A blue logo with a bird and text

Description automatically generated

**Benemérita Universidad Autónoma de Puebla**

Facultad de Ciencias de la Computación

Integrantes:

José Jesús Ramírez Cruz 202052478

Gabriel Reyes Leal 202053516

Alfonso Saldaña Campos 202056307

Periodo: Primavera 2024

Materia: Minería de Datos

Manual de Usuario

Nombre del Docente: María Beatriz Bernabé Loranca

**Manual de Usuario para el Análisis de Sentimientos**

**INTRODUCCION**

Este manual de usuario está diseñado para guiarte en el uso de un script de Python que realiza análisis de sentimientos en tweets utilizando técnicas de procesamiento de lenguaje natural (NLP) y machine learning. A través de este manual, aprenderás cómo preparar los datos, ejecutar el análisis y visualizar los resultados.

**REQUISITOS PREVIOS**

Antes de comenzar, asegúrate de tener instalados los siguientes requisitos:

* Python 3.x
* Librerías de Python: nltk, openpyxl, scikit-learn, joblib, unidecode, matplotlib, numpy, csv, re

Puedes instalar las librerías necesarias utilizando pip:

Texto

Descripción generada automáticamente

**PASO 1: PREPARACIÓN DEL ENTORNO**

Descargar Recursos de nltk

Para que el script funcione correctamente, necesitarás descargar ciertos recursos de nltk. Ejecuta el siguiente código en Python para descargarlos:

Texto

Descripción generada automáticamente

Archivos de Datos

Asegúrate de tener los archivos de Excel con los tweets que deseas analizar. El script está diseñado para leer los datos de archivos de Excel (.xlsx).

**PASO 2: ESTRUCTURA DEL SCRIPT**

El script se compone de varias funciones organizadas de manera que permiten:

1. Preprocesar el texto: Limpieza y normalización de los tweets.
2. Clasificar sentimientos: Etiquetar los tweets como positivos, negativos o neutrales.
3. Entrenar un modelo Naive Bayes: Para clasificar automáticamente los sentimientos.
4. Visualizar los resultados: Mostrar los porcentajes de sentimientos en gráficos de barras.

Variables y Diccionarios

* diccionario\_principal: Contiene las listas de palabras positivas y negativas.
* stop\_words: Lista de palabras comunes que se eliminarán durante el preprocesamiento.

Funciones Clave

1. preprocess\_text(text): Limpia y tokeniza el texto.
2. get\_sentiment\_labels(text): Determina el sentimiento basado en el diccionario.
3. read\_tweets\_from\_excel(file\_path): Lee los tweets de un archivo de Excel.
4. process\_tweets\_and\_get\_sentiments(tweets): Procesa los tweets y obtiene las etiquetas de sentimientos.
5. calculate\_sentiment\_percentages(sentiment\_labels1, sentiment\_labels2): Calcula los porcentajes de sentimientos.
6. plot\_comparison(...): Visualiza la comparación de sentimientos.
7. train\_naive\_bayes\_model(X, y): Entrena y guarda el modelo Naive Bayes.
8. clean\_text(text): Limpia y normaliza el texto eliminando caracteres no alfabéticos.
9. extract\_words\_from\_excel(input\_file): Extrae palabras de un archivo de Excel.
10. write\_top\_words\_to\_csv(words, output\_file): Escribe las palabras más comunes a un archivo CSV.

Paso 3: Ejecución del Script

Leer y Preprocesar los Tweets

Primero, lee los tweets desde los archivos de Excel:

Texto

Descripción generada automáticamente

Obtener Etiquetas de Sentimiento

Procesa los tweets y obtén las etiquetas de sentimiento:

Pantalla de computadora con letras

Descripción generada automáticamente con confianza baja

Calcular Porcentajes de Sentimientos

Calcula los porcentajes de sentimientos para ambos conjuntos de tweets:



Visualizar Resultados

Visualiza la comparación de sentimientos:

Texto

Descripción generada automáticamente

Entrenar el Modelo Naive Bayes

Entrena el modelo Naive Bayes con los datos de los tweets:

Texto

Descripción generada automáticamente con confianza media

Analizar Nuevos Tweets (Tiempo 2)

Lee los nuevos tweets y preprocesa:

Texto

Descripción generada automáticamente

Carga el modelo y el vectorizador, y realiza las predicciones:

Texto

Descripción generada automáticamente

Calcula y visualiza los sentimientos para el nuevo conjunto de datos:

Texto

Descripción generada automáticamente

**PASO 4: ANÁLISIS DE PALABRAS COMUNES**

Extrae las palabras más comunes y escribe a CSV:

Texto

Descripción generada automáticamente

Repite para el segundo conjunto de datos:

Texto

Descripción generada automáticamente