

ACTIVIDAD FORMATIVA/FOCUSED COURSE

Información de la AF	
Título	2.11 Representación avanzada del lenguaje y modelos modernos para datos no estructurados
Program	CETIS
ECTS Credits	2
University	Universidad Pontificia Comillas
Descriptor	Esta actividad formativa introduce a los estudiantes en enfoques avanzados para la representación y el análisis del lenguaje en el contexto de la investigación con datos no estructurados. Abarca la transición desde las representaciones numéricas tradicionales del texto hacia los word embeddings, los embeddings contextuales y los modelos basados en transformers. Los estudiantes aprenderán los fundamentos conceptuales de las representaciones vectoriales densas, el papel de la similitud semántica y las ventajas de los modelos modernos de lenguaje para capturar el significado contextual de los datos textuales. El curso también aborda el uso de técnicas de deep learning para el análisis de texto y sus aplicaciones en investigación en empresa, economía y gestión. Se prestará especial atención a la interpretación, evaluación y uso crítico de estos modelos, considerando tanto su potencial metodológico como sus limitaciones en investigaciones empíricas con datos no estructurados.

Profesor(es)	
Nombres	
Departamentos	Métodos Cuantitativos
Despacho	
e-mail	
Horarios de tutoría	

Prerrequisitos
AF: Análisis textual clásico y técnicas interpretativas para datos no estructurados
OBJETIVOS DE APRENDIZAJE Y COMPETENCIAS
CN03: Los principios y elementos integrantes del diseño de una investigación en el ámbito de la Gestión, la Economía y la Empresa. CN04: Los principales enfoques, métodos y técnicas de investigación que permiten responder a problemas de investigación propios de la diversidad de áreas y disciplinas ligadas al territorio de la Gestión, la Economía y la Empresa. CM01: Identificar y formular de modo claro, fundamentado y consistente problemas de investigación relevantes tanto desde la perspectiva científico-académica como social, desde los principios y valores democráticos y el marco de los Objetivos de Desarrollo Sostenible como referentes. CM03: Aplicar enfoques, métodos, análisis y técnicas de investigación propios de las áreas de Gestión, Economía y Finanzas de modo consistente y coherente, desde principios y valores democráticos y con orientación hacia el logro de los Objetivos de Desarrollo Sostenible.

CM04: Reportar y difundir resultados fundamentados de la investigación a través de una comunicación eficaz, oral y escrita, tanto a la comunidad investigadora y profesional relacionada con las problemáticas estudiadas como a públicos no especializados.

CB11: Comprensión sistemática de un campo de estudio y dominio de las habilidades y métodos de investigación relacionados con dicho campo.

CB12: Capacidad de concebir, diseñar o crear, poner en práctica y adoptar un proceso sustancial de investigación o creativo.

CB14: Capacidad de realizar un análisis crítico y de evaluación y síntesis de ideas nuevas y complejas.

Habilidades o destrezas

HA01: Aportar perspectivas originales, críticas y holísticas en el desarrollo y aplicación de investigaciones que permitan abordar la complejidad de los problemas de investigación.

HA02: Apreciar las implicaciones de los enfoques metodológicos alternativos, así como de sus fundamentos onto-epistemológicos y axiológicos, sobre el diseño de la investigación, el análisis, la interpretación y el reporte de datos, y la construcción de teorías.

HA03: Evaluar críticamente los diversos métodos y técnicas de investigación, así como discernir su utilidad para cada contexto de investigación.

HA04: Ser consistentes en el uso de las herramientas de investigación dentro del diseño y el contexto de un proyecto de investigación.

HA06: Saber aplicar e integrar fundamentos teóricos y metodológicos para resolver problemas en entornos nuevos y complejos relacionados con los campos de estudio de la Gestión, la Economía y las Finanzas.

CONTENIDOS

- **Representación Numérica del Texto:** Introducción a las principales formas de transformar textos en representaciones analizables.
- **Word Embeddings:** Fundamentos de las representaciones vectoriales densas del lenguaje.
- **Embeddings Contextuales:** Introducción a las representaciones contextuales del lenguaje.
- **Transformers y Modelos Modernos de Lenguaje:** Fundamentos conceptuales de la arquitectura transformer y de los modelos de lenguaje actuales.

MÉTODOS DE ENSEÑANZA Y APRENDIZAJE

La asignatura está diseñada con una orientación eminentemente práctica y aplicada. Los estudiantes deberán prepararse antes de cada clase mediante la revisión de los contenidos que se tratarán, lo que permitirá un mejor aprovechamiento de las sesiones presenciales. Durante estas sesiones, se reforzarán los contenidos teóricos y se explicarán los conceptos fundamentales de la representación avanzada del lenguaje y del uso de modelos modernos para el análisis de datos no estructurados.

Las clases combinarán lecciones expositivas con ejercicios prácticos y resolución de casos aplicados. Los estudiantes trabajarán con ejemplos sencillos que les permitirán comprender la lógica de las representaciones vectoriales del texto, los embeddings, los modelos basados en transformers y sus principales aplicaciones en contextos de investigación.

Las actividades prácticas constituirán un componente clave de la asignatura y se centrarán en el uso guiado de herramientas no-code o low-code para la representación y análisis de datos no estructurados mediante embeddings, embeddings contextuales y modelos modernos de lenguaje. Se realizarán prácticas tanto en clase como fuera del aula, lo que permitirá a los estudiantes desarrollar y perfeccionar sus habilidades de forma continua, sin que la programación avanzada sea un requisito central. Además, los estudiantes trabajarán en un proyecto práctico aplicado a problemas empresariales reales, facilitando la interpretación crítica de técnicas avanzadas en contextos empresariales y económicos.

ACTIVIDADES REALIZADAS EN EL AULA	COMPETENCIAS
- 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.	CN03, CN04, CM01, CM03, CM04, CB11, CB12, CB14

ACTIVIDADES FUERA DEL AULA	COMPETENCIAS
- Estudio personal - Realización de casos y trabajos de aplicación práctica	CN04, CM01, CM03, CM04, CB11, CB12, CB14

METODOS DE EVALUACION		
ACTIVIDAD DE EVALUACIÓN	CRITERIO	PESO
Examen Final de la actividad 30%	Examen Final: Calificación numérica de 0 a 10.	60%
Trabajo final de aplicación práctica 30%	Trabajo Final de aplicación práctica: Uso de datos reales y programación Calificación mediante rúbrica Propuesta 10%, Trabajo final 40%, presentación 50%	
Evaluación continua Individual 30%	Calificación numérica de 0 a 10	30%
Casos prácticos periódicos en grupo 10%	A realizar en clase-casa Calificación numérica 0-10	10%

La no asistencia al 75% de las horas presenciales supone un no apto. Las AF solo tienen una convocatoria. Si no se superan se deberá tomar el curso siguiente.

Se permite el uso responsable y declarado de herramientas de inteligencia artificial generativa como apoyo auxiliar para el aprendizaje, la mejora de redacción, la comprensión de conceptos o la asistencia técnica básica. No obstante, el estudiante deberá mantener la autoría intelectual del trabajo presentado y será responsable de verificar la calidad, exactitud y pertinencia de cualquier contenido generado con apoyo de estas herramientas.

Todo uso de IA en trabajos evaluables deberá declararse explícitamente, indicando la herramienta empleada, la finalidad de uso y las partes del trabajo en las que ha intervenido. No se permitirá el uso de IA para generar entregables completos, respuestas de examen, análisis, interpretaciones, conclusiones o código que se presenten como propios sin supervisión crítica y sin declaración expresa.

El uso indebido de ChatGPT u otras herramientas de inteligencia artificial generativa (IAG) será considerado una falta grave según lo establecido en el Reglamento General de la Universidad, art. 168.2.e, como: «la realización de acciones encaminadas a falsear o defraudar los sistemas de evaluación del rendimiento académico».

Las consecuencias de ello serán «la expulsión temporal de hasta tres meses o la prohibición de presentarse a un examen en la convocatoria siguiente a la imposición de la sanción, en una o varias asignaturas en las que el estudiante esté matriculado, [...] además de la calificación de suspenso (0) en la asignatura correspondiente, [...] [y] la prohibición de presentarse a examen de dicha asignatura en la siguiente convocatoria».

RESUMEN DEL TRABAJO DEL ESTUDIANTE		
HORAS PRESENCIALES		
CLASES MAGISTRALES	EJERCICIOS Y OTRAS ACTIVIDADES INDIVIDUALES EN EL AULA	EXAMEN O TEST
6	3	5
OUT OF CLASS HOURS		
ESTUDIO O TRABAJO INDEPENDIENTE	TRABAJO EN GRUPO	
24	12	
CRÉDITOS ECTS :		2
RECURSOS DE APRENDIZAJE		
• Ignatow, G., & Mihalcea, R. (2017). An introduction to text mining: Research design, data collection, and analysis. Sage Publications. • Ihrmark, Daniel, and Jukka Tyrkko. "Learning text analytics without coding? An introduction to KNIME." Education for Information 39.2 (2023): 121-137. • Zong, C., Xia, R., & Zhang, J. (2021). Text data mining (Vol. 711, p. 712). Singapore: Springer. • Hofmann, Markus, and Andrew Chisholm, eds. Text mining and visualization: Case studies using open-source tools. CRC Press, 2016.		

COURSE INFORMATION	
Title	2.11. Advanced Language Representation and Modern Models for Unstructured Data
Program	CETIS PHD
ECTS Credits	2
University	Universidad Pontificia Comillas
Descriptor	This focused course introduces students to the foundations of classical text analytics for the analysis of unstructured data. It covers the main stages of the text analysis workflow, including corpus construction, text cleaning, preprocessing, tokenization, normalization and exploratory analysis. Students will learn to describe textual collections through frequencies, n-grams and visualizations, and to apply interpretive techniques such as topic modeling and sentiment analysis. The activity emphasizes the use of these methods to identify latent themes, patterns, opinions and perceptions in textual data, with applications to research problems in business, economics and management.

TEACHERS	
Name	
Department	
Office	
e-mail	
Tutorials	

Prerequisites
AF: Classical Text Analytics for Topic Modeling and Sentiment Analysis
LEARNING GOALS
CN03: The principles and constituent elements of research design in the field of Management, Economics and Business. CN04: The main research approaches, methods and techniques that make it possible to address research problems specific to the diverse areas and disciplines related to the field of Management, Economics and Business. Competences CM01: To identify and formulate, in a clear, well-founded and consistent manner, relevant research problems from both a scientific-academic and social perspective, taking democratic principles and values and the framework of the Sustainable Development Goals as points of reference. CM03: To apply research approaches, methods, analyses and techniques specific to the areas of Management, Economics and Finance in a consistent and coherent manner, based on democratic principles and values and oriented towards the achievement of the Sustainable Development Goals. CM04: To report and disseminate research findings through effective oral and written communication, both to the research and professional communities related to the issues studied and to non-specialized audiences. CB11: Systematic understanding of a field of study and mastery of the research skills and methods related to that field. CB12: Ability to conceive, design or create, implement and adopt a substantial research or creative process. CB14: Ability to carry out critical analysis, evaluation and synthesis of new and complex ideas. Skills or abilities

HA01: To contribute original, critical and holistic perspectives in the development and application of research that addresses the complexity of research problems.

HA02: To appreciate the implications of alternative methodological approaches, as well as their onto-epistemological and axiological foundations, for research design, data analysis, interpretation and reporting, and theory building.

HA03: To critically evaluate different research methods and techniques, and to discern their usefulness in each research context.

HA04: To use research tools consistently within the design and context of a research project.

HA06: To apply and integrate theoretical and methodological foundations to solve problems in new and complex environments related to the fields of Management, Economics and Finance.

CONTENTS

- **Numerical Representation of Text:** Introduction to the main ways of transforming texts into analyzable representations.
- **Word Embeddings:** Foundations of dense vector representations of language.
- **Contextual Embeddings:** Introduction to contextual language representations.
- **Transformers and Modern Language Models:** Conceptual foundations of transformer architecture and current language models.

LEARNING AND TEACHING METHOD

The course is designed with a strongly practical and applied orientation. Students are expected to prepare before each class by reviewing the contents to be covered, allowing them to make the most of the face-to-face sessions. During these sessions, theoretical contents will be reinforced and the fundamental concepts of advanced language representation and the use of modern models for unstructured data analysis will be explained.

Classes will combine lectures with practical exercises and the resolution of applied cases. Students will work on simple examples that will allow them to understand the logic of vector representations of text, embeddings, transformer-based models and their main applications in research contexts.

Practical activities will be a key component of the course and will focus on the guided use of no-code or low-code tools for the representation and analysis of unstructured data through embeddings, contextual embeddings and modern language models. Practical exercises will be carried out both in class and outside the classroom, enabling students to develop and refine their skills on an ongoing basis, without advanced programming being a central requirement. In addition, students will work on a practical project applied to real business problems, facilitating the critical interpretation of advanced techniques in business and economic contexts.

CLASS-BASED ACTIVITIES	COMPETENCIES
• Presentation of the main theoretical concepts • Development of simple application examples • Group discussion and correction of cases and problems completed by students • Assessment activities • Activities such as seminars, workshops, simulations or group dynamics	CN03, CN04, CM01, CM03, CM04, CB11, CB12, CB14
OUT OF CLASS ACTIVITIES	COMPETENCIES
• Personal study • Completion of cases and practical application assignments	CN04, CM01, CM03, CM04, CB11, CB12, CB14, CB15

ASSESSMENTS AND ASSESSMENT CRITERIA

ASSESSMENT ACTIVITIES	CRITERIA	WEIGHT
Final exam for the activity: 30% Final practical application project: 30%	numerical grade from 0 to 10.	60%
Individual continuous assessment	numerical grade from 0 to 10.	30%
Periodic group practical cases	To be completed in class and at home. Numerical grade from 0 to 10.	10%

Non-attendance to 75% of the lectures/workshops will imply a non-pass. Students will have to take the AF again the following year, as there is only one sitting.

The responsible and declared use of generative artificial intelligence tools is permitted as auxiliary support for learning, writing improvement, concept understanding or basic technical assistance. However, students must retain intellectual authorship of the work submitted and will be responsible for verifying the quality, accuracy and relevance of any content generated with the support of these tools.

Any use of AI in assessed work must be explicitly declared, indicating the tool used, the purpose of its use and the parts of the work in which it was involved. The use of AI to generate complete submissions, exam answers, analyses, interpretations, conclusions or code presented as the student's own work without critical supervision and explicit disclosure will not be permitted.

The improper use of ChatGPT or other generative artificial intelligence (GAI) will be considered a serious offense as stated in the General Regulations of the University, art. 168.2.e as: "carrying out actions aimed at falsifying or defrauding the evaluation systems of academic performance". The consequences of this will be "temporary expulsion for up to three months or the prohibition to take an exam in the following call to the imposition of the sanction, in one or more subjects in which the student is enrolled, [...] apart from the grade of failure (0) in the respective subject, [...] [and] the prohibition to take an exam in that subject in the following call".

SUMMARY OF STUDENT WORKLOAD		
CLASS BASED HOURS		
LECTURES	ACTIVITY BASED	TESTS
6	3	5
OUT OF CLASS HOURS		
INDEPENDENT STUDY/WORK	GROUP WORK	
24	12	
ECTS CREDITS:		2
RESOURCES		
- Ignatow, G., & Mihalcea, R. (2017). An introduction to text mining: Research design, data collection, and analysis. Sage Publications. - Ihrmark, Daniel, and Jukka Tyrkko. "Learning text analytics without coding? An introduction to KNIME." Education for Information 39.2 (2023): 121-137. - Zong, C., Xia, R., & Zhang, J. (2021). Text data mining (Vol. 711, p. 712). Singapore: Springer. - Hofmann, Markus, and Andrew Chisholm, eds. Text mining and visualization: Case studies using open-source tools. CRC Press, 2016.		