

IIC 2440 – Procesamiento de Datos Masivos Actividad Evaluada 2

Instrucciones. Esta actividad debe ser resuelta en forma individual.

El formato de entrega consta de los siguientes archivos

- Un documento PDF con la respuesta a las preguntas.
- Un archivo donde esté el código o notebook con el que sacaste la información.

Fechas. La fecha de entrega de la tarea es el 16 de Mayo, a las 20:00 hrs.

1. Evaluando la sensibilidad del tamaño del contexto en RAG

Tienes un archivo con una gran cantidad de noticias. En esta actividad debes:

- Realizar un corto análisis exploratorio de los documentos, encontrando 5 temáticas de noticias presentes en los datos.
- Para cada una de esas temáticas, elaborar una pregunta en el contexto de esa temática. Por ejemplo, si tu temática fuera la situación de generación de energía en Chile, podrías preguntar ¿que proyectos de plantas energéticas están en evaluación de impacto ambiental?
- Carga las noticias en Chroma, y ejecuta un proceso de RAG con cada una de esas 5 preguntas, y para cada pregunta, cambiando la cantidad de documentos que te devuelve el retriever, para valores $k = 1, 3, 5, 10$. Anota las respuestas.
- Para cada pregunta, y k , evalúa la calidad de la respuesta (en base a lo que tú sabes), con una nota de 1 a 7.
- Grafica el promedio de las notas obtenidas agrupadas por valor de k (son 5 preguntas, por lo que debes promediar 5 notas por cada k). ¿Qué puedes concluir al respecto?