

Este trabajo permitió que se aplicara conceptos fundamentales del procesamiento de lenguaje natural (NLP) utilizando Python. A través del proceso de limpieza y transformación del texto, logró analizar un conjunto de oraciones de manera automatizada, identificando las palabras más relevantes y frecuentes. Además, representó gráficamente los resultados, lo que facilitó la interpretación de la información obtenida.

Se utilizaron distintas herramientas y bibliotecas, cada una con una función específica:

Python: fue el lenguaje de programación principal utilizado para implementar todo el proyecto.

NLTK (Natural Language Toolkit): biblioteca especializada en procesamiento de texto que permitió realizar tareas como tokenización, eliminación de palabras vacías (stopwords) y lematización.

word_tokenize: función de NLTK que divide un texto en palabras individuales, facilitando el análisis palabra por palabra.

stopwords: conjunto de palabras comunes en inglés (como “the”, “and”, “is”) que se eliminan porque no aportan significado útil al análisis.

WordNetLemmatizer: herramienta de NLTK que transforma las palabras a su forma raíz, lo que ayuda a unificar distintas formas de una misma palabra (por ejemplo, “running” pasa a “run”).

Scikit-learn (TfidfVectorizer): se utilizó para calcular la matriz TF-IDF, que indica la importancia de cada palabra dentro del conjunto de oraciones analizado.

Matplotlib: biblioteca que permitió crear gráficos (de barras, torta y líneas) para visualizar de manera clara los resultados del análisis de palabras.