

Informe de análisis de datos de tweets

1. Definición de Constantes para Interactuar con la Base de Datos

Las constantes para la conexión de la base de datos son: el identificador de la base de datos, el usuario, la contraseña, el servidor y el puerto. Estas constantes se establecen en el código para configuraciones consistentes y centralizadas.

DB_HOST: 'localhost'

DB_PORT: '33060'

DB_NAME: 'no'

DB_USER: 'raíz'

DB_PASSWORD: 'secreto'

Estas constantes facilitan una conexión rápida y simplifican los ajustes a la configuración de almacenamiento de datos cuando cambian.

2. Bibliotecas utilizadas

Para este análisis de datos, se utilizan las siguientes bibliotecas

mysqlconnector: para conectar y manipular datos en la base de datos MySQL

coleccionContador: Para contar la frecuencia de palabras en los textos de los tweets

Para analizar expresiones regulares y optimizar la extracción y el filtrado de palabras

Para la evaluación de sentimientos, específicamente la positividad e imparcialidad del texto dado

La codificación del texto está configurada en utf8mb4, lo que garantiza una gestión precisa de los caracteres especiales en los tweets

3. Funciones para Interactuar con la Base de Datos

Se implementan varias funciones para simplificar la interacción con la base de datos

Función de enlace de base de datos: establece una conexión con MySQL a través de parámetros de configuración

La verificación del enlace activo antes de su terminación garantiza la optimización de los recursos

Ejecutar comandos SQL: ejecuta búsquedas en la base de datos y genera los resultados

Estas operaciones organizan la vinculación y la implementación de consultas de manera ordenada, facilitando la reutilización del código para diversos fines analíticos

4. Creación de una nueva tabla para análisis posteriores

En la base de datos, se establece una tabla `tweet_analysis` para retener los datos procesados. Esta tabla incluye las siguientes columnas.

`id`: Identificador único de cada entrada

`tweet_text`: Texto del tweet analizado

Puntuación de polaridad del sentimiento (un valor que oscila entre -1 y 1, donde -1 significa extremadamente negativo, 0 denota neutralidad y 1 representa extremadamente positivo)

Nivel de subjetividad: Escala de imparcialidad en el contenido (un continuo que va desde 0 para absoluto desapego hasta 1 para absoluto sesgo) En esta oración reescrita, "subjetividad" se ha reemplazado por "nivel de subjetividad" y "objetivo" por "imparcialidad" Además, "valor" se ha cambiado a "escala", "texto"

Esto facilita la conservación y referencia de los tweets, incorporando sus métricas de sentimiento y objetividad, lo que permite un escrutinio adicional sin la necesidad de un cálculo repetido de estas métricas

5. Cuestiones de reflexión y análisis

A ¿Cuales son las palabras más utilizadas?

Para identificar los términos predominantes, se ejecutó un análisis de texto para recuperar cada término distinto de los tweets. Se excluyeron las palabras comunes y de baja relevancia, como artículos y preposiciones. Los resultados muestran que las 10 palabras más recurrentes corresponden al tema principal de los tweets, con términos específicos acentuados.

B ¿Qué sentimiento es más intenso?

El análisis de sentimiento se basa en la polaridad de cada tweet. La polaridad mide la positividad o negatividad de una emoción, mientras que la intensidad está determinada por la magnitud absoluta de la polaridad. El tweet más intenso es el que muestra con fuerza una opinión o sentimiento sobre el tema, con una puntuación de 0,9 o superior. Este grado de polaridad indica una postura emocional significativa, ya sea favorable o adversa, hacia el tema de referencia.

C ¿Qué tan objetivos son los Tweets sobre el tema?

La objetividad se cuantifica a través de la función subjetiva, donde un recuento de 0 significa una declaración totalmente imparcial y un recuento de 1 significa una declaración totalmente sesgada. La porción predominante de los tweets en esta evaluación demuestra una propensión a la subjetividad, lo que indica que los puntos de vista y perspectivas individuales prevalecen en estas discusiones sobre el tema. Esto está en consonancia con la conducta predominante en las plataformas sociales, donde las personas suelen compartir puntos de vista y anécdotas personales.

D ¿Qué tan positivos se esperan que sean los comentarios mañana?

Para pronosticar la actitud de los próximos comentarios, se tuvo en cuenta el sentimiento medio de las publicaciones existentes. Si la gente sigue diciendo cosas buenas, probablemente lo volverán a hacer mañana porque lo que le gusta a la gente en las redes sociales a menudo permanece igual por un tiempo. Por el contrario, si la tendencia disminuye, se podrían esperar comentarios de tono análogo para el día siguiente.