Trabajo N°2 - Procesamiento de Texto con NLP

# Objetivo General:

Aplicar técnicas de procesamiento de lenguaje natural (NLP) para analizar y extraer información relevante de textos en español, usando herramientas de Python.

# Instrucciones Generales:

- El trabajo puede realizarse de forma individual o en parejas.  
- Deben entregar un informe en Jupyter Notebook (.ipynb) con comentarios, explicaciones y resultados visuales y el dataset escogido.  
- Se evaluará el código, la limpieza del texto, la visualización y el análisis final.

# Estructura del Trabajo:

## 1. Carga y exploración del dataset

- Descargar o cargar un archivo .csv o .txt con textos en español.  
- Mostrar ejemplos de algunas filas o textos.

## 2. Preprocesamiento de texto

Aplicar las siguientes transformaciones:  
- Conversión a minúsculas  
- Eliminación de signos de puntuación  
- Eliminación de stopwords  
- Tokenización  
- Lematización (usando spaCy)

## 3. Análisis exploratorio

- Palabras más frecuentes (antes y después del preprocesamiento)  
- Nube de palabras (wordcloud)  
- Distribución de longitud de textos

## 4. Aplicación de técnicas NLP

Aplicar al menos dos de las siguientes técnicas:  
- Análisis de sentimiento  
- Extracción de entidades (NER)

## 5. Conclusiones

- Insights obtenidos del análisis.  
- Ideas para continuar o mejorar el análisis.

# Rúbrica de Evaluación (20 puntos + Bonus):

|  |  |
| --- | --- |
| Criterio | Puntaje |
| Carga y exploración de datos | 2 pts |
| Preprocesamiento completo | 4 pts |
| Visualizaciones efectivas | 3 pts |
| Técnicas NLP aplicadas | 6 pts |
| Conclusiones y presentación final | 3 pts |
| Bonus: Clasificador funcional | +2 pts |

# Fecha de Entrega:

Martes 22 de abril   
Formato de entrega: .ipynb (Notebook) + PDF con resultados visibles.

Enviar a: diego.miranda@academico.cftlr.cl