

## Entrega 2 – EDA Avanzado, Representación y Propuesta Metodológica

Profundizar en el análisis del corpus, mostrar la representación de los textos según el contexto del proyecto y proponer una metodología de modelado inicial.

---

### 1. Correcciones y modificaciones respecto a la Entrega 1

- Indicar qué ajustes se realizaron al corpus, a las visualizaciones o al planteamiento inicial.
- Explicar brevemente las razones de los cambios.

---

### 2. Análisis Exploratorio Avanzado

- Ampliar el estudio descriptivo del corpus.
- Según el caso, integrar **ontologías**, **tesauros** u otros **recursos léxicos**.
- Incluir métricas de diversidad **léxica**, coocurrencias, **bigramas/trigramas** u otros análisis pertinentes.

---

### 3. Conjunto de datos medido

- Entregar el corpus en su versión actualizada (ya preprocesado y acompañado de las métricas obtenidas).
- El dataset debe estar organizado en un formato legible (CSV, JSON, parquet, etc.) e incluir las variables relevantes (ej.: id documento, texto limpio, tokens, longitudes, categorías iniciales, etc.).
- Adjuntar el archivo en la carpeta/repositorio junto al informe.

---

### 4. Representación de los textos

Realizar un análisis inicial del corpus que incluya:

- Seleccionar y justificar la(s) técnica(s) de representación más adecuadas al proyecto (**BoW**, **TF-IDF**, **embeddings**, representaciones contextuales, etc.).
- Mostrar ejemplos de cómo los documentos fueron transformados en vectores o representaciones numéricas.

---

### 5. Propuesta metodológica de los modelos

- Definir qué tipo de enfoque(s) de modelado se planea usar en la fase final (clasificación supervisada, clustering, topic modeling, análisis de sentimiento, etc.).
- Justificar por qué esa metodología es adecuada para el objetivo.
- Especificar si se planea comparar más de un modelo y bajo qué criterios.

---

### 6. Presentación en PPT o Canva

- Crear una presentación que resuma los ítems más relevantes del avance del proyecto.
- Extensión sugerida: 10 –15 diapositivas.
- Debe ser clara, sintética y visual.

---

### 7. Formato de entrega

- Informe en PDF.
- Repositorio/carpeta compartida con los notebooks o scripts utilizados y los datos.