

GUÍA DOCENTE

PROCESAMIENTO DE LENGUAJE NATURAL Y MINERÍA DE TEXTO

GRADO EN CIENCIA E INGENIERÍA DE DATOS

CURSO 2026-27

Fecha de publicación: 02-07-2026



I.-Identificación de la Asignatura	
Tipo	OPTATIVA
Período de impartición	4 curso, 2Q semestre
Nº de créditos	4.5
Idioma en el que se imparte	Castellano

II.-Presentación
Esta asignatura introduce al estudiantado en los fundamentos y aplicaciones principales del análisis automático de información textual. A lo largo de la asignatura se estudiarán los conceptos básicos del procesamiento del lenguaje natural y la minería de texto, incluyendo los modelos y recursos lingüísticos más habituales, así como las principales técnicas utilizadas para diseñar y desarrollar soluciones de análisis de textos en lenguaje natural. El objetivo principal de la asignatura es que el estudiantado adquiera una visión completa del procesamiento del lenguaje natural y la minería de texto, comprendiendo qué son, cuáles son sus aplicaciones más relevantes y qué tecnologías permiten abordar problemas basados en datos textuales. Dentro del Grado en Ciencia e Ingeniería de Datos, esta asignatura permite aplicar conocimientos previos de programación, aprendizaje automático y análisis de datos a textos en lenguaje natural. Al finalizar la asignatura, el estudiantado será capaz de analizar problemas de ciencia de datos relacionados con datos textuales, identificar cuándo resulta adecuado aplicar técnicas de procesamiento del lenguaje natural y minería de texto, proponer soluciones concretas y desarrollar prototipos para el análisis automático de textos. Esta asignatura también contribuirá al desarrollo de competencias transversales como la resolución de problemas con iniciativa y creatividad, el aprendizaje autónomo, el trabajo en equipo y la comunicación efectiva de conocimientos y resultados relacionados con la materia. Se recomienda haber cursado previamente las asignaturas de programación de cursos anteriores, así como las asignaturas de Aprendizaje Automático I y Aprendizaje Automático II.
II.B-Presentación en inglés



The course introduces students to the fundamentals and main applications of the automatic analysis of textual information. Throughout the course, the basic concepts of Natural Language Processing and Text Mining will be studied, including the most common models and linguistic resources, as well as the main techniques used to design and develop solutions for the analysis of texts in natural language. The main objective of the course is for students to acquire a comprehensive overview of Natural Language Processing and Text Mining, understanding what they are, what their most relevant applications are, and which technologies make it possible to address problems based on textual data. Within the Bachelor's Degree in Data Science and Engineering, this course enables students to apply previous knowledge of programming, machine learning, and data analysis to texts in natural language. By the end of the course, students will be able to analyse data science problems related to textual data, identify when it is appropriate to apply Natural Language Processing and Text Mining techniques, propose specific solutions, and develop prototypes for the automatic analysis of texts. This course will also contribute to the development of transferable skills such as problem-solving with initiative and creativity, autonomous learning, teamwork, and effective communication of knowledge and results related to the subject. It is recommended that students have previously taken the programming courses from previous years, as well as the courses Machine Learning I and Machine Learning II.

III.-Resultados de Aprendizaje

CB05. Que los estudiantes hayan desarrollado aquellas habilidades de aprendizaje necesarias para emprender estudios posteriores con un alto grado de autonomía

CG04. Utilizar y aplicar productos tecnológicos y tendencias de la tecnología, asociados al ámbito de la ciencia e ingeniería de datos.

CG05. Resolver problemas con iniciativa y creatividad.

CG06. Ser capaces de construir modelos, realizar cálculos, redactar informes, planificar tareas y otros trabajos análogos en el ámbito específico de la ciencia e ingeniería de datos, atendiendo a los principios básicos de capacidad de análisis y síntesis, claridad expositiva y rigor científico.

CG09. Acceder y gestionar la información en diferentes formatos para su posterior análisis con el fin de obtener conocimiento a partir de datos

CE10. Emplear distintas técnicas de modelado de datos, análisis de datos y aprendizaje automático: elegir y emplear técnicas adecuadas, evaluando la calidad de los modelos, validándolos, interpretándolos y proponiendo mejoras.

CE11. Diseñar e implementar la toma de datos, su integración, transformación, selección, comprobación de su calidad y veracidad a partir de distintas fuentes, teniendo en cuenta su carácter, heterogeneidad y variabilidad.

- Leyenda para las titulaciones no adaptadas al RD 822/21: **CB** - competencias básicas, **CG** - competencias generales, **CE** - competencias específicas, **CT** - competencias transversales.
- Leyenda para las titulaciones adaptadas al RD 822/21: **CON** - conocimientos, **COM** - competencias, **HAB** - habilidades.



IV.-Contenido

IV.A.-Temario de la asignatura

Tema 1: Introducción al procesamiento del lenguaje natural y la minería de texto

Definición de procesamiento del lenguaje natural y minería de texto. Aplicaciones del procesamiento del lenguaje natural y la minería de texto. Entorno de desarrollo, principales librerías y lenguaje de programación

Tema 2: Análisis textual y recursos lingüísticos

Niveles básicos de análisis textual. Técnicas de preprocesamiento textual. Tareas asociadas a cada nivel de análisis. Herramientas y recursos lingüísticos de apoyo.

Tema 3: Representación del texto

Introducción a la representación computacional del texto. Modelos basados en bolsa de palabras y frecuencias. Representación semántica y modelado de temas. Embeddings estáticos y contextuales.

Tema 4: Modelos de lenguaje

Introducción a las redes neuronales aplicadas al procesamiento de texto. Arquitectura Transformer y mecanismos de atención. Modelos de lenguaje de gran tamaño. Técnicas de postentrenamiento.

Tema 5: Inteligencia artificial generativa

Ingeniería de prompts. Generación aumentada por recuperación. Agentes basados en modelos de lenguaje.

IV.B.-Actividades formativas

Tipo	Descripción
Asistencia a clases teóricas	Lecciones magistrales
Resolución de ejercicios, problemas, casos	Ejercicios con enfoque aplicado
Prácticas	Prácticas con ordenador
Trabajos colectivos	Trabajos grupales con defensa oral
Realización de pruebas	Examen teórico-práctico final
Tutorías académicas	Tutorías personalizadas
Otras actividades	Estudio individual o en grupo



V.-Tiempo de trabajo del alumnado (30h grado y 25h máster)

Clases teóricas	21
Clases de resolución de ejercicios, problemas, casos, etc.	10
Prácticas en laboratorios experimentales, tecnológicos, clínicos, campo, etc.	10
Realización de pruebas	4
Tutorías académicas	32
Actividades relacionadas: jornadas, seminarios, etc.	4
Preparación de clases teóricas	15
Preparación de prácticas/ejercicios/casos	27
Preparación de pruebas	12
Total de horas de trabajo del alumnado	135

VI.-Metodología y plan de trabajo

Tipo	Periodo	Contenido
Clases Teóricas	Semana 1 a Semana 15	Clases magistrales en horario de clase
Resolución de ejercicios, problemas, casos	Semana 1 a Semana 15	Clases de resolución de ejercicios en horario de clase
Prácticas	Semana 1 a Semana 15	Prácticas con ordenador en horario de clase
Trabajos colectivos	Semana 1 a Semana 8	Desarrollo del primer trabajo grupal con defensa oral
Trabajos colectivos	Semana 9 a Semana 15	Desarrollo del segundo trabajo grupal con defensa oral
Tutorías académicas	Semana 1 a Semana 15	Tutorías personalizadas
Otras actividades	Semana 1 a Semana 15	Estudio individual o en grupo para la preparación de la asignatura



VII.-Método de evaluación

El modelo de evaluación general es la evaluación continua, tal como establece el Reglamento de evaluación de los resultados de aprendizaje de la Universidad Rey Juan Carlos.

Deberán utilizarse todos los sistemas de evaluación establecidos para la asignatura en la memoria de la titulación, excepto aquellos que tuviesen una ponderación mínima del 0%, que podrán utilizarse en los cursos académicos en los que el profesorado lo considere oportuno. Cada uno de los sistemas de evaluación podrá ser aplicado mediante una o más actividades de evaluación, coherentes con ese sistema. Ninguna de las actividades de evaluación podrá superar individualmente el 60% de la calificación global de la asignatura.

La suma de las actividades de evaluación no revaluables no podrá superar el 40% de la calificación global de la asignatura y, en general, no deberían tener nota mínima (salvo en el caso de actividades de carácter práctico en las que, estrictamente, no pudieran reproducirse en la convocatoria extraordinaria las condiciones de evaluación de la convocatoria ordinaria).

El alumnado que no consiga superar la asignatura en la convocatoria ordinaria, o no se hayan presentado, podrán presentarse a la convocatoria extraordinaria únicamente a las actividades de evaluación revaluables no superadas.

El uso de herramientas basadas en inteligencia artificial está prohibido en cualquier ejercicio y prueba de evaluación de la asignatura salvo que el profesor lo indique de forma explícita en el enunciado o en el material de esta. En caso de que esté permitido, se darán indicaciones precisas de qué usos están permitidos y cuáles no.

La distribución y características de las actividades de evaluación son las que se describen a continuación.

VII.A.- Descripción de las actividades de evaluación y su ponderación

Convocatoria ordinaria

Sistema de Evaluación	Reevaluable en Extraordinaria	Ponderación	Actividad de evaluación	Nota mínima	Contenidos	Fecha
SE3 – Evaluación de trabajos	Sí. Se seguirá el mismo formato que en la convocatoria ordinaria.	50%	Trabajo grupal con defensa oral (30%)	5	Temas 1-4	Semanas 1-8
			Trabajo grupal con defensa oral (20%)	5	Tema 5	Semanas 8-15
SE1 – Pruebas individuales	Sí. Se seguirá el mismo formato que en la convocatoria ordinaria.	50%	Prueba escrita relacionada con los contenidos teóricos y prácticos de la asignatura (50%)	4	Temas 1-5	Periodo oficial de exámenes

Cálculo de la nota final

La nota final será la media ponderada de las notas de las actividades de evaluación según los porcentajes indicados, siempre y cuando se hayan superado con la nota mínima indicada para cada una de ellas. Teniendo en cuenta lo anterior, para aprobar la asignatura la media ponderada debe ser no inferior a 5. Si una o varias actividades de evaluación no se han superado con la nota mínima, la nota final de la asignatura será la nota media obtenida entre todas las actividades de evaluación realizadas, teniendo en cuenta que no podrá superar un 4. La nota final de quien no se haya presentado a ninguna actividad de evaluación será "No presentado".

Convocatoria extraordinaria

En convocatoria extraordinaria los estudiantes solamente se presentarán a las actividades de evaluación no superadas. Para el cálculo de la nota final en esta convocatoria se utilizará la calificación de las actividades de evaluación superadas en convocatoria ordinaria y las notas obtenidas en las actividades revaluadas. Asimismo, podrán presentarse a la prueba SE1 cuando hayan alcanzado la nota mínima establecida en todas las actividades revaluables, pero la media ponderada final sea inferior a 5. El cálculo de la nota final se realiza tal y como se indica en el apartado anterior.

Convocatoria adelantada

La convocatoria adelantada se evaluará siguiendo el mismo método descrito para la convocatoria ordinaria. Se recomienda al alumnado que vaya a solicitar este tipo de convocatoria que se ponga en contacto con el profesorado con la mayor antelación posible, de forma que pueda tener realizados los trabajos evaluables en la fecha de dicha convocatoria.

Dispensa académica

En el caso de estudiantes con dispensa académica concedida, ésta solo les exime de la asistencia a clase, teniendo que realizar todas las pruebas de evaluación cuyas fechas serán comunicadas al inicio de la asignatura.

Conducta académica

En el caso de **fraude académico** en alguna actividad de evaluación, se otorgará una calificación de cero puntos en dicha actividad lo que, para aquellas actividades con nota mínima superior a cero implica el suspenso en la convocatoria correspondiente.



Se recuerda además que, atendiendo al artículo 9.d). de la **Normativa de Convivencia de la Universidad Rey Juan Carlos** (<https://www.urjc.es/images/Universidad/Presentacion/normativa/normativa%20convivencia%20universitaria.pdf>) el fraude académico en alguna actividad de evaluación se considera falta grave. Las sanciones correspondientes a las faltas graves, según el artículo 11.2 de la referida normativa, son la expulsión temporal de la Universidad, y la pérdida en su caso de los derechos de matrícula.

VII.B.- Evaluación del alumnado con dispensa académica de asistencia a clase

La concesión de Dispensa Académica de Asistencia a Clase (DAAC) no implica que se quede automáticamente eximido de participar en las actividades de evaluación continua ni en las actividades formativas presenciales de asistencia obligatoria establecidas en la guía docente. Una vez concedida la dispensa, el alumnado deberá contactar con el docente, que podría proponerle las adaptaciones que considere convenientes, siempre que garanticen la adquisición y adecuada evaluación de los resultados de aprendizaje previstos. El alumnado deberá mantener a lo largo de curso una comunicación fluida con el docente para que este le proporcione información sobre las fechas en que se realizarán esas actividades formativas y de evaluación, en caso de que su programación no estuviese ya fijada y a disposición del alumnado en el momento de la concesión de la dispensa.

Asignatura con posibilidad de dispensa: Sí

VII.C.- Revisión y reclamación de las actividades de evaluación

Se realizará conforme al Reglamento de evaluación de los resultados de aprendizaje de la Universidad Rey Juan Carlos.

VII.D.- Alumnado con discapacidad o necesidades educativas especiales

Con el fin de garantizar la igualdad de oportunidades, la no discriminación, la accesibilidad universal y la mayor garantía de éxito académico, las adaptaciones curriculares para alumnado con discapacidad o con necesidades educativas especiales serán pautadas por la Unidad de Atención a Personas con Discapacidad, de acuerdo con la normativa que regula el servicio de Atención a Estudiantes con Discapacidad, aprobada por Consejo de Gobierno de la Universidad Rey Juan Carlos.

Para que esas adaptaciones puedan realizarse, será requisito la emisión de un informe de adaptaciones curriculares por parte de dicha Unidad, por lo que el alumnado con discapacidad o necesidades educativas especiales deberán contactar con este servicio para analizar conjuntamente las distintas opciones.

VII.E.- Conducta académica, integridad y honestidad académica

La Universidad Rey Juan Carlos está plenamente comprometida con los más altos estándares de integridad y honestidad académica, por lo que estudiar en la URJC supone asumir y suscribir los valores de integridad y la honestidad académica recogidos en el Código Ético de la Universidad.

Para acompañar este proceso, la Universidad dispone de la Normativa de Convivencia de la Universidad Rey Juan Carlos y de diferentes herramientas (antiplagio, supervisión) que ofrecen una garantía colectiva para el completo desarrollo de estos valores esenciales.

Las herramientas basadas en inteligencia artificial pueden mejorar la experiencia de aprendizaje, pero debe utilizarse de forma ética, legal, honesta y transparente. Por este motivo, deben seguirse las recomendaciones de uso adecuado indicadas por los docentes de la asignatura.



VIII.-Recursos y materiales didácticos

Bibliografía básica

- Daniel Jurafsky y James H. Martin. 2026. Speech and Language Processing (3rd edition).
- Roberto Infante. 2026. AI Agents and Applications: With LangChain, LangGraph, and MCP. Manning
- Jay Alamar y Maarten Grootendorst. 2024. Hands-On Large Language Models. O'Reilly.
- Sebastian Raschka. 2024. Build a Large Language Model (From Scratch). Manning.
- Hobson Lane, Cole Howard y Hannes Max Hapke. 2019. Natural Language Processing in Action. Understanding, analyzing, and generating text with Python. Manning.
- Steven Bird, Ewan Klein, Edward Loper. 2009. Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit. O'Reilly.
- Christopher D. Manning y Hinrich Schütze (1999). Foundations of Statistical Natural Language Processing. The MIT Press.

Bibliografía complementaria

- Christopher M. Bishop y Hugh Bishop. 2024. Deep Learning: Foundations and Concepts 2024th Edition. Springer.
- Sohom Ghosh y Dwight Gunning. 2019. Natural Language Processing Fundamentals. Packt Publishing.
- Akshay Kulkarni, Adarsha Shivananda. 2019. Natural Language Processing Recipes: Unlocking Text Data with Machine Learning and Deep Learning using Python. Apress.
- Ian Goodfellow, Yoshua Bengio y Aaron Courville. 2016. Deep Learning. MIT Press.

IX.-Profesorado

Nombre y apellidos	NATALIA MADRUEÑO SIERRO
Correo electrónico	natalia.madrueño@urjc.es
Departamento	Informática y Estadística
Campus de impartición	Fuenlabrada
Categoría	Profesor/a Ayudante Doctor/a
Titulación académica	Doctor
Responsable de asignatura	Si
Horario de tutorías	Para consultar las tutorías póngase en contacto con el/la profesor/a a través de correo electrónico
Nº de Quinquenios	0
Nº de Sexenios	0
Nº de Sexenios de transferencia	0
Nº de evaluaciones positivas Docencia	0

