

<h1>"VÓRTICE DE ODIO: ANÁLISIS DE COMENTARIOS EN YOUTUBE"</h1>		
Nombre del proyecto:	Irunela	
Director del proyecto:	Iruma A.Bragado	

Consideraciones Iniciales	<ul style="list-style-type: none">Las relaciones sociales experimentaron cambios significativos en plataformas virtuales en la última década. El anonimato y la interacción a través de pantallas favorecieron la expresión de emociones intensas. Este proyecto se centró en interacciones de contenido negativo, especialmente en la plataforma de YouTube en español.
Pregunta de Investigación:	<ul style="list-style-type: none">¿Generó la política un gran volumen de comentarios de odio?
Hipótesis Principal: Hipótesis Alternativa:	<ul style="list-style-type: none">Existió una relación significativa entre la cantidad de comentarios de odio y una línea política específica.Los comentarios o publicaciones negativas generaron más interacción que los positivos.
Objetivo General:	<ul style="list-style-type: none">Identificar y clasificar el contenido de los comentarios de odio.

ASPECTOS ESPECÍFICOS DEL PROYECTO

- Obtener publicaciones y comentarios asociados a diferentes tópicos.
- Explorar el contenido de los comentarios.
- Identificar la frecuencia e intensidad de odio en los comentarios.
- Relacionar comentarios con usuarios.
- Predecir la ocurrencia por categoría de comentarios negativos.
- Comparar la prevalencia y naturaleza de comentarios de odio en diferentes plataformas sociales respecto a la política.

METODOLOGÍA

Mediante el Procesamiento de Lenguaje Natural (NLP) con la librería Pysentimiento, analicé la valoración del contenido de las publicaciones, devolviendo una clasificación de negativo, neutro o positivo. Me centré en comentarios negativos para analizar la intensidad de odio con la librería Speech Haters, que proporciona un puntaje de hateful y aggressive. Con los resultados, generé una función de categorización para identificar la temática del mensaje, basada en la frecuencia de comentarios relacionados con cada categoría (ideología política, racismo y machismo). Exploré diferentes modelos de Machine Learning para análisis multicategoría, destacando el modelo de Support Vector Machine (SVC) para categorizar comentarios y BERT para la clasificación de comentarios.

RECOLECCIÓN Y ANÁLISIS DE DATOS

A través de API	extraje información de páginas de YouTube, basándome en tópicos políticos predefinidos como inmigración, economía, educación, salud, vivienda, género, violencia de género, territorio, amnistía, feminismo y monarquía. Utilicé librerías como Pandas, Numpy, Matplotlib y Seaborn para análisis estadísticos y técnicas de procesamiento de datos, normalización y tokenización de comentarios.
-----------------	---

METRICAS Y CONCLUSIONES

Método conservador	Considerando la investigación, aproximadamente el 60% de los 110,000 comentarios totales fueron negativos, y dentro de este grupo, el 48% exhibió mayores niveles de hateful y aggressive. Respecto a las hipótesis planteadas, no se pudo afirmar que hubiera más interacciones según el tema del video ni identificar fácilmente la afiliación política de los usuarios que emitieron estos comentarios. Sin embargo, la categoría ideología política mostró la mayor frecuencia y valores más altos de hateful y aggressive. En cuanto al objetivo general del proyecto (identificar y clasificar el contenido de los comentarios de odio), el enfoque más conservador con el modelo SVC arrojó una precisión del 0.67, con valores de precisión alrededor de 0.90 y un valor de recall superior a 0.85 (Tabla 1).
--------------------	---

Tabla 1: predicciones en el conjunto de prueba SVC

	precisión	recall	f1-score	support
ideología política	0.88	0.89	0.89	65
machismo	0.90	0.97	0.93	71
racismo	0.95	0.85	0.90	66
accuracy			0.91	202
macro avg	0.91	0.90	0.90	202
weighted avg	0.91	0.91	0.91	202

bert

Lo anterior indica que si bien el valor de predicción no es elevado, al considerar los valores de *presicion* y *recall* demuestran que el modelo se ajusta bien a las categorías.

Partiendo de un enfoque menos conservador e implementado el modelo pre entrenado de BERT para predecir las mismas categorías. Por razones de coste computacional hemos definido *epochs* = 3, y *batch_size*=32, lo que nos devolvió valores de precisión de 0.63 (accuracy), *precision* en torno a los 0.60 y *recall* cercanos a .70 (tabla 2).

En observación con estos datos, considero que para esta problemática es mejor utilizar un enfoque convencional por su beneficio en tiempo y coste computacional.

	precisión	recall	f1-score	support
ideología política	0.65	0.42	0.51	74
machismo	0.59	0.74	0.66	57
racismo	0.66	0.77	0.71	71
accuracy			0.63	202
macro avg	0.63	0.64	0.63	202
weighted avg	0.64	0.63	0.62	202

oportunidades de mejora

Con todo lo anterior, puedo pensar que a la hora de analizar el impacto de los comentarios negativos, agresivos y de odio en las redes sociales, identificar y/o etiquetar el contenido del comentario favorece el entendimiento del mismo. A su vez, puede pensarse como una herramienta de higiene en el tipo de consumo de esas redes sociales, si se advierte al usuario el posible contenido de esos comentarios.

Como posibilidades de mejora entiendo que mejorar la calidad del dataset para el entrenamiento de los modelos permitirá ajustar los parámetros y a su vez las métricas. Otras opciones de mejora en torno a los datos sería Investigar si las medidas tomadas por las plataformas (eliminación de contenido, suspensión de cuentas, etc.) tienen un impacto real en la reducción del odio; geolocalización de los comentarios; identificar usuarios con mayor actividad y comparar sus interacciones con diferentes tópicos; comparar temáticas similares en diferentes plataformas virtuales.