia la mitad de la asignatura
Trabajo práctico final		Última semana de clases

BIBLIOGRAFÍA Y RECURSOS

Bibliografía Básica

Anandarajan, Murugan, Chelsey Hill, and Thomas Nolan. "Practical text analytics." Maximizing the Value of Text Data.(Advances in Analytics and Data Science. Vol. 2.) Springer (2019): 45-59.

Singh, J. (2023). Natural Language Processing in the Real World: Text Processing, Analytics, and Classification. United Kingdom: CRC Press.

Goldberg, Yoav. Neural network methods for natural language processing. Springer Nature, 2022.

Bibliografía Complementaria

Srinivasa-Desikan, B. (2018). Natural Language Processing and Computational Linguistics: A Practical Guide to Text Analysis with Python, Gensim, SpaCy, and Keras. India: Packt Publishing.

Silge, Julia, and David Robinson. Text mining with R: A tidy approach. " O'Reilly Media, Inc.", 2017.

Vajjala, Sowmya, et al. Practical natural language processing: a comprehensive guide to building real-world NLP systems. O'Reilly Media, 2020.