



COMILLAS

UNIVERSIDAD PONTIFICIA

ICAI

ICADE

CIHS

GUÍA DOCENTE

2026 - 2027

FICHA TÉCNICA DE LA ASIGNATURA

Datos de la asignatura	
Nombre completo	Minería avanzada de textos y Procesamiento de Lenguaje Natural/Advanced Text Mining and Natural Language Processing
Código	E000015167
Impartido en	Grado en Análisis de Negocios/Bachelor in Business Analytics [Cuarto Curso] Grado en Admin. y Dirección de Emp. y Grado en Análisis de Negocios/Bachelor in Business Analytics [Cuarto Curso]
Nivel	Reglada Grado Europeo
Cuatrimestre	Anual
Créditos	3,0 ECTS
Carácter	Obligatoria (Grado)
Departamento / Área	Departamento de Métodos Cuantitativos

Datos del profesorado	
Profesor	
Nombre	Jenny Alexandra Cifuentes Quintero
Departamento / Área	Departamento de Métodos Cuantitativos
Correo electrónico	jacifuentes@icade.comillas.edu
Profesor	
Nombre	Amir Dehdarirad
Departamento / Área	Departamento de Métodos Cuantitativos
Correo electrónico	adehdarirad@comillas.edu
Profesor	
Nombre	Leonardo Dulcetti
Departamento / Área	Departamento de Métodos Cuantitativos
Correo electrónico	ldulcetti@comillas.edu
Profesor	
Nombre	Víctor Pérez Segura
Departamento / Área	Departamento de Métodos Cuantitativos
Correo electrónico	vperezs@comillas.edu



DATOS ESPECÍFICOS DE LA ASIGNATURA

Contextualización de la asignatura

Aportación al perfil profesional de la titulación

Una parte creciente de la información utilizada en los procesos de toma de decisiones empresariales se encuentra en formatos no estructurados. Opiniones de clientes, publicaciones en redes sociales, noticias, informes, reseñas, correos electrónicos y otros documentos textuales contienen información relevante sobre las preferencias de los consumidores, la reputación corporativa, las tendencias del mercado y los riesgos que afectan a las organizaciones.

La asignatura de Análisis de Datos No Estructurados introduce al estudiante en los fundamentos del procesamiento y análisis de información textual mediante técnicas de procesamiento del lenguaje natural y aprendizaje automático. A lo largo del curso, los estudiantes aprenderán a preparar, representar y analizar colecciones de documentos, utilizando métodos tradicionales y modelos contemporáneos basados en *transformers*.

La asignatura presenta técnicas de preprocesamiento textual, representación numérica de documentos, modelado de tópicos, análisis de sentimientos y emociones, así como una introducción al funcionamiento y uso responsable de los modelos de lenguaje de gran tamaño. Estos contenidos permitirán al estudiante convertir información textual en resultados cuantitativos que puedan incorporarse a procesos de análisis y toma de decisiones.

La orientación de la asignatura será eminentemente práctica. Los estudiantes aplicarán las técnicas estudiadas mediante Python y trabajarán con conjuntos de datos reales vinculados con problemas empresariales y sociales. De este modo, la asignatura contribuirá al desarrollo de competencias analíticas, tecnológicas y comunicativas relevantes para el ejercicio profesional en áreas como marketing, consultoría, inteligencia de negocio, investigación de mercados, análisis de riesgos y estrategia empresarial.

Prerrequisitos

Para cursar adecuadamente la asignatura, se requiere haber adquirido previamente los conocimientos y competencias desarrollados en las asignaturas **Machine Learning I** y **Machine Learning III**.

En particular, se espera que los estudiantes conozcan los fundamentos del aprendizaje supervisado y no supervisado, las principales etapas de preparación y análisis de datos, los conceptos básicos de clasificación y clustering, así como los criterios generales de evaluación e interpretación de modelos. También deberán contar con competencias previas de programación en Python y manejo de datos, necesarias para implementar las técnicas de minería de textos y procesamiento del lenguaje natural abordadas en la asignatura.

Competencias - Objetivos

Competencias

GENERAL

CG02 Capacidad de análisis de datos complejos y no estructurados procedentes de diversas fuentes

CG03 Resolución de problemas y toma de decisiones en un entorno de datos masivos tanto cuantitativos como cualitativos.

RA1 Saber seleccionar para cada problema la técnica o técnicas de análisis de datos más adecuada para poder convertir los datos en bruto en información y ésta en conocimiento que ayude a la toma de decisiones y a mejorar la gestión.



CG04	Capacidad para elaborar proyectos e informes de manera oral y escrita, difundiendo estas ideas a través de canales digitales
RA1	Ser capaz de resumir, sintetizar y comunicar de una forma atractiva y eficaz los resultados de la aplicación de las técnicas de análisis de datos, incluso de las más sofisticadas, de manera que resulten comprensibles a destinatarios no técnicos y ayuden de forma eficiente a la toma de decisiones empresariales.
CG09	Compromiso ético en la sociedad de la información
RA1	Persigue la excelencia en las actuaciones profesionales
RA2	Se preocupa por las consecuencias que su actividad y su conducta pueden tener para los demás
RA3	Incorpora en su discurso y en sus propuestas de actuaciones, las consecuencias que las mismas pueden tener para los distintos stakeholders de una organización global

BLOQUES TEMÁTICOS Y CONTENIDOS

Contenidos – Bloques Temáticos

Introducción a la minería de textos y preprocesamiento del lenguaje natural

Representación numérica del texto

Representaciones contextuales y modelos basados en transformers

Modelado de tópicos

Análisis de sentimientos y emociones

Introducción y uso aplicado de los modelos extensos de lenguaje

METODOLOGÍA DOCENTE

Aspectos metodológicos generales de la asignatura

Metodología Presencial: Actividades

Exposición de los principales conceptos teóricos

Realización y discusión de ejemplos utilizando programación.

Puesta en común y corrección de casos/problemas realizados por los alumnos (usando algún lenguaje de programación)

Actividades de evaluación

Metodología No presencial: Actividades

Estudio personal

Realización de casos y trabajos de aplicación práctica, utilizando programación aplicada a diferentes conjuntos de datos



COMILLAS

UNIVERSIDAD PONTIFICIA

ICAI

ICADE

CIHS

GUÍA DOCENTE

2026 - 2027

RESUMEN HORAS DE TRABAJO DEL ALUMNO

Horas Presenciales		
Lecciones de Carácter expositivo	Ejercicios y resolución de casos y de problemas	
15	15	
Horas No Presenciales		
Ejercicios y resolución de casos y de problemas	Estudios individual y/o en grupo, y lectura organizada	Trabajos monográficos y de investigación, individuales o colectivos
15	15	15
CRÉDITOS ECTS: 3 (75 horas)		

EVALUACIÓN Y CRITERIOS DE CALIFICACIÓN

Actividades de evaluación	Criterios de evaluación	Peso
Examen final	Calificación de 0 a 10	
	Es necesario un mínimo de 5 puntos en el examen final para hacer media con el resto de ítems y poder aprobar la asignatura	40
Prueba Corta Intermedia / Midterm 15%	Calificación numérica de 0 a 10	15
Evaluación Continua 10%	Calificación numérica de 0 a 10	10
Trabajo de aplicación práctica. 25% El uso indebido de inteligencia artificial generativa (como por ejemplo ChatGPT) será causa de sanción. Véase NOTA al respecto.	Uso de datos reales y programación Calificación mediante rúbrica Propuesta 20%, 25 defensa 50%, reporta final 30%	
Casos prácticos periódicos 10%	A realizar en clase-casa Calificación numérica 0-10	10

Calificaciones

Para la evaluación de ambas convocatorias, se seguirá el mismo sistema de calificación, con las mismas ponderaciones indicadas previamente. Este sistema responde a la filosofía de que el trabajo continuado forma parte esencial del proceso de aprendizaje de la asignatura.

La evaluación continua, con una ponderación del 10% sobre la calificación final, podrá incluir cualquier actividad propuesta por el profesor a lo largo del curso. Entre estas actividades podrán considerarse, entre otras, la participación activa en clase, la asistencia, la realización de quizzes, talleres, ejercicios prácticos, entregas individuales o grupales, así como la actitud académica y el grado de implicación del estudiante en el desarrollo de la asignatura.

Es necesario obtener una calificación mínima de 5 puntos sobre 10 en el examen final para poder hacer media con el resto de ítems y aprobar la asignatura, tanto en convocatoria ordinaria como en convocatoria extraordinaria.

Es obligatorio realizar el trabajo final, con un mínimo de calidad, para poder aprobar la asignatura tanto en convocatoria ordinaria como en convocatoria extraordinaria. El trabajo constituye una forma de demostrar que se han adquirido las habilidades necesarias para aplicar los conceptos teóricos a datos reales mediante técnicas de análisis adecuadas, así como las competencias de programación requeridas.

ALUMNOS EN CONVOCATORIA EXTRAORDINARIA (2ª): se aplica el mismo sistema de calificación que en convocatoria ordinaria, con



COMILLAS

UNIVERSIDAD PONTIFICIA

ICAI

ICADE

CIHS

GUÍA DOCENTE
2026 - 2027

los mismos porcentajes. El alumno que no haya realizado o que haya suspendido el trabajo práctico deberá realizarlo nuevamente para esta convocatoria.

ALUMNOS EN TERCERA Y POSTERIORES CONVOCATORIAS: en tercera y posteriores convocatorias, el trabajo final será obligatorio. La calificación final se calculará aplicando la siguiente ponderación: 70% examen final y 30% trabajo final. Para poder aprobar la asignatura, será necesario obtener una calificación mínima de 5 puntos sobre 10 tanto en el examen final como en el trabajo final.

ALUMNOS DE INTERCAMBIO (IN): mismo régimen que los alumnos ordinarios.

ALUMNOS DE INTERCAMBIO (OUT): mismo régimen que los alumnos en tercera y posteriores convocatorias.

PLAN DE TRABAJO Y CRONOGRAMA

Actividades	Fecha de realización	Fecha de entrega
Propuesta de Trabajo Final		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.