



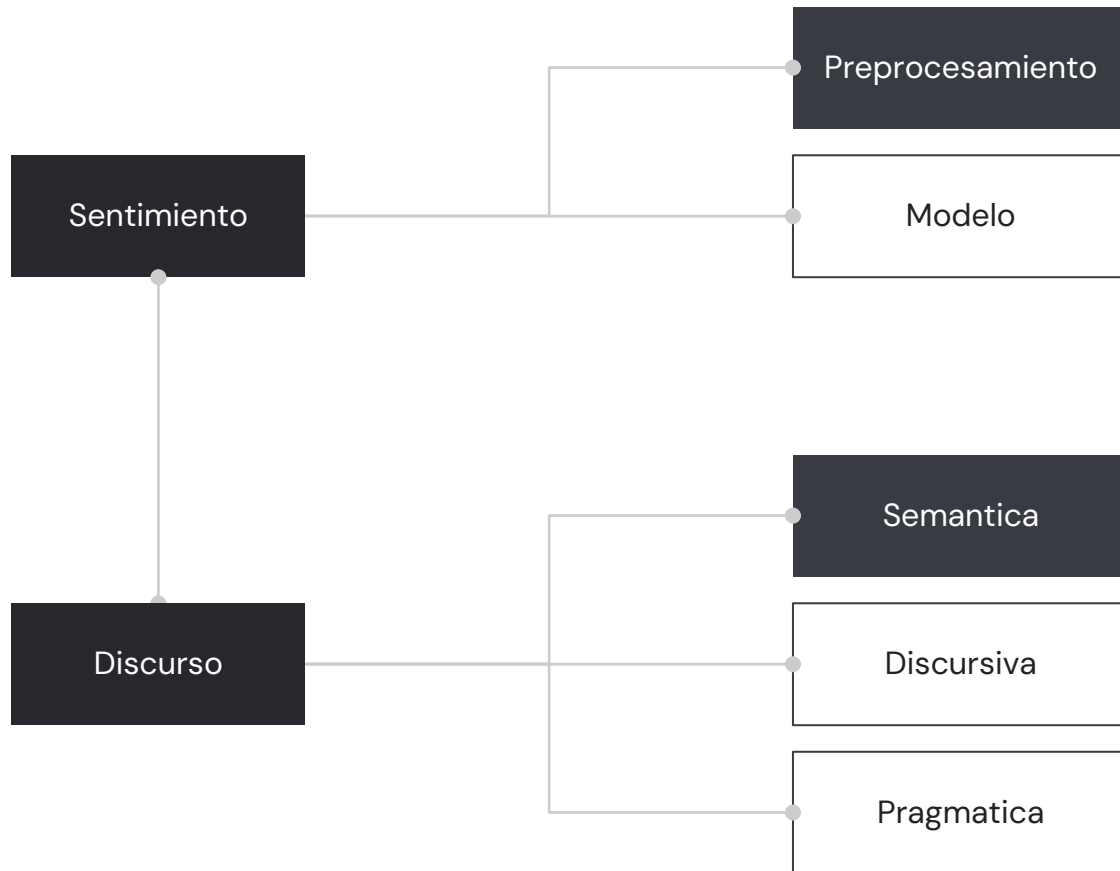
**¡Les damos la
bienvenida!**

¿Comenzamos?

Clase 4. NLP & Deep Learning aplicado a Ciencia de Datos

Taller NLP

MAPA DE CONCEPTOS





Preprocesamiento de Texto

Deberás entregar un dataset del estilo texto (libro, paper, documento, colección de tweets, etc) donde se desarrollen de mínima dos de las tareas más usuales de preprocesamiento de NLP: stemming, tokenización, remoción de stopwords, lematización, etc.

Preprocesamiento de Texto

Consigna

- ✓ Seleccionar un cuerpo de texto que cumpla con alguno de los estilos comentados en la clase (ej, libro).
- ✓ Preferentemente hostear el mismo en GitHub o vincularlo via URL.
- ✓ Cargar el mismo en la notebook usando la función que corresponda.
- ✓ Aplicar al menos dos de las funciones presentadas cómo preprocesamiento de NLP (recordando que es un mínimo sugerido pueden aplicarse más):
 - Lematización;
 - Tokenización;
 - Remoción de stopwords;
 - Pos-tagging.
 - Etc.

Para trabajo final si se opta por NLP debe adicionarse alguna de las siguientes tareas:

- ✓ Realizar un sentiment analysis/TF-IDF vectorizer/otra sobre el texto seleccionado.
- ✓ Evaluar la polaridad de alguna de las sentencias involucradas.

Preprocesamiento de Texto

Formato

- ✓ Entregar un archivo con formato .ipynb.
Debe tener el nombre
"Datasets+Apellido.ipynb".

Sugerencias

- ✓ Preparar el código y probar los resultados con distintas entradas



¿Aún quieres conocer más?
Te recomendamos el
siguiente material



MATERIAL AMPLIADO

Recursos Adicionales

Título

- ✓ [NLP – Chapter 1](#) | **Hugging Face**
- ✓ [Speech and Language Processing](#) | **Daniel Jurafsky**
- ✓ [CS 224 deep learning](#) | **Stanford**

¿Preguntas?

Opina y valora
esta clase

Muchas gracias.

Educación digital para el mundo real

CODERHOUSE