

Texto como Datos: Métodos Computacionales en Ciencias Sociales

Universidad Católica del Uruguay

Julio 2026

Profesora:	Ayelén Vanegas	Email:	avanegas@ucu.edu.uy
Horarios de Consulta:	A definir	Oficina:	402 (o Microsoft Teams)
Reservar una reunión:			

Este curso introduce a estudiantes y profesionales de las ciencias sociales en el uso de técnicas computacionales para el análisis de grandes volúmenes de información textual. En un contexto en el que los discursos de política pública, las noticias y las redes sociales ocupan un lugar central en la comprensión de los fenómenos sociales, el análisis computacional de texto se ha convertido en una herramienta clave para la investigación empírica.

Se trata de un curso introductorio que articula enfoques provenientes del *Natural Language Processing* (NLP), el *machine learning* y las ciencias sociales, con énfasis en aplicaciones prácticas al análisis del discurso político, la detección de sentimientos y la comunicación de políticas públicas. A lo largo del curso, los participantes aprenderán técnicas esenciales de preprocesamiento textual (tokenización, stemming y eliminación de *stopwords*), así como metodologías básicas de aprendizaje no supervisado, aprendizaje supervisado y NLP neuronal.

Objetivos del curso

Al finalizar el curso, los estudiantes serán capaces de:

- Aplicar herramientas computacionales para analizar grandes corpus textuales.
- Utilizar modelos estadísticos y de *machine learning* para representar y analizar contenido textual.
- Interpretar fenómenos sociales a partir de datos textuales de manera crítica y metodológicamente informada.

Estructura del curso

Clase 1: Introducción a Machine Learning y Análisis de Texto

- Introducción a *Text as Data* en las ciencias sociales.
- Procesamiento y comparación de textos: librerías para tokenización y vectorización.

Referencias

Justin Grimmer, Margaret E. Roberts y Brandon M. Stewart, *Text as Data: A New Framework for Machine Learning and the Social Sciences* (Princeton, NJ: Princeton University Press, 2022).

Clase 2: Preprocesamiento de Texto y Aprendizaje No Supervisado

- *Web scraping* para la recolección de datos textuales.

- Aprendizaje no supervisado: modelos de tópicos.

Referencias

Roberts, M.E., Stewart, B.M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S.K., Albertson, B. and Rand, D.G. (2014), Structural Topic Models for Open-Ended Survey Responses. *American Journal of Political Science*, 58: 1064-1082.

Clase 3: Aprendizaje Supervisado

- Diccionarios, léxicos y palabras clave.
- Análisis de sentimientos.
- *Keyword-Assisted Topic Models*.

Referencias

Bestvater SE, Monroe BL. Sentiment is Not Stance: Target-Aware Opinion Classification for Political Text Analysis. *Political Analysis*. 2023;31(2):235-256.

Eshima, S., Imai, K. and Sasaki, T. (2024), Keyword-Assisted Topic Models. *American Journal of Political Science*, 68: 730-750.

Young, L., Soroka, S. (2012). Affective News: The Automated Coding of Sentiment in Political Texts. *Political Communication*, 29(2), 205–231.

Clase 4: NLP Neuronal y Embeddings

- Word2Vec y GloVe.
- Modelos de regresión con embeddings (*Embedding Regression*).

Referencias

Rodriguez, P. L., Spirling, A., & Stewart, B. M. (2023). Embedding Regression: Models for Context-Specific Description and Inference. *American Political Science Review*, 117(4), 1255–1274.

Rodriguez, P. L., & Spirling, A. (2022). Word Embeddings: What Works, What Doesn't, and How to Tell the Difference for Applied Research. *The Journal of Politics*, 84(1), 101–115.

Clase 5: Trabajo final y presentación de pósters

Actividad de cierre del curso orientada a la aplicación práctica de las herramientas aprendidas. Los/as participantes desarrollarán un análisis de texto aplicado a un problema social, institucional o comunicacional, utilizando al menos dos de las técnicas vistas en el curso (preprocesamiento, aprendizaje supervisado o no supervisado, embeddings). El trabajo se presentará en formato póster, enfatizando la claridad en la comunicación de resultados, la interpretación sustantiva y la reflexión crítica sobre las limitaciones del análisis.