**Informe de Técnicas Utilizadas en el Análisis de Conversaciones de WhatsApp**

**Objetivo del proyecto**

El propósito de este análisis fue aplicar técnicas básicas de **procesamiento de lenguaje natural (PLN)** a una conversación de WhatsApp para extraer información útil sobre el contenido textual de los mensajes. Se utilizó un archivo .txt exportado desde WhatsApp como fuente de datos, respetando el formato original de la aplicación.

**Técnicas aplicadas**

**1. Lectura y estructuración de datos**

Se diseñó un script que lee línea por línea el archivo .txt, extrayendo:

* Fecha del mensaje
* Hora
* Autor
* Contenido del mensaje

Esto permitió convertir el texto plano en una estructura tipo DataFrame para facilitar el análisis posterior.

**2. Limpieza de texto**

Antes de analizar el lenguaje, se aplicaron transformaciones al contenido para dejarlo en una forma adecuada:

* Conversión a minúsculas
* Eliminación de signos de puntuación y números
* Eliminación de caracteres especiales y emojis
* Eliminación de palabras vacías (stopwords en español)
* Eliminación manual de palabras irrelevantes como multimedia y omitido

Esta limpieza se hizo utilizando expresiones regulares y una lista de stopwords personalizada.

**3. Conteo de palabras más frecuentes**

Una vez limpiado el texto, se generó un conteo de las palabras más utilizadas en toda la conversación. Esto permitió detectar patrones de vocabulario general, jerga frecuente o temas recurrentes entre los participantes.

**4. Nube de palabras (Word Cloud)**

Se utilizaron visualizaciones tipo “nube de palabras” para mostrar de manera intuitiva y estética la frecuencia de aparición de los términos. Se generaron dos tipos de nubes:

* **General**, con todos los mensajes juntos.
* **Por autor**, para analizar diferencias individuales en el lenguaje utilizado.

La nube asigna mayor tamaño visual a las palabras más repetidas.

**Herramientas utilizadas**

* Python
* Pandas (manejo de datos)
* re (expresiones regulares)
* NLTK (stopwords en español)
* WordCloud y Matplotlib (visualización)

**Conclusión**

Este proyecto demostró cómo técnicas simples de PLN permiten analizar y visualizar información valiosa contenida en conversaciones cotidianas. A pesar de no incluir análisis semántico ni temporal, los resultados ofrecen una primera aproximación al estudio del lenguaje en contextos reales.