

Datos Alternativos en Finanzas

ENFIN829		3 Créditos
Profesor (es)	Ramiro H. Gálvez	
Ayudantes	-	

DESCRIPCIÓN DEL CURSO

En los últimos años, ha habido un aumento significativo en la cantidad de datos en formato de texto disponibles electrónicamente. En el ámbito financiero, estos datos incluyen informes anuales, comunicados de prensa, artículos de periódicos y contenido generado por usuarios en redes sociales. Estos datos son una valiosa fuente de información para la toma de decisiones, pero su procesamiento y análisis requiere habilidades adicionales en comparación con los datos tradicionales.

El curso "*Datos Alternativos en Finanzas*" ofrece una introducción al uso de técnicas provenientes del procesamiento del lenguaje natural (NLP) aplicadas al análisis financiero. Los participantes explorarán un amplio espectro de técnicas, tanto introductorias como avanzadas, que incluyen el preprocesamiento de textos, la representación de textos mediante bolsas de palabras, la representación de palabras a través de embeddings, así como también una introducción a los modelos de atención y a los grandes modelos de lenguaje (LLMs).

El curso combina teoría y práctica de manera integral (*hands-on*), haciendo foco en trabajar con aplicaciones reales de análisis financiero. Utilizando Python como herramienta principal, los estudiantes adquirirán habilidades prácticas para implementar soluciones de NLP en una variedad de aplicaciones financieras, como el análisis de sentimiento, la clasificación de documentos y la detección de entidades en textos financieros, entre otras.

Al finalizar el curso, los participantes no solo habrán adquirido conocimientos teóricos sólidos, sino que también se llevarán consigo una amplia base de código útil y aplicable en sus ámbitos laborales.

I.- OBJETIVOS DE APRENDIZAJE DEL CURSO

Proporcionar a los participantes una comprensión integral de cómo aplicar técnicas de procesamiento del lenguaje natural (NLP) en el análisis financiero, capacitándolos para utilizar eficazmente herramientas y metodologías en la era de los datos alternativos.

1. Familiarizar a los participantes con los fundamentos teóricos del NLP y su aplicación en el ámbito financiero.

2.	Dotar a los participantes con habilidades prácticas para recolectar, preprocesar y analizar datos financieros utilizando técnicas avanzadas de NLP.	
3.	Capacitar a los participantes para implementar soluciones efectivas de NLP en diversas aplicaciones financieras, utilizando Python como herramienta principal.	

II.- CONTENIDOS		
Tema	Contenido	Lectura Obligatoria
1	Introducción al Procesamiento de Lenguaje Natural (NLP) y sus aplicaciones en finanzas	Artículos seleccionados.
2	Preprocesamiento de texto: Tokenización y normalización de texto (eliminación de stopwords, <i>stemming</i> , lematización), representación <i>bag-of-words</i> , ponderación <i>TF-IDF</i> , similitud coseno.	Dan Jurafsky & James H. Martin “ Speech and Language Processing ” (3rd ed. draft)
3	Breve introducción al aprendizaje supervisado: Definición de aprendizaje supervisado (clasificación, vs. regresión), importancia de disponer de conjunto de validación, repaso de regresión logística regularizada, métricas de evaluación (<i>accuracy</i> , AUC-ROC).	James, Witten, Hastie & Tibshirani, “ An Introduction to Statistical Learning with Applications in Python ”, Springer, 2023 (ISLP)
4	Embeddings no contextuales de palabras: Introducción a Word2Vec, alternativas a Word2Vec (GloVe, fast-text, etc.), limitaciones del enfoque no contextual	Dan Jurafsky & James H. Martin “ Speech and Language Processing ” (3rd ed. draft)
5	Breve introducción a la arquitectura transformers: Intuiciones detrás del mecanismo de atención y auto-atención, modelos BERT, <i>fine-tuning</i> de modelos transformers pre-entrenados, otras arquitecturas (RoBERTa, GPT, etc.)	Dan Jurafsky & James H. Martin “ Speech and Language Processing ” (3rd ed. draft)
6	Chatbots basados en modelos grandes de lenguaje (LLMs): Introducción a los Chatbots basados en LLMs, prompt-engineering, <i>one-shot learning</i> y <i>few-shot learning</i> utilizando chatbots basados en LLMs, introducción a la API de Gemini.	Dan Jurafsky & James H. Martin “ Speech and Language Processing ” (3rd ed. draft) Xianzhi Li et. al “ Are ChatGPT and GPT-4 General-Purpose Solvers for Financial Text Analytics? A Study on Several Typical Tasks ”. EMNLP: Industry Track.
7	En caso de disponer de tiempo Introducción a scraping: urllib, estructura del código HTML, BeautifulSoup, presentación de selenium	Slides autocontenidas

III.- METODOLOGÍA, EVALUACIÓN Y NORMATIVA BÁSICA

3.1.- Metodología:

En casa clase combinará tanto contenido teórico como aplicaciones prácticas. Se utilizará el lenguaje de programación Python. Se pide que todos los alumnos asistan a clases con una computadora portátil y que, antes de la primera clase, la misma tenga instalada y corriendo adecuadamente la distribución Anaconda de Python (se puede descargar en <https://www.anaconda.com/download>).

3.2.- Evaluación:

- Participación y asistencia a las actividades del curso. 20% de la nota final.
- Durante la cursada se entregarán tres talleres para realizar. Los alumnos, en grupos de no más de dos personas, deberán elegir uno de dichos talleres y entregarlo resuelto. El taller podrá ser aprobado o reprobado. En caso de reprobalo, se podrá re-entregar una única vez. Es condición para la aprobación del curso haber aprobado el taller. 30% de la nota final
- Se tomará un examen final con preguntas de opción múltiple y preguntas de tipo verdadero o falso. El mismo se realizará a través del campus virtual de la Universidad. 50% de la nota final.

3.3- Normativa Básica

1. Los/las estudiantes deberán tener al menos un 75% de asistencia en el curso para poder aprobarlo.
2. Las clases serán los días y horas indicados por la dirección del programa.
3. Para cada clase los/las estudiantes deberán haber leído y estudiado anticipadamente la bibliografía correspondiente.
4. La calificación de todas las evaluaciones se hará con nota de 1 a 7.
5. El/la profesor/a se reserva el derecho de agregar, eliminar o reemplazar bibliografía durante el transcurso del programa si así lo estimara conveniente para la buena marcha de la asignatura.
6. La ausencia injustificada de un/a estudiante a una exigencia será calificada con nota 1.
7. Es importante enfatizar que cada estudiante debe asumir su propia responsabilidad en el cumplimiento del programa, especialmente en lo relativo a:
 - a. Estar al día en el desarrollo de la materia y de las diversas indicaciones que entregue tanto el/la profesor/a como la coordinación del curso. Por ejemplo, la ausencia a una sesión de clases no lo exime de las obligaciones académicas señaladas ese día.
 - b. Velar por el fiel cumplimiento de las fechas y plazos establecidos para las distintas actividades de evaluación. Una vez fijadas y conocidas no se procederá a modificarlas.
 - c. Obtener el material de apoyo indicado para la cátedra cuando corresponda.
8. Todos los trabajos que se presenten en el transcurso del programa solo tendrán valor en la medida en que su autor sea capaz de explicar y respaldarlos personalmente. No se aceptan entregas que contradigan lo anterior. Toda justificación médica correspondiente a la inasistencia a una exigencia debe ser presentada a través de los canales regulares establecidos por la Universidad.
9. Toda forma de **copia y/o plagio** está penalizada y en caso de identificarse esta situación, se

seguirá el [procedimiento disciplinario respectivo](#).

IV.- BIBLIOGRAFÍA

Lecturas Obligatorias

- Dan Jurafsky & James H. Martin, "[Speech and Language Processing](#)" (3rd ed. draft).
- James, Witten, Hastie & Tibshirani, "[An Introduction to Statistical Learning with Applications in Python](#)", Springer, 2023 (ISLP).
- Artículos seleccionados.

*Programa sujeto a cambios