

## **Semana 8: (8 al 12 de Mayo)**

### **Tareas realizadas:**

- ☒ Se realiza el preprocesamiento de un dataset en inglés. Esto nos permitirá comparar los resultados obtenidos en ambos idiomas, para así determinar si el lenguaje del corpus puede influir en los resultados. El dataset ha sido procesado para que esté en igualdad de condiciones que el dataset en español, es decir, misma cantidad de datos por categoría, mismo preprocesamiento, entre otros. Puede consultarse el notebook con el preprocesamiento en el repositorio del proyecto<sup>1</sup>.
- ☒ Búsqueda y lectura de referencias para redacción del documento.
- ☒ Se realizó el avance en el documento en una reunión, incluyendo el apartado de metodología y se realizaron las correcciones.

### **Para la próxima semana:**

- Experimentación con el dataset en inglés
- Reporte de los resultados de experimentación

---

<sup>1</sup> Enlace al notebook:

[https://github.com/Emmazch22/Paula\\_y\\_Emanuel\\_CI0134/blob/main/Notebooks/english\\_dataset\\_preparation.ipynb](https://github.com/Emmazch22/Paula_y_Emanuel_CI0134/blob/main/Notebooks/english_dataset_preparation.ipynb)