

**Nombre del curso: Análisis Computacional de Texto para las Ciencias Sociales**

Escuela de Invierno en Métodos

Universidad Católica del Uruguay

Modalidad: presencial

Docente: Naim Bro

Correo: naim.bro.k@uai.cl

**Descripción general del curso:** Este curso explora herramientas y metodologías para el análisis computacional de texto aplicadas a las ciencias sociales. A través de una combinación de conceptos teóricos y prácticos, los participantes aprenderán a extraer, procesar y analizar datos textuales utilizando herramientas de machine learning, así como a interpretar los resultados desde una perspectiva social.

**Objetivos del curso:**

1. Entender los fundamentos teóricos del análisis computacional de texto en ciencias sociales.
2. Manejar herramientas básicas para la extracción y preprocesamiento de datos textuales.
3. Interpretar resultados y comunicar hallazgos de forma efectiva.

**Público objetivo:** Estudiantes, investigadores y profesionales interesados en aplicar el análisis computacional de texto en el ámbito de las ciencias sociales.

**Prerrequisitos:** No hay prerrequisitos.

**Modalidad:** El curso es presencial e incluirá actividades como clases lectivas, talleres prácticos y discusiones grupales. Cada sesión combinará teórica y práctica.

**Estructura del curso:**

**Módulo 1: Fundamentos del Análisis de Texto**

- Descripción: Introducción al análisis de texto y conceptos básicos de procesamiento de lenguaje natural (NLP) aplicados a las ciencias sociales.
- Técnicas: Modelos de lenguaje básicos, introducción a Python, NLTK y Jupyter Notebook, tokenización, lematización y stemming.
- Lecturas:
  - Grimmer, J., Roberts, M. E., & Stewart, B. M. (2022). Text as Data: A New Framework for Machine Learning and the Social Sciences.

**Módulo 2: Representaciones Textuales y Modelado de Tópicos**

- Descripción: Exploración de técnicas para representar textos y descubrir patrones latentes en grandes volúmenes de datos.
- Técnicas: TF-IDF y extracción de características textuales, modelado de tópicos (LDA, Correlated Topic Models), representaciones distribuidas (Word2Vec, Doc2Vec, y GloVe), detección de relaciones semánticas con embeddings.
- Lecturas:
  - Sarmiento et al (2022), "*Identifying and Characterizing New Expressions of Community Framing during Polarization*"

### **Módulo 3: Análisis de Redes y Opinión Pública**

- Descripción: Técnicas avanzadas para analizar conexiones sociales y opiniones utilizando datos textuales.
- Técnicas: Extracción de redes sociales desde texto, análisis de sentimientos (léxicos y supervisados), reconocimiento de entidades nombradas (NER), combinación de NLP y análisis de redes.
- Lecturas:
  - Bro (2025), "*A frustratingly easy way of extracting political networks from text*"

### **Módulo 4: LLMs para el Análisis Cualitativo**

- Descripción: Uso de LLMs para el análisis cualitativo avanzado en ciencias sociales.
- Técnicas: Análisis cualitativo con modelos generativos, comunicación de hallazgos.
- Lecturas:
  - Bosley (2025), "*Improving Probabilistic Models in Text Classification via Active Learning*"

### **Bibliografía:**

- Grimmer, J., Roberts, M. E., & Stewart, B. M. (2022). Text as Data: A New Framework for Machine Learning and the Social Sciences.
- Bro (2025), "*A frustratingly easy way of extracting political networks from text*"
- Bosley (2025), "*Improving Probabilistic Models in Text Classification via Active Learning*"
- Sarmiento et al (2022), "*Identifying and Characterizing New Expressions of Community Framing during Polarization*"