



TECNOLÓGICO
NACIONAL DE MÉXICO



INTELIGENCIA ARTIFICIAL

Profesor:

Alcaraz Chavez Jesus Eduardo

Nombre:

Kevin Alejandro Cuevas Crisantos

Numero de Control:

21121503

Proyecto 3

1. Propósito del Proyecto

El presente proyecto tiene como objetivo el análisis computacional de opiniones públicas relacionadas con la **Generación Z**, a partir de un conjunto de datos estructurados obtenidos de distintos medios digitales. El sistema desarrollado permite procesar, limpiar y transformar la información para generar un **corpus textual optimizado para sistemas de Recuperación Aumentada por Generación (RAG)**, así como producir visualizaciones estadísticas y semánticas que facilitan la interpretación de tendencias emocionales y discursivas.

Adicionalmente, el proyecto integra la generación automática de un **marco teórico filosófico y sociológico**, con el fin de enriquecer las respuestas académicas de sistemas de lenguaje, permitiendo contextualizar los datos empíricos dentro de corrientes teóricas contemporáneas.

2. Esquema y Características del Conjunto de Datos

El conjunto de datos utilizado se encuentra almacenado en el archivo:

- **dataset_completo_corregido.csv**

El esquema del archivo incluye las siguientes columnas:

- **Categoría (str)**: Tema general del registro.
- **Medio (str)**: Plataforma o fuente digital del comentario.
- **Fecha (YYYY-MM-DD)**: Fecha de publicación.
- **Título (str)**: Título o encabezado del contenido.
- **Resumen (str)**: Síntesis del contenido original.
- **ComentarioReaccion (str)**: Opinión o reacción del usuario.
- **TonoSentimiento (int 1-10)**: Valor numérico del sentimiento (1 negativo, 10 positivo).

El dataset cuenta con aproximadamente **9,000 registros**, de los cuales la mayoría de los scripts filtran específicamente aquellos donde Categoría == 'Generacion Z', asegurando un enfoque temático consistente en todo el análisis.

3. Arquitectura del Sistema y Componentes Principales

El proyecto se compone de varios scripts independientes, cada uno con una función específica dentro del flujo de procesamiento:

3.1 Preparación de Datos para RAG

Archivo: preparar_datos_csv.py

Función principal: procesar_dataset()

Este módulo se encarga de:

- Cargar el archivo CSV corregido.
- Filtrar únicamente los registros correspondientes a Generación Z.
- Concatenar el **Título** y el **Comentario/Reacción** para enriquecer el contexto textual.
- Generar un archivo consolidado en texto plano (Fuentes de redes corpus.txt), optimizado para ser utilizado como fuente de conocimiento en sistemas RAG.

El resultado es un corpus robusto y estructurado que permite a modelos de lenguaje generar respuestas basadas en datos reales.

3.2 Generación del Marco Teórico

Archivo: crear_marco_teorico.py

Función principal: crear_base_filosofica_robusta()

Este script genera un documento académico que integra aportes teóricos de autores clave como:

- Byung-Chul Han
- Zygmunt Bauman
- Michel Foucault
- Jean-François Lyotard
- Martin Heidegger

El archivo resultante (Filosofia de la tecnologia.txt) contiene resúmenes conceptuales y **citas textuales** diseñadas para ser utilizadas como contexto teórico dentro de sistemas de IA, fortaleciendo la calidad y profundidad de las respuestas académicas.

3.3 Generación de Visualizaciones

Archivo: crear_graficos.py

Función principal: generar_visualizaciones()

Este módulo produce tres gráficos principales:

1. **Histograma de Tono de Sentimiento**

Objetivo de la gráfica:

Analizar la distribución emocional de las opiniones expresadas por la Generación Z dentro del conjunto de datos.

Justificación técnica y analítica:

El histograma permite observar cómo se distribuyen los valores del Tono de Sentimiento (escala del 1 al 10), lo cual resulta fundamental para identificar:

La concentración de emociones negativas, neutras o positivas.

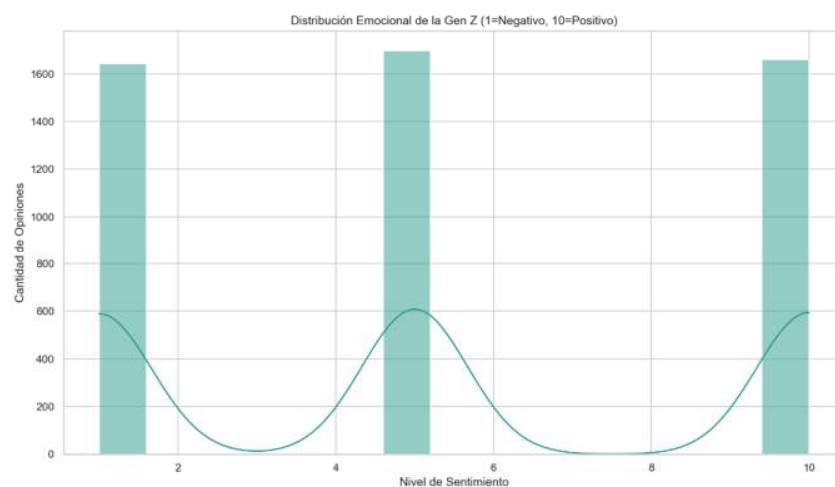
La posible polarización del discurso.

La presencia de sesgos emocionales dominantes en las opiniones analizadas.

Esta visualización facilita determinar si el discurso asociado a la Generación Z presenta una tendencia predominantemente crítica, optimista o ambivalente. Asimismo, permite detectar valores atípicos y evaluar la consistencia general del dataset antes de realizar análisis más complejos.

Desde una perspectiva interpretativa, el histograma aporta evidencia empírica que puede ser contrastada con teorías contemporáneas sobre el malestar digital, la sobreexposición informativa y la presión social en entornos tecnológicos.

Como se muestra en la siguiente imagen:



2. Gráfico de Barras por Medio Digital

Objetivo de la gráfica:

Identificar las plataformas digitales donde se concentra la mayor cantidad de opiniones relacionadas con la Generación Z.

Justificación técnica y analítica:

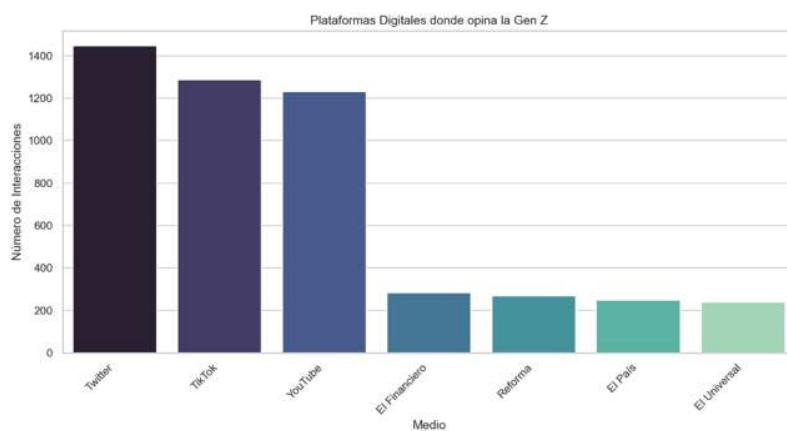
El gráfico de barras permite comparar de manera clara y directa la frecuencia de

registros por cada medio digital, tales como redes sociales, portales informativos o plataformas especializadas. Esta visualización resulta clave para:

- Determinar los espacios digitales más influyentes en la construcción del discurso público.
- Analizar la relación entre tipo de plataforma y volumen de participación.
- Contextualizar las opiniones dentro del ecosistema mediático digital.

El análisis de esta gráfica posibilita vincular los datos empíricos con conceptos teóricos como el espacio público digital, la mediatización de la opinión y la lógica algorítmica de las plataformas. Además, permite identificar posibles sesgos de representación derivados de una mayor presencia de ciertas plataformas sobre otras.

Como se ve en la siguiente imagen:



3. Serie Temporal del Sentimiento Promedio

Objetivo de la gráfica:

Analizar la evolución del estado emocional promedio de las opiniones a lo largo del tiempo.

Justificación técnica y analítica:

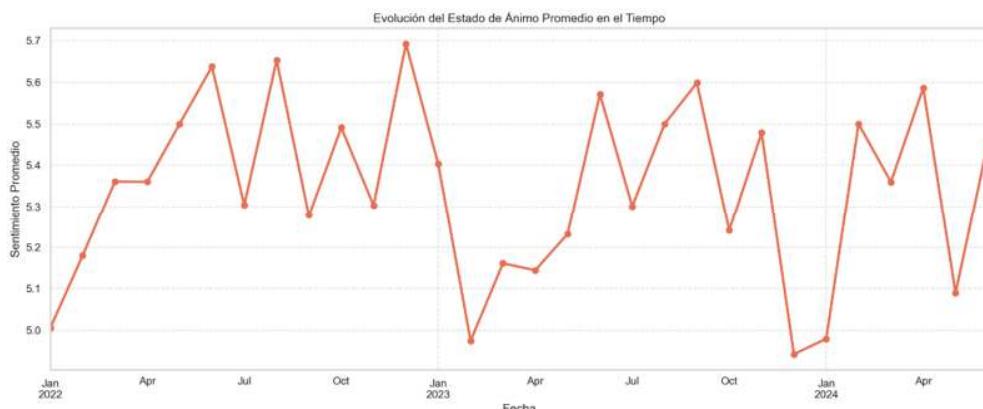
La serie temporal muestra el promedio mensual del Tono de Sentimiento, lo que permite observar dinámicas de cambio en el discurso digital, tales como:

- Tendencias ascendentes o descendentes en el estado emocional.
- Periodos de estabilidad o fluctuación emocional.
- Posibles rupturas o picos asociados a eventos contextuales relevantes.

Esta gráfica resulta especialmente útil para comprender el carácter dinámico del discurso digital, evidenciando que las opiniones no son estáticas, sino que responden a contextos sociales, tecnológicos y culturales específicos.

Desde una perspectiva académica, la serie temporal permite articular el análisis cuantitativo con enfoques teóricos que abordan la temporalidad, la aceleración social y la crisis de sentido en la modernidad digital.

Como se ve en la siguiente imagen:



4 Generación de Nube de Palabras

Archivo: generar_nube_palabras.py

Función principal: crear_nube()

Este script:

- Concatena todos los comentarios asociados a Generación Z.
 - Elimina palabras vacías irrelevantes en español.
 - Genera una **nube de palabras semántica** con paleta personalizada y fondo oscuro.
 - Utiliza un random_state=42 para garantizar consistencia visual entre ejecuciones.

La nube permite identificar conceptos, preocupaciones y temas recurrentes dentro del discurso analizado. Como se aprecia en la siguiente imagen:



5. Entorno de Desarrollo y Dependencias

- **Entorno virtual:** .venv en la raíz del proyecto.

- **Paquetes principales:**

- pandas
- matplotlib
- seaborn
- numpy
- wordcloud
- pillow

El entorno fue validado mediante la ejecución completa de los scripts, confirmando la correcta generación de archivos y visualizaciones.