

ANÁLISIS DE SENTIMIENTO DE LOS DATOS DE TWITTER DE COVID-19 UTILIZANDO MODELOS DE APRENDIZAJE PROFUNDO Y APRENDIZAJE DE MÁQUINA

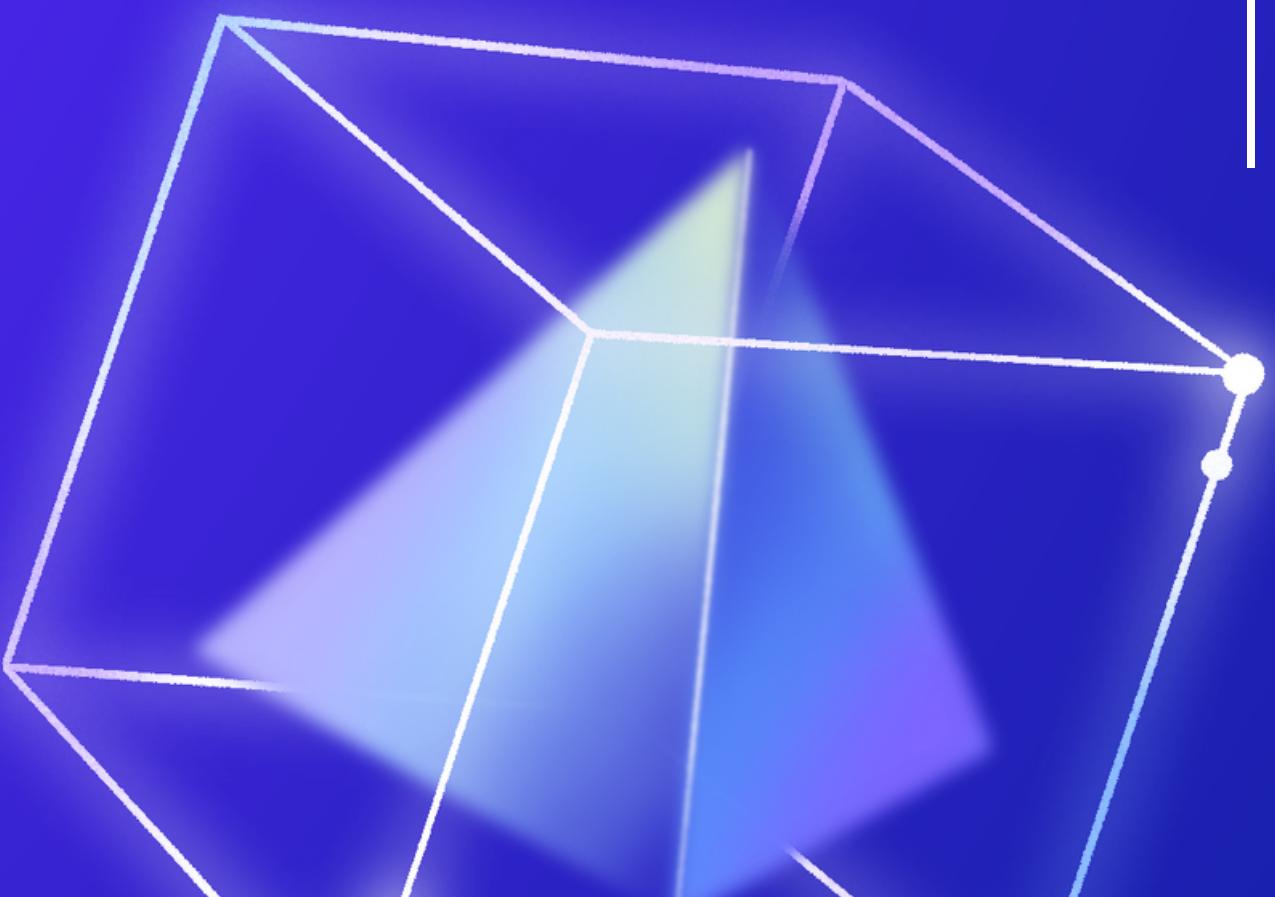
Simran Daradl, Sridhar Krishnan





TABLA DE CONTENIDO

- Resumen
- Introducción
- Materiales y métodos
- Resultados y discusión
- Conclusiones





RESUMEN

El análisis del sentimiento se llevó a cabo mediante un modelo de aprendizaje profundo conocido como (BERT) y varios modelos de aprendizaje automático para el análisis de texto y el rendimiento, que luego se compararon entre sí. Los modelos ML utilizados son Bayes ingenuas, regresión logística, bosque aleatorio y máquinas vectoriales de soporte.



INTRODUCCIÓN

Esta investigación se enfocó en un conjunto de datos que contiene tweets de Twitter y tweets a los que se accedió relacionados con la “Pandemia de COVID-19”.

Por lo tanto, hay dos formas de etiquetar los tweets extraídos utilizando la API de Twitter con tweepy.

- La primera forma es entrenar un modelo BERT y varios modelos de aprendizaje automático con datos ya etiquetados, evaluando qué modelo clasificador podría etiquetar correctamente los tweets, y luego usar ese modelo para etiquetar el texto de los tweets extraídos.
- La segunda forma de encontrar el sentimiento es usar VADER, una librería predesarrollada en código abierto para análisis de sentimiento.



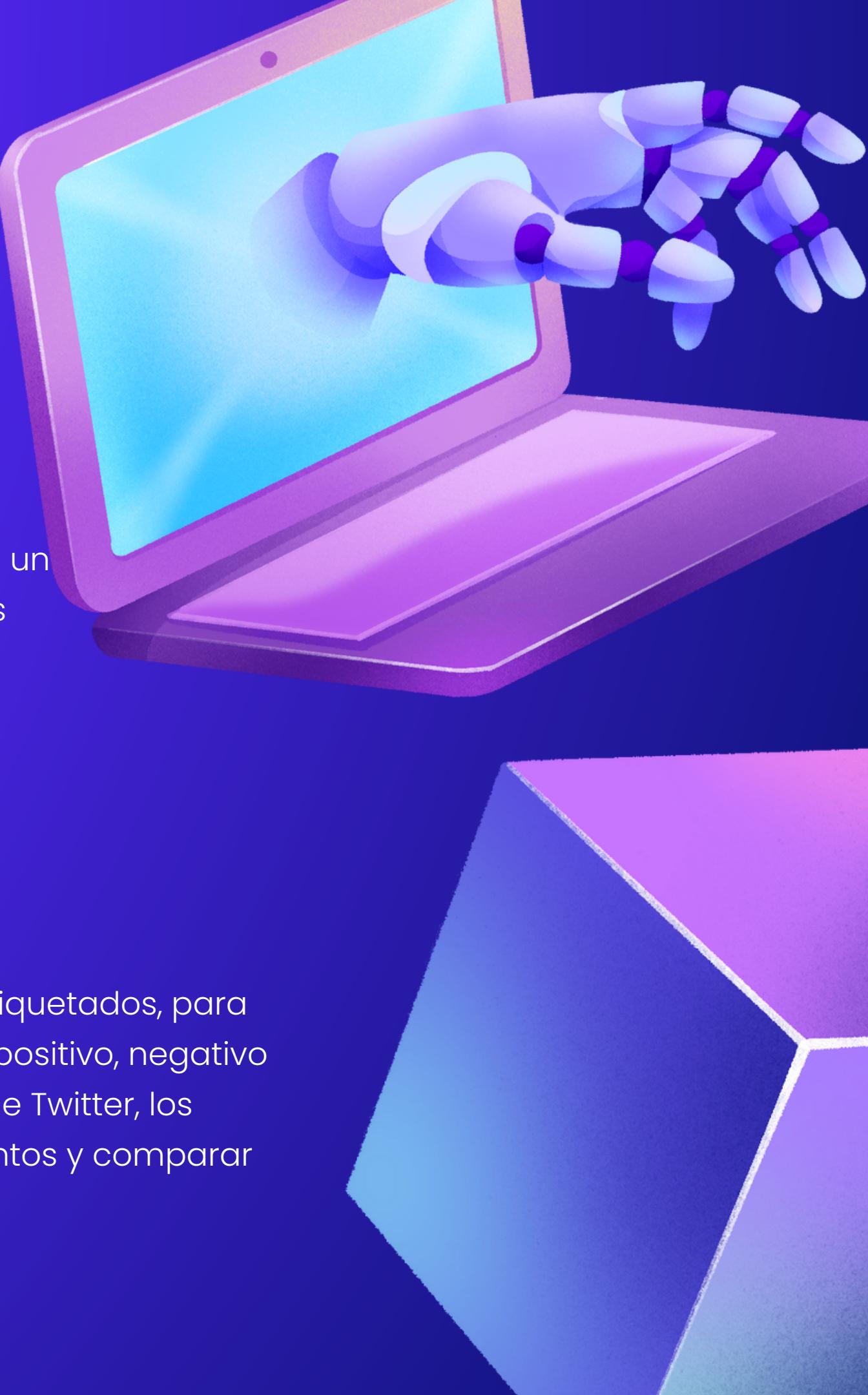
MATERIALES Y MÉTODOS

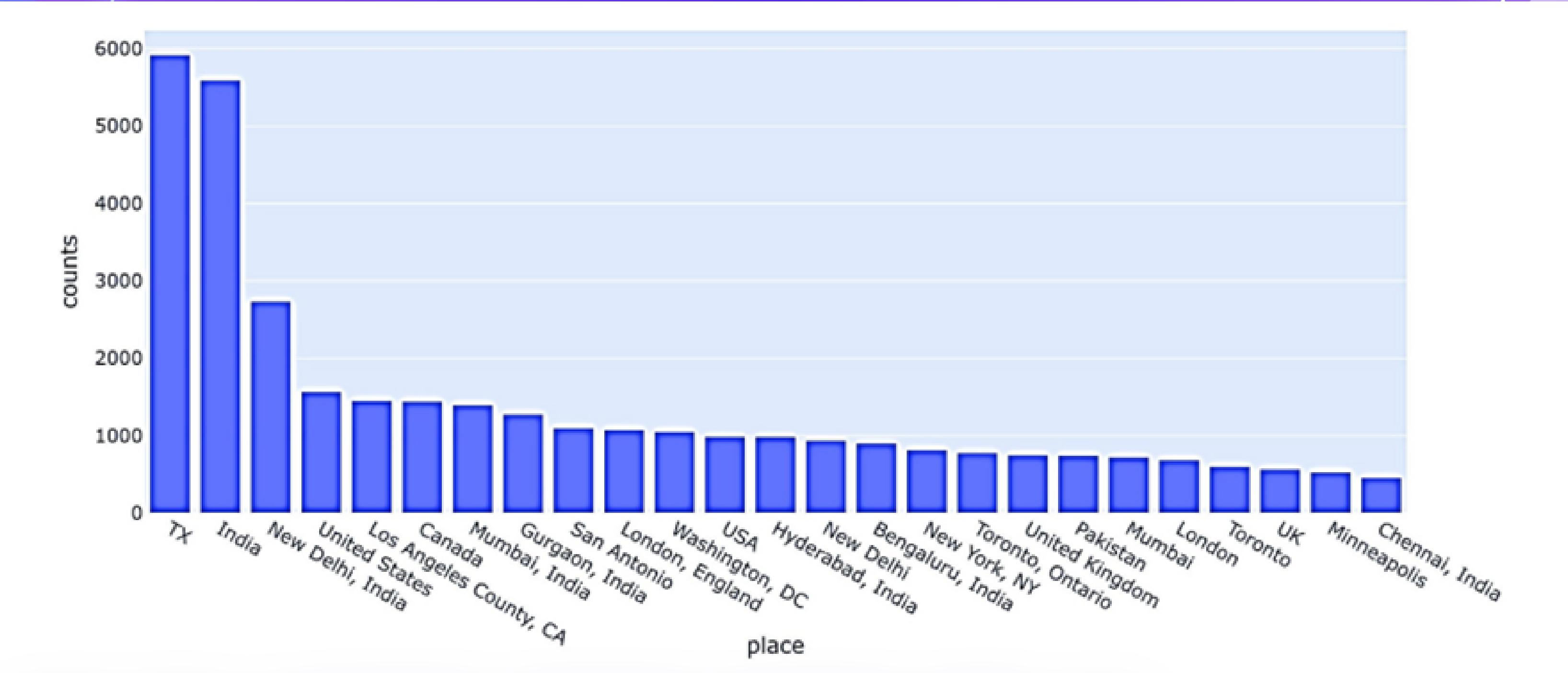
Materiales

La data para este trabajo fue adquirida de Twitter, utilizando su API tweepy. Tweepy es un paquete de Python de código abierto de fácil uso, para acceder a las funcionalidades proporcionadas por la API de Twitter.

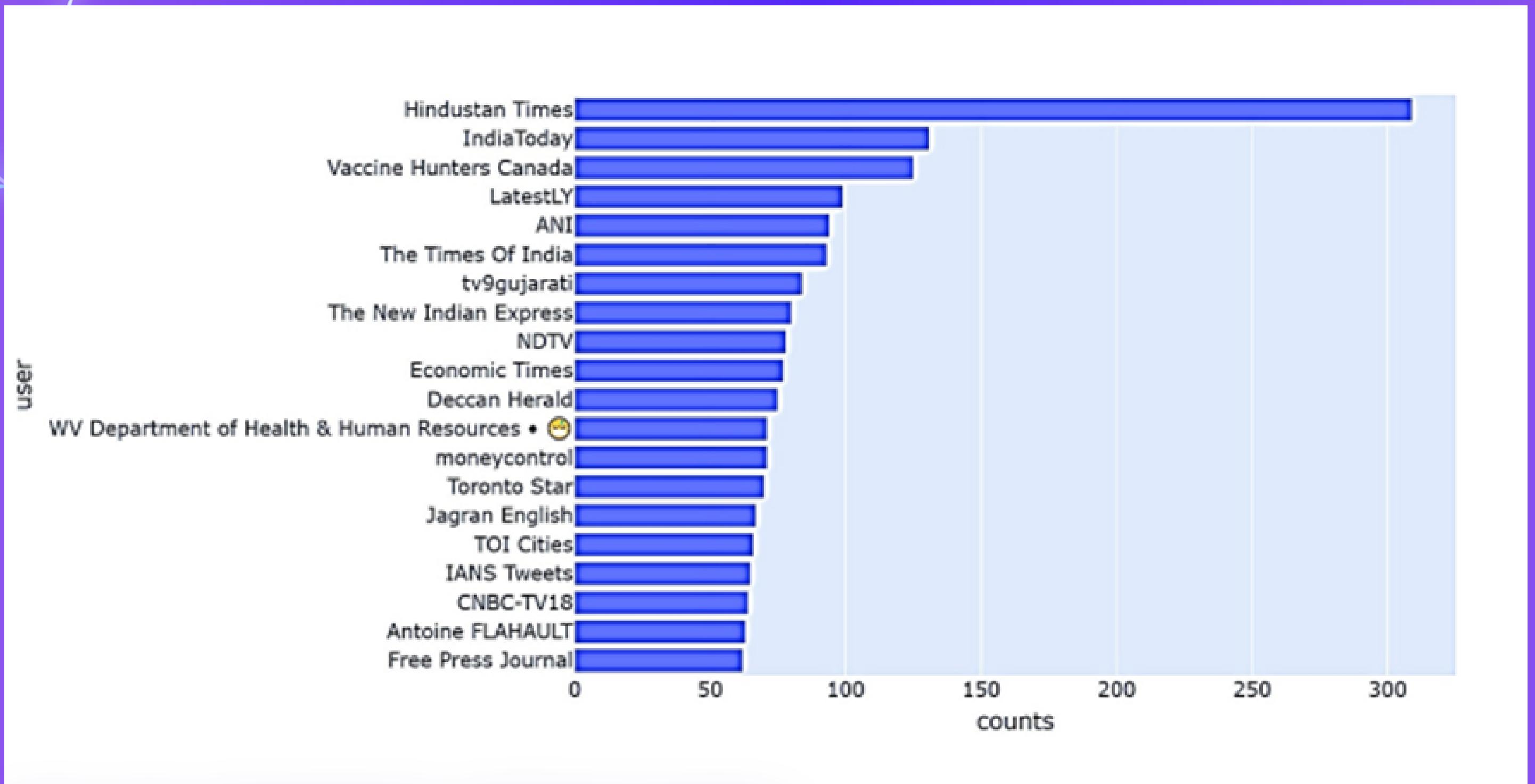
Métodos

El objetivo de este estudio es entrenar utilizando texto de los tweets etiquetados, para evaluar automáticamente si el tweet no etiquetado recolectado si es positivo, negativo o neutral. Después de entrenar los modelos sobre datos etiquetados de Twitter, los modelos fueron aplicados a data extraída para etiquetar los sentimientos y comparar los resultados de los diferentes algoritmos.





LOS 25 LUGARES DESDE DONDE MÁS SE
ORIGINARON LOS TWEETS



LOS 20 USUARIOS VERIFICADOS
QUE MÁS TUITEARON

RESULTADOS Y DISCUSIÓN



RESULTADOS

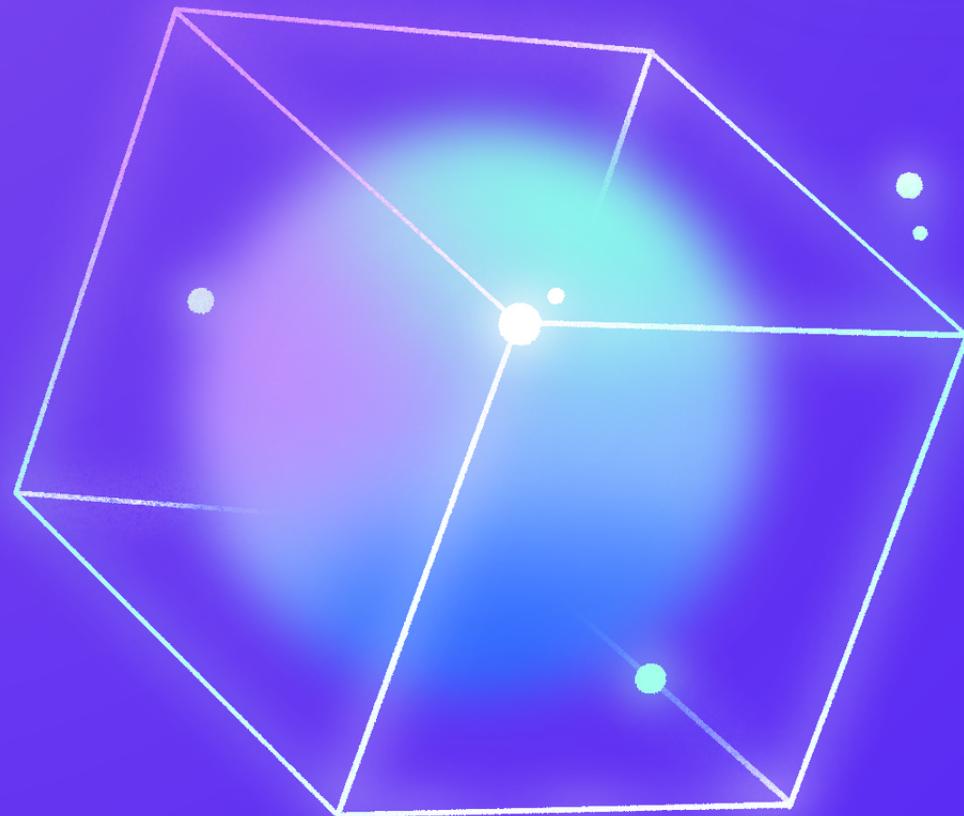
Se aplicó la clasificación multiclase de los diferentes modelos a la data de entrenamiento para encontrar la exactitud de la etiqueta correcta en el conjunto de prueba.

DISCUSIÓN

Este estudio puede ser utilizado para analizar los sentimientos cambiantes de la gente alrededor del mundo, y verificar si hay variaciones importantes en ellos en el período de tiempo junto con el aumento en el suministro de vacunas.

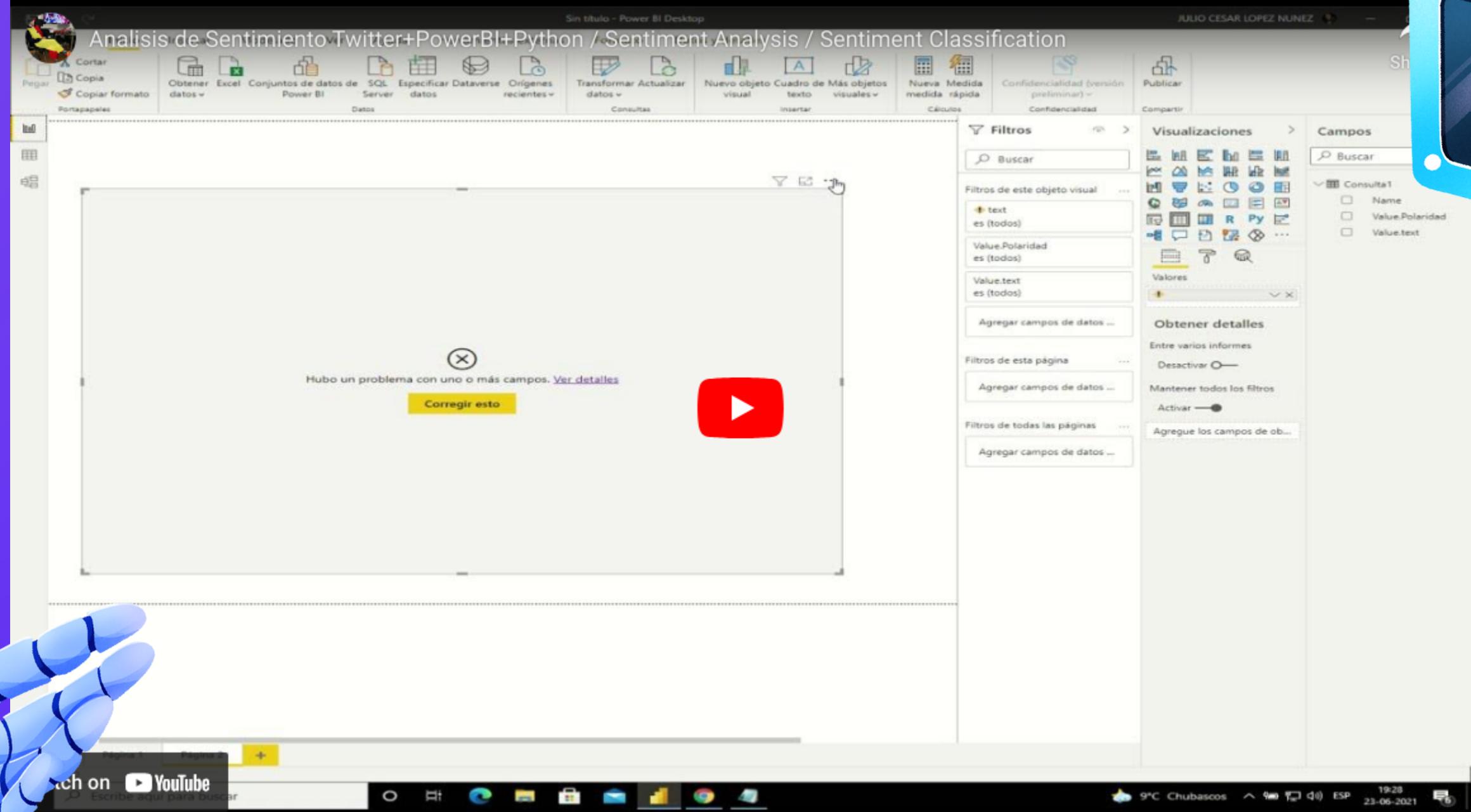


CONCLUSIÓN



En la predicción de texto en cualquier red social es casi imposible alcanzar una exactitud perfecta. A través de esto, se puede aprender la cuestión principal para ayudar a los proveedores de salud a identificar algún tipo de enfermedad mental antes que sea demasiado tarde.





GRACIAS



AUTOMATION PROCESS



se realizó minería de datos en Twitter para recolectar un total de 107.990 tweets relacionados con COVID-19, entre el 13 de diciembre de 2019 y el 9 de marzo de 2020.



Se utilizó un enfoque de procesamiento de Lenguaje Natural (NLP, Natural Language Processing en inglés) y el algoritmo de asignación latente de Dirichlet para identificar los tópicos más comunes de los tweets, así como también categorizar clústeres e identificar temas en base



Lorem ipsum dolor sit amet, consectetur adipiscing elit. Proin varius, eros nec efficitur euismod, lectus turpis sollicitudin augue, non ultricies enim nunc sit amet augue.



Lorem ipsum dolor sit amet, consectetur adipiscing elit. Proin varius, eros nec efficitur euismod, lectus turpis sollicitudin augue, non ultricies enim nunc sit amet augue.