

Análisis y Clasificación de Sentimientos de Tweets sobre el COVID-19 y la Aplicación de la Vacuna COVID-19

Cynthia Martínez Espinoza
Matrícula: 1011238
cynthia.martineze@uanl.edu.mx

23 de enero de 2025

Introducción

La pandemia de COVID-19 ha generado una de las crisis sanitarias más significativas en la historia reciente, afectando a millones de personas en todo el mundo. En este contexto, las vacunas se han posicionado como una herramienta fundamental para mitigar el impacto del virus, reducir la gravedad de la enfermedad y limitar su propagación. La vacuna contra el COVID-19 ha transformado nuestras vidas.

Desde su invención, las vacunas han sido una herramienta clave en la salud pública. La historia nos muestra cómo la inmunización ha erradicado enfermedades y salvado millones de vidas. Reflexionemos sobre su impacto a lo largo de los años.

Las vacunas funcionan mediante la estimulación del sistema inmunológico. Al introducir un patógeno inactivado o atenuado, el cuerpo produce anticuerpos. Esta respuesta prepara al organismo para combatir futuras infecciones.

La inmunización no solo protege al individuo, sino que también contribuye a la inmunidad colectiva. Al vacunar a la población, se reduce la propagación del virus, protegiendo a los más vulnerables y evitando brotes.

La inmunización no solo protege al individuo, sino que también contribuye a la inmunidad colectiva. Al vacunar a la población, se reduce la propagación del virus, protegiendo a los más vulnerables y evitando brotes.

A pesar de los avances, existen desafíos en la distribución y aceptación de las vacunas. La desinformación y el acceso desigual son obstáculos que debemos enfrentar para lograr una inmunización efectiva. Sin embargo, cada individuo es responsable de decidir, si o no aplicarse la vacuna.

Planteamiento del Problema

Este informe documenta el análisis y clasificación de sentimientos aplicado a un conjunto de 6000 tweets relacionados con el COVID-19 y las vacunas. El análisis incluye tareas de procesamiento de texto, visualización de resultados y evaluación del desempeño de un modelo de clasificación basado en algoritmos de aprendizaje automático.

Metodología

Recopilación de Datos

Los tweets fueron obtenidos de un conjunto de datos que incluía menciones a diversas vacunas, como Pfizer/BioNTech, Sinopharm, Sinovac, Moderna, Oxford/AstraZeneca, Covaxin y Sputnik V.

Identificar y clasificar los sentimientos de los tweets, comprender las tendencias de opinión y evaluar el desempeño del modelo Random Forest.

Preparación de los Datos

- **Descarte de Palabras:** Se eliminaron palabras comunes irrelevantes utilizando listas de stopwords en inglés.
- **Tokenización:** Los tweets fueron divididos en palabras individuales.
- **Stemming:** Se redujeron las palabras a su forma base para unificar variantes.
- **Vectorización:** Se generó una representación numérica de los textos, incluyendo unigramas y bigramas.

Análisis de Sentimientos

- **Polaridad:** Indica el sentimiento general del texto, variando de -1 (negativo) a 1 (positivo).

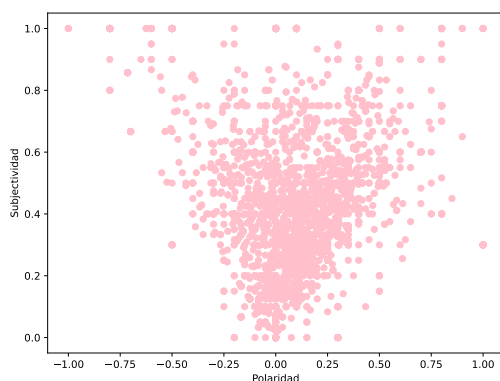


Figura 3: Dispersion Polaridad Subjetividad

Distribución de Sentimientos

- **Positivos:** 2235 tweets (37.25 %).
 - **Negativos:** 822 tweets (13.7 %).
 - **Neutrales:** 2943 tweets (49.05 %).
- Figura 2: Distribución de sentimientos.

Relación entre polaridad (eje X) y subjetividad (eje Y).

Polaridad (¿qué tan positivo o negativo es el sentimiento?).

Subjetividad (¿es una opinión o un hecho?).

Cada punto en el gráfico representa un texto analizado, y su posición depende de los valores de polaridad y subjetividad asociados.

Mide el sentimiento: Eje X (Polaridad) Izquierda (-1): Negativo. Centro (0): Neutro. Derecha (1): Positivo.

Mide la subjetividad: Eje Y (Subjetividad): Inferior (0): Objetivo (hechos). Superior (1): Subjetivo (opiniones).

La mayoría de los textos analizados son neutrales en términos de polaridad, pero con un grado moderado de subjetividad.

Existe una cantidad considerable de textos positivos y negativos, pero su proporción es menor en comparación con los neutros. La distribución de subjetividad sugiere que los textos contienen tanto opiniones como referencias a hechos, aunque las opiniones predominan. **Figura 3: Dispersion Polaridad Subjetividad**

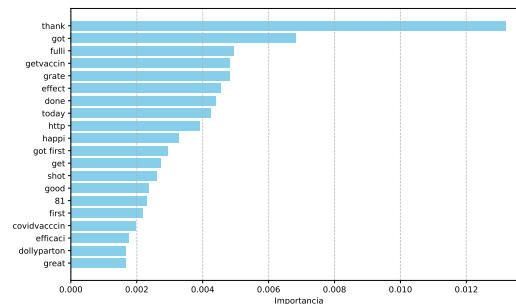


Figura 4: Enter Caption

Importancia de las Características

El modelo Random Forest permitió identificar los términos más relevantes en la clasificación, como *side effect*, *thank*, y *vaccination*. Esto facilita entender cómo las palabras clave influyen en las decisiones del modelo.

Este gráfico proporciona una interpretación clara de cómo el modelo clasifica los textos basándose en las características más representativas. Ayuda a comprender qué palabras o frases influyen más en las decisiones del modelo. Los términos destacados pueden utilizarse para identificar tendencias en los datos, mejorar el preprocesamiento del texto (por ejemplo, considerando sinónimos o términos específicos) y ajustar futuros modelos.

El modelo de Random Forest no solo realiza clasificaciones precisas, sino que también ofrece una forma interpretable de entender los datos a través de la importancia de las características.

Figura 4: Características Relevantes

Evaluación del Modelo

- **Precisión en Entrenamiento:** 99 %.
- **Precisión en Prueba:** 76 %.

Conclusiones

- La conversación sobre las vacunas se centra en aspectos logísticos, como dosis, disponibilidad y aprobación.
- India destaca como un foco importante debido a su rol en la producción de vacunas.
- Los tweets reflejan una mezcla de gratitud y preocupaciones sobre efectos secundarios.
- El modelo Random Forest es efectivo para identificar patrones en los datos y ofrece una

interpretación clara de los términos más influyentes.