

Análisis de sentimientos y tópicos sobre las vacunas cubanas Soberana 02 y Abdala en Twitter

Sentiment and topics analysis about the cuban vaccines Soberana 02 and Abdala on Twitter

Antonio Rodríguez Sánchez^{1*} <https://orcid.org/0000-0001-7742-8871>

Luis Manuel Rivero González² <https://orcid.org/0000-0001-9544-551X>

Sandra Soca Lozano³ <https://orcid.org/0000-0003-0198-9298>

¹Escuela Nacional de Salud Pública, Departamento de Bioestadística. La Habana, Cuba

²Universidad de La Habana, Facultad de Derecho. La Habana, Cuba.

³Universidad del Sur de la Florida, Escuela de Psicología. Florida, Estados Unidos.

*Autor para la correspondencia: tonypsico14@gmail.com

RESUMEN

El análisis de sentimientos o minería de opiniones es una rama de la computación que permite analizar opiniones, sentimientos y emociones en ciertas áreas de interés social como productos, servicios, organizaciones, compañías, eventos y temas de interés actual. En tal sentido se propuso identificar los sentimientos y tópicos presentes en los *tweets* que hicieron mención a las vacunas cubanas Soberana 02 y Abdala en la red social Twitter. Se optó por los lenguajes de programación *Python* y *R* con sus librerías específicas para la ciencia de datos. La primera parte del estudio, que abarcó desde el *web scraping* hasta la cuantificación de las palabras más usadas, se realizó con *Python* y las siguientes librerías: *tweepy*, *pandas*, *re*, *nltk* y *matplotlib*. Mientras que la segunda, que fue la del análisis de sentimientos y

detección de tópicos, se implementó con *R* y se utilizó: *tokenizers*, *tm*, *syuzhet*, *topic modeling*, *tidyverse*, *barplot* y *wordcloud*. Se obtuvo que entre los términos con que más se dialoga en Twitter están dosis, vacunas, eficacia, cubanos, candidatos, millones, país, personas, recibido y población. En los *tweets* las emociones predominantes fueron el miedo y, ligeramente por encima, la confianza; en la polaridad predominó la positiva, como expresión del contexto vivido en el cual se desarrolló la campaña de vacunación. A partir de los tópicos identificados y los términos que se relacionaron con las emociones predominantes, así como por la polaridad, se aprecia consenso en torno a las vacunas Soberana 02 y Abdala.

Palabras clave: análisis de sentimientos; tópicos; procesamiento de lenguaje natural; Twitter; vacunas.

ABSTRACT

Sentiment analysis or opinion mining is a branch of computing that allows analyzing opinions, feelings and emotions in certain areas of social interest such as products, services, organizations, companies, events and topics of current interest. In this sense, the objective of this paper was to identify the feelings and topics present in the tweets mentioning the Cuban vaccines Soberana 02 and Abdala on Twitter social network. The programming languages Python and R with their specific libraries for data science were chosen. The first part of the study, which ranged from web scraping to the quantification of the most used words, was carried out with Python and the libraries *tweepy*, *pandas*, *re*, *nltk* and *matplotlib*. While the second, which was the sentiment analysis and topic detection, was implemented with R and used *tokenizers*, *tm*, *syuzhet*, *topic modeling*, *tidyverse*, *barplot*, and *wordcloud*. It was obtained that among the terms with which there is more dialogue on Twitter are doses, vaccines, efficacy, Cubans, candidates, millions, country, people, received and population. In the tweets, the predominant emotions were fear and confidence, slightly above it; in the polarity, the positive one predominated, as an expression of the lived context in which the vaccination campaign was developed. A consensus can be perceived around the vaccines Soberana 02 and Abdala, from the identified topics and the terms that were related to the predominant emotions, as well as the polarity.

Keywords: sentiment analysis; topics; natural language processing; Twitter; vaccines.

Recibido: 07/09/2022

Aceptado: 13/11/2022

Introducción

El análisis de sentimientos o minería de opiniones es una rama de la computación que permite analizar opiniones, sentimientos y emociones en ciertas áreas de interés social como productos, servicios, organizaciones, compañías, eventos y temas de interés actual. Se realiza mediante la aplicación del Procesamiento de Lenguaje Natural (PNL, por sus siglas en inglés), Aprendizaje Automático e Inteligencia Artificial (IA), dado que dichas opiniones pueden ser expresadas mediante *tweets*, clasificadas en emociones o polaridades.⁽¹⁾

Relacionados específicamente con la salud, se han realizados estudios sobre el impacto de la pandemia COVID-19,⁽²⁾ su efecto en la salud mental,⁽³⁾ acerca del aislamiento social,^(4,5,6,7,8) otros que además incorporaron la detección de temas de conversación^(9,10,11,12,13,14) y en relación con el uso del dióxido de cloro como tratamiento de la enfermedad.⁽¹⁵⁾ También se identificó una investigación sobre la farmacoterapia.⁽¹⁶⁾ En Cuba solo se halló un trabajo sobre el diálogo en Twitter acerca de las brigadas médicas cubanas en el contexto de la pandemia, pero en el año 2021.⁽¹⁷⁾

Entonces, el evento de interés resultó ser la campaña de vacunación masiva vivida en Cuba como parte del enfrentamiento a la enfermedad de la COVID-19, la cual comenzó como intervención sanitaria en mayo del 2021 para grupos y territorios de riesgo y en junio se inició su uso de emergencia para la totalidad de la población cuando el país atravesaba por un empeoramiento de la situación epidemiológica, debido al aumento sin precedentes de las tasas de incidencia y mortalidad y el riesgo de muerte, para abril de esa fecha.⁽¹⁸⁾ Los candidatos vacunales más avanzados en aquel momento, de acuerdo con el cumplimiento de las diferentes etapas del ensayo clínico, eran Abdala y Soberana 02. Según el cronograma divulgado habían concluido la última fase en mayo y contaban con la aprobación del Centro para el Control Estatal de Medicamentos, Equipos y Dispositivos Médicos (CECMED),⁽¹⁸⁾ que es la entidad rectora y reguladora de este proceso.

Por tanto, debido al impacto mediático que tal suceso generó, se indagó en las redes sociales, específicamente en Twitter, con el objetivo de identificar los sentimientos y los tópicos presentes en los *tweets* que hicieron mención a las vacunas cubanas Soberana 02 y Abdala. Teniendo en cuenta que esta red social es la tercera más usada en Cuba⁽¹⁹⁾ y, además, es de las más flexibles para el acceso a la información de sus usuarios con fines investigativos. Los estudios en las redes sociales favorecen el acceso a la producción ideográfica de los sujetos en condiciones de total espontaneidad, inmediatez y simultaneidad por lo que es factible de convertirse en un medidor con alta sensibilidad para indagar en cómo reaccionan las personas a los acontecimientos sociales.

Métodos

Se realizó un estudio de análisis de sentimientos y tópicos sobre los *tweets* que hicieron mención a las vacunas cubanas Soberana 02 y Abdala en el período del 11 de julio al 21 de septiembre del 2021. Para la realización de la investigación se optó por los lenguajes de programación *Python* y *R* con sus librerías específicas para la ciencia de datos. Para *Python* se escogió como entorno de desarrollo integrado el *Visual Studio Code* y, en el caso de *R*, el *RStudio*. La primera parte del estudio, que abarcó desde el *web scraping* hasta la cuantificación de las palabras más usadas, se realizó con *Python*; mientras que la segunda, el análisis de sentimientos y detección de tópicos, se implementó con *R*.

El primer paso y condición necesaria para iniciar la investigación es ser usuario de *Twitter*. Esto permite la habilitación de una cuenta de desarrollador⁽²⁰⁾ y el ulterior registro de una aplicación a la que se nombró “Proyecto análisis” para obtener el acceso a su API (Interfaces de Programación de Aplicaciones), el cual proporciona unas credenciales de acceso y uso de uno sus servicios como es la búsqueda de contenido. Teniendo en cuenta el tipo de API de las que provee la red social, se usó el API *rest* para acceder a los *tweets* que se han generado desde el pasado reciente hasta el presente: los últimos siete días, de acuerdo con sus restricciones.

A través de *tweepy*⁽²¹⁾ se inició con la conexión y autenticación a la API *rest* de *Twitter* para dar lugar a la extracción de los *tweets*. Se estableció como *query* o término de búsqueda los

hashtags #Soberana02 y #Abdala. La recolección de *tweets* ocurrió entre el 11 de julio y el 21 de septiembre. En este período se realizaron capturas dentro de un rango máximo de 15 minutos, limitación impuesta por *Twitter*. Los *tweets* recolectados fueron cargados en un archivo *json* y se organizaron en un *dataframe*, a través de *pandas*.⁽²²⁾ La estructura de metadatos quedó formada por un *dataset*, cuyos campos fueron el id del *tweet*, autor, fecha/hora y texto del *tweet*.

Para iniciar la etapa de minería de texto con la aplicación de las herramientas del NLP se creó un nuevo *dataframe* en formato *csv* a partir del filtro aplicado al campo “texto del *tweet*”, ya que en lo adelante sería la unidad de análisis. Posteriormente, se realizó el preprocesamiento que no es más que la limpieza de datos. En las tareas 1 y 2 se empleó la librería *re*⁽²³⁾ y en la 3 y 4 con *nlTK*,⁽²⁴⁾ esta última es imprescindible para pasar de un objeto de estructura oracional a uno de *bag of words* o bolsa de palabras. Una vez concluida la normalización del texto, se calculó la frecuencia de las palabras más usadas y se graficó con *matplotlib*.⁽²⁵⁾

1. Eliminación de saltos de líneas, signos de puntuación, hashtags, menciones, hipertextos, emoticones, palabras incompletas y caracteres sueltos.
2. Conversión de todas las palabras a minúsculas.
3. Eliminación de *stopwords* (palabras enlace) y palabras derivadas.
4. Tokenización: división del texto en palabras.

Para el análisis de sentimientos se implementó el siguiente procedimiento de acuerdo con las librerías elegidas. Con *tokenizers*⁽²⁶⁾ se tokenizaron nuevamente las palabras para permitir la vectorización. En la identificación de sentimientos se usó la librería *syuzhet*,^(27,28) que opera con un diccionario de términos en español como parte del *NRC Word-Emotion Association Lexicon*, compuesto por una lista de palabras y sus asociaciones con ocho emociones (ira, miedo, anticipación, confianza, sorpresa, tristeza, alegría y disgusto) y dos sentimientos (negativo y positivo).⁽²⁷⁾ Esto genera una matriz que está integrada por un vector de palabras, las ocho emociones y la polaridad negativa y positiva, donde 0 indica que la palabra del listado no existe en el diccionario y ≥ 1 su correspondencia-valencia con las emociones; a esto se le calculó el porcentaje.

También se abordó la polaridad a nivel longitudinal para identificar su evolución a lo largo de todo el conjunto de términos; según sea negativa se le asigna - 1 y si es positiva 1. Se emplean tres procedimientos estadísticos para su realización: la media móvil, la regresión ponderada localmente y la transformación de coseno discreta.⁽²⁷⁾ En uno y otro caso el ploteo se hizo con *barplot*⁽²⁹⁾ y *simple_plot*,⁽²⁷⁾ respectivamente. Para mostrar la distribución de los términos por tipo de emoción y polaridad se optó por hacerlo a través de una gráfica de nubes de palabras con *wordcloud*.⁽³⁰⁾ Esto requirió de *tm*⁽³¹⁾ para para la conversión del *dataset tokenizado* en un vector de palabras, posteriormente, a un corpus y de este a una matriz de término-documento.

El modelado de tópicos para la identificación de temas en los *tweets* se realizó con *topicmodels*, que es un método de aprendizaje no supervisado que opera con el algoritmo matemático LDA (Asignación Latente de Dirichlet).⁽³²⁾ Este se fundamenta, como parte de las características básicas de los algoritmos sobre la detección de temas, en que cada documento se compone de varios temas o *topics* y que cada tema supone un conjunto de palabras que lo representa,⁽³²⁾ a partir de la coocurrencia de las palabras. De esta manera se puede medir la distancia semántica entre ellas en una estructura de *bag of words*, por lo que se prioriza el aspecto conceptual del lenguaje en detrimento del sintáctico.

Por tanto, su agrupamiento está delimitado por la presencia de tópicos que esperan sean descubiertos, porque son la estructura del discurso. Se sigue una lógica inversa en la generación de temas, al plantear que no son las palabras las que los determinan sino al revés y, en consecuencia, cada palabra tiene distinto peso en dichos tópicos; de lo que se deriva que sean más relevantes en uno que en otros. Cada palabra es el resultado de un encadenamiento de distribuciones y luego se realiza la inferencia hacia atrás para calcular la distribución más probable, dada las palabras y los documentos.⁽³³⁾ El algoritmo calcula la proporción de palabras en cada documento asignado a tema, es decir, [*p* (tema T|documento D)] y después la proporción de las palabras que se asignaron a un tema sobre todos los documentos, o sea, [*p* (palabra W|tema T)].⁽³⁴⁾ Su implementación se realiza sobre la matriz de término-documento que, a través de *tidyverse*,⁽³⁵⁾ se convirtió en un objeto *tidy* y se seleccionó el Gibbs⁽³²⁾ como método de muestreo y la beta, el tipo de distribución. La elección de la cantidad de tópicos se prefijó en cinco. La visualización se realizó con *ggplot2*.⁽³⁶⁾ No se requiere del uso de etiquetas *a posteriori* para identificar los tópicos.

El código y los *datasets* pueden ser consultados de forma pública.⁽³⁷⁾ Únicamente no está presente en el código las claves de acceso provistas por la API de Twitter por advertirse de su uso privado. Y en el caso del dataset de *tweets*, en estos no se divulga información que comprometa la privacidad de los usuarios investigados, en correspondencia con el cumplimiento de los requerimientos de *Twitter* acerca del uso responsable de los datos personales de sus usuarios. Lo anterior quedó refrendado en el formulario de obligatorio llenado que debía ser completado para la asignación de la cuenta de desarrollador.

Resultados

En el período establecido para las capturas de los *tweets* con los *hashtags* #Soberana02 y #Abdala se obtuvo una muestra de 3251 entre *retweets* (2133) y *tweets* (1118). En función de la representación gráfica se excluyeron aquellas con frecuencia \leq , lo que dio como resultado 67 palabras con mayor frecuencia.

Entre los términos con que más se dialoga en *Twitter*, por solo mencionar los 10 primeros, se encuentran: dosis (383), vacunas (373), eficacia (163), cubanos (153), candidatos (152), millones (131), país (125), personas (122), recibido (95) y población (94). Como es de apreciar se refleja el contexto que se vivió el país acerca de la campaña de inmunización, el cual es representado en el discurso digital como un proceso colectivo (fig. 1).

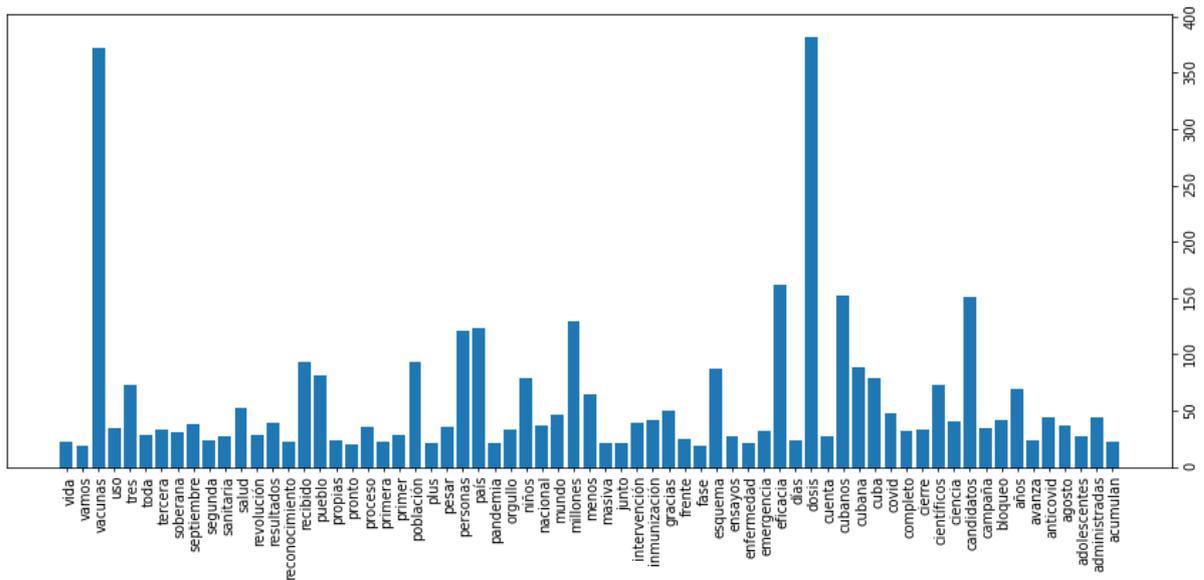
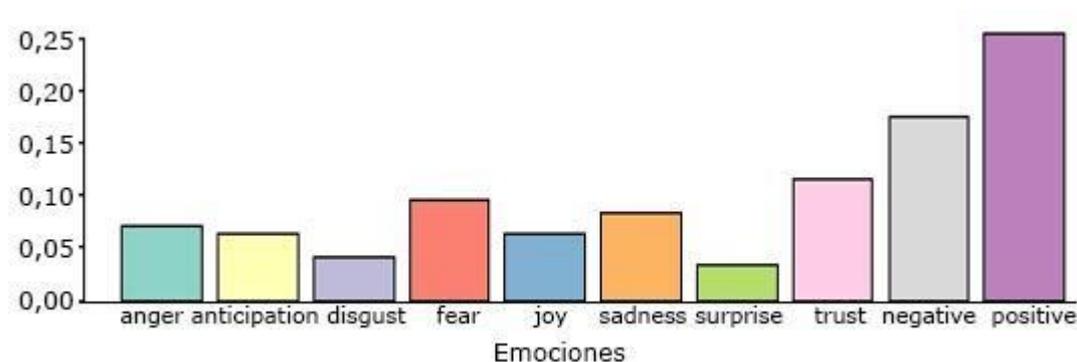


Fig. 1 – Distribución de los términos más empleados.

En cuanto a los sentimientos, según los términos que formaron el contexto semántico, se identificaron aquellos que expresaron emociones como la confianza (12 %) y el miedo (10 %), que también se aprecia en la polaridad, donde predominan *tweets* con una polaridad positiva (25 %) (fig. 2).



Nota al pie: (anger = enojo, anticipation = anticipación, disgust = disgusto, fear = miedo, joy = alegría, sadness = tristeza, surprise = sorpresa, trust = confianza, negative = negativo, positive = positivo)

Fig. 2 – Distribución de las emociones y la polaridad.

A continuación, se muestran algunos de los términos más frecuentes, según las emociones representadas, así como la polaridad. Se identificó tristeza a partir de términos como:

emergencia, pandemia, cierre; en tanto que la confianza fue en las palabras: solidaridad, esperanza, presidente y la polaridad positiva de los *tweets* en: eficacia, recibido, orgullo, completo (fig. 3).



Fig. 3 – Distribución de las palabras por tipo de emoción y polaridad.

De manera longitudinal se puede apreciar que en los primeros *tweets* la polaridad del sentimiento comienza siendo negativo hasta que se va estableciendo una tendencia hacia una polaridad positiva (fig. 4).

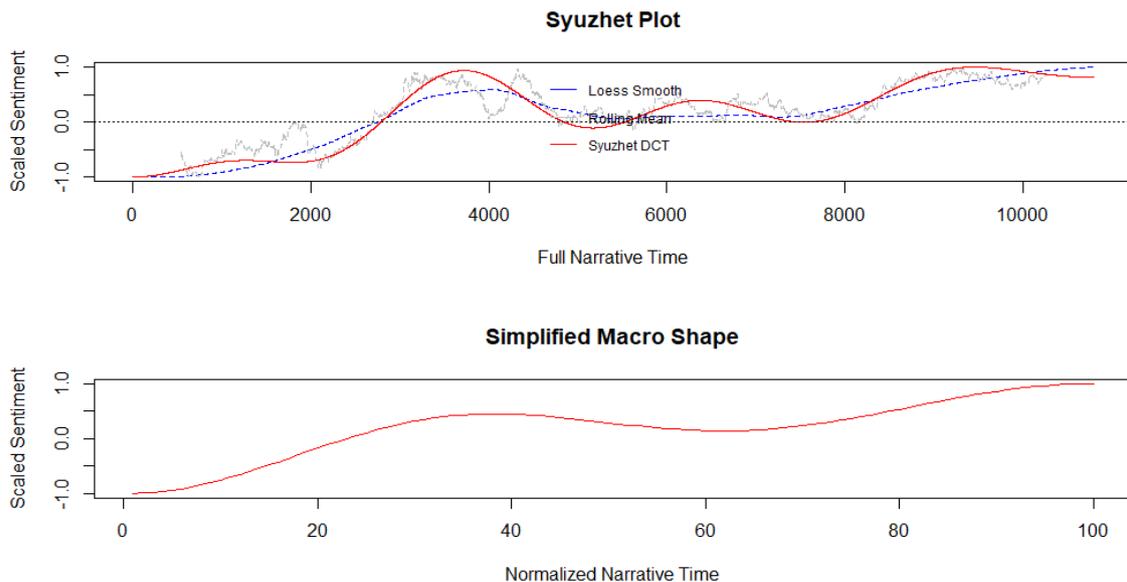


Fig. 4 – Evolución de la polaridad en los *tweets*.

En cuanto a los tópicos se identificó que los temas están compuestos por los siguientes términos:

- Tema 1: uso, plus, COVID, proceso, pandemia.
- Tema 2: revolución, agosto, pediátrica, reconocimiento, responsabilidad.
- Tema 3: pueblo, esperanza, días, bloqueo, candidatos, autorizo.
- Tema 4: masiva, enfermedad, toda, resultado, cubanos.
- Tema 5: dosis, vacunas, cubanos, candidatos, millones (fig. 5).

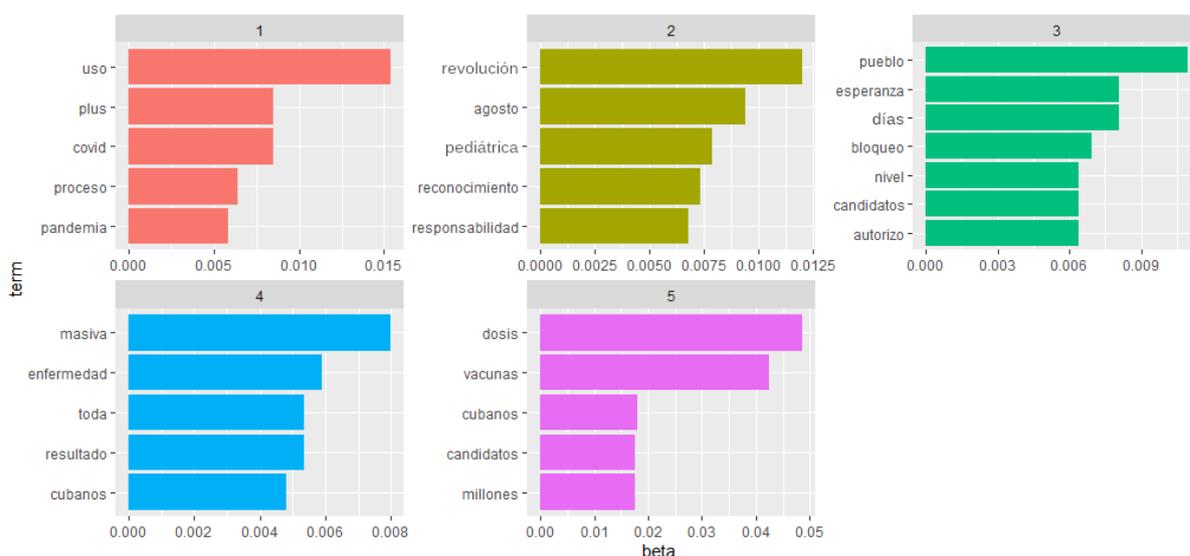


Fig. 5 – Distribución probabilística de palabras por tópicos.

Discusión

La medida de confinamiento social como respuesta a la pandemia de la COVID-19 precipitó la presencia de la Internet en la población cubana. Entre el 2021 y el 2022 hubo un incremento del 5,1 % de celulares conectados y una penetración de la Internet del 68 %.⁽³⁸⁾ Por lo tanto, el escenario digital, como entorno donde también se desarrolla la conversación pública, es susceptible de ser abordado por las Ciencias Sociales computacionales para dar

cuenta de cómo se interactúa con los acontecimientos colectivos. En este estudio se abordó lo relativo a un evento de salud, la vacunación masiva ocurrida en el país, al ser la estrategia más democrática para producir inmunidad biológica colectiva y afrontar en mejores condiciones el virus.

Rodríguez y otros⁽³⁹⁾ en su estudio con 38,034 *tweets*, recopilados entre el 12 de mayo y el 30 de septiembre del 2020 de usuarios colombianos, mostraron que un 42 % manifestó miedo, el 38 % ira, el 15 % alegría y el 5 % tristeza, en relación con la vacuna. Aunque en este caso las reacciones estudiadas se corresponden, teniendo en cuenta la fecha, con el anuncio de la vacuna en sus distintas fases de ensayo clínico, por lo que aún se desconocía su eficacia y efectividad en los seres humanos. Por lo tanto, los resultados responden a las anticipaciones que se generan en contextos de incertidumbre, además de la fuerte polarización política vivida en ese país en torno a la gestión gubernamental de la pandemia. En este mismo país Arias García y Doria Pérez⁽⁴⁰⁾ estudiaron el impacto de la vacunación durante el período del 15 de marzo hasta el 25 de abril de 2021 con una muestra de 1504 *tweets*. Se clasificaron 612 *tweets* positivos, 486 negativos y 407 neutrales con un porcentaje del 40,66 %, 32,29 % y 27,04 %, respectivamente.⁽⁴⁰⁾

Vallejo⁽⁴¹⁾ en su estudio sobre el Ecuador y tres provincias específicas, como parte de la primera parte del plan de vacunación entre los meses de enero y julio del 2021, encontró que en la polaridad de los *tweets* predominó la positiva y, en menor medida, la neutral; la negativa fue más baja. Lo anterior también se reflejó en el análisis temporal realizado, aunque no llegó a determinar si la polaridad de los *tweets* se relacionaba, de forma estadísticamente significativa, con la pertenencia a alguna de las provincias.

Rodríguez y otros⁽⁴²⁾ con una muestra de 1 millón de *tweets* correspondientes a países y regiones, cuya lengua es el inglés o español, halló un 35 % de polaridad, tanto positiva como negativa. A través de la técnica de *clustering* encontraron que una mayor polaridad negativa se encuentra en temas como la desigualdad en el acceso a las vacunas y el mantenimiento de las medidas sanitarias a pesar de la vacunación. Entre tanto la polaridad positiva fue predominante en la inoculación en figuras públicas, la gratuidad de las vacunas y los efectos sobre la salud. La polaridad neutra fue mayor en temas como las *fake news* y la presión política sobre la vacunación.

Liu y Liu⁽⁴³⁾ caracterizaron los sentimientos en 2 678 372 *tweets* provenientes de los Estados Unidos entre el 20 de noviembre del 2020 y el 31 de enero del 2021. Se obtuvo que el 42,8

% fueron positivos, 26,9 % neutrales y el 30,3 % negativos. La polaridad positiva se relaciona con tópicos como los resultados de los ensayos, la administración, la vida, la información y la eficacia. Mientras que en los de orientación negativa se coincidió, con relación a los temas que también generaron sentimientos positivos, en los resultados de ensayos y la administración; en tanto que se difirió en los siguientes temas: conspiración, confianza y efectividad.

Povedano y otros⁽⁴⁴⁾ realizaron un estudio en 4 000 000 de *tweets* geolocalizados, en su mayoría, en los Estados Unidos entre el 15 de noviembre y el 16 de diciembre del 2020. Reportan un porcentaje importante de *tweets* negativos, pertenecientes a los usuarios de estados del centro y sur;⁽⁴⁴⁾ mientras que los de sentimiento positivo se ubicaban en el este. Entre los tópicos que se detectaron como los principales temas y preocupaciones relacionados con la vacuna fueron: la efectividad de la vacuna, la seguridad de la vacuna, los efectos secundarios (en especial como puede influir la vacuna en la fertilidad y las posibles reacciones alérgicas), la distribución de la vacuna y qué grupo de población será el primero en recibir las dosis iniciales.⁽⁴⁴⁾

Con una muestra de 2 970 *tweets* *Carrasco-Polaino* y otros⁽⁴⁵⁾ investigaron acerca de la polaridad de los sentimientos hacia las primeras vacunas durante el tiempo en que fueron anunciados sus resultados de eficacia. Los autores refieren que, de manera general, la polaridad fue positiva, aunque por vacunas hubo diferencias estadísticamente significativas en el gradiente de positividad: la Pfizer y Moderna tuvieron más alto nivel de positividad, seguido de las vacunas chinas y la Oxford-AstraZeneca con nivel medio y la Sputnik V en medio bajo. También reportaron que las vacunas occidentales, dígase Oxford-AstraZeneca, Pfizer y Moderna en ese orden, tuvieron mayores índices de favorabilidad, en contraste con el alto índice de polémica que tuvieron las vacunas chinas y la rusa.

Roe y otros⁽⁴⁶⁾ detectaron la polaridad del sentimiento en una muestra de 137 781 *tweets* del continente europeo entre el 1ro y el 21 de julio del 2021. Los resultados indican que 53 899 fueron negativos, 53 071 positivos y 30 811 neutrales. Según el análisis temporal reportado de las tres semanas, la polaridad negativa predominó en las dos últimas. Entre los términos que identificaron asociados a la polaridad se encuentran: el *hashtag* #*covid19* y la palabra persona; ambos vinculados, tanto a los sentimientos negativo, como positivo y neutral; en tanto que con una polaridad diferente, asociada a la positiva están las palabras: obtener, ayuda y vacuna, respectivamente.

Es válido resaltar que en todos los trabajos citados se refiere la polaridad del *tweet*, mientras que en los resultados hallados en el presente estudio se identificó la polaridad y varias emociones. Además, se mostraron aquellos términos que las representaban, por lo que se parte de esta distinción para contrastar la evidencia. Se aprecia diferencias con los hallazgos de *Povedano y otros*⁽⁴⁴⁾ y *Roe y otros*,⁽⁴⁶⁾ quienes describen tópicos relacionados, fundamentalmente, con la inseguridad ante las distintas circunstancias en el contexto de la enfermedad, de la que forma parte la vacunación; mientras que en el presente estudio predominó la polaridad positiva y los temas detectados están vinculados con la aplicación de las vacunas como proceso colectivo y de consenso

En relación con el contenido de los tópicos se han encontrado coincidencias con *Rodríguez y otros*⁽⁴²⁾ y *Liu y otros*,⁽⁴³⁾ aunque el primer estudio reporta una igualdad en la polaridades positivo-negativo, en tanto el segundo muestra que predominó del sentimiento positivo. Por tanto, la evidencia refleja varias contradicciones. Se debe tener en cuenta que en el hemisferio occidental es donde más influencia poseen los grupos antivacunas y las redes sociales emergen como espacios de intercambio comunicativo con ciertas desregulaciones que han conducido a que las grandes empresas de redes sociales implementen algoritmos para detectar información falsa. Sin embargo, en el estudio de *Rodríguez y otros*⁽³⁹⁾ y *Vallejo*⁽⁴¹⁾ existieron semejanzas en el predominio de la polaridad, pues analizaron *tweets* geolocalizados en Latinoamérica y fue mayoritario el sentimiento positivo hacia las vacunas.

Conclusiones

En el estudio se identificaron cinco tópicos que reflejan el contexto de la vacunación como un proceso colectivo. De manera ligeramente predominante hay presencia de emociones como la confianza y el miedo con una polaridad positiva. Los términos empleados se corresponden con el contexto vivido en el cual se desarrolló la campaña de vacunación. En sentido general, de acuerdo con los términos más frecuentes empleados para referirse a las vacunas, los tópicos identificados, los términos que se relacionaron con las emociones predominantes, así como por la polaridad, se apreció consenso en torno a las vacunas Soberana 02 y Abdala.

Referencias bibliográficas

1. Mejía GK. Enriquecimiento del modelo basado en reglas Vader a través de lexicones (Tesis de Diploma). México: Universidad Autónoma del Estado de México; 2018 [acceso 21/06/2022]. Disponible en: <http://ri.uaemex.mx/bitstream/handle/20.500.11799/99592/2018-Kevin+Mejia+Gonzalez-Tesis.pdf?sequence=1>
2. Ashish Kumar MBBS, Safi U Khan MD, Ankur Kalra MD FACP FACC FSCAI. COVID-19 pandemic: a sentiment analysis: A short review of the emotional effects produced by social media posts during this global crisis. European Heart Journal. 2021;41(39):3782-3. DOI: <https://doi.org/10.1093/eurheartj/ehaa597>
3. Valdez D, ten Thij M, Bathina K, Rutter LA, Bollen J. Social Media Insights into US Mental Health During the COVID-19 Pandemic: Longitudinal Analysis of Twitter Data. J Med Internet Res. 2020;22(12):e21418. DOI: <https://doi.org/doi:10.2196/21418>
4. Arango Pastrana CA, Osorio Andrade CF. Aislamiento social obligatorio: un análisis de sentimientos mediante machine learning. Suma de Negocios. 2021;12(26):1-13. DOI: <https://doi.org/10.14349/sumneg/2021.v12.n26.a1>
5. Saleh SN, Lehmann CU, McDonald SA, Basit MA, Medford RJ. Understanding public perception of coronavirus disease 2019 (COVID-19) social distancing on Twitter. Infection Control & Hospital Epidemiology. 2021;42(2):131-8. DOI: <https://doi.org/10.1017/ice.2020.406>
6. Salaberry N. Análisis de contenido en Twitter y el aislamiento social obligatorio. Revista de Investigación en Modelos Matemáticos Aplicados a la Gestión y la Economía. 2020 [acceso 23/06/2022];7(1):1-15. Disponible: <http://www.economicas.uba.ar/wp-content/uploads/2016/04/Salaberry-Natalia.pdf>
7. Vela Delfa C, Cantamutto L, Núñez-Cansado M. Análisis del hashtag #MeQuedoEnCasa: la conversación digital por twitter® ante la emergencia producida por la COVID-19. Íkala, Revista de Lenguaje y Cultura. 2021;26(2):365-84. DOI: <https://doi.org/10.17533/udea.ikala.v26n2a08>

8. Ferreyra SG, Nieto AA, Juarez WI. Mar del Plata en Twitter: comunidades y tópicos durante la cuarentena. *Revista Enlace Universitario*. 2020;35(7):1-15. Disponible en: <https://www.mdp.edu.ar/attachments/article/82/Enlace35.pdf>
9. Sued Palmeiro GE, Cebral Loureda M. Voces autorizadas en Twitter durante la pandemia de COVID-19: actores, léxico y sentimientos como marco interpretativo para usuarios ordinarios. *Revista de Comunicación y Salud*. 2020 [acceso 23/06/2022];10 (2):549-68. DOI: [https://doi.org/10.35669/reys.2020.10\(2\).549-568](https://doi.org/10.35669/reys.2020.10(2).549-568)
10. Chandrasekaran R, Mehta V, Valkunde T, Moustakas E. Topics, Trends, and Sentiments of Tweets About the COVID-19 Pandemic: Temporal Infoveillance Study. *J Med Internet Res*. 2020;22(10):e22624. DOI: <https://doi.org/10.2196/22624>
11. Xue J, Chen J, Hu R, Chen C, Zheng C, Su Y, *et al.* Twitter Discussions and Emotions about the COVID-19 Pandemic: Machine Learning Approach. *J Med Internet Res*. 2020;22(11):e20550. DOI: <https://doi.org/10.2196/20550>
12. Cebral-Loureda M, Sued-Palmeiro GE. Análisis computacional de la conversación pública en lengua española. *Cuadernos Info*. 2021(49):1-25. DOI: <https://doi.org/10.7764/cdi.49.27467>
13. Torres JA. Análisis de opinión sobre tuits del COVID-19 generados por usuarios ecuatorianos. *CEDAMAZ*. 2021 [acceso 24/06/2022];11(1):70-7. Disponible en: <https://revistas.unl.edu.ec/index.php/cedamaz/article/download/1039/791>
14. Alles Torrent S, del Rio Riande G, De León R, Fila M, Hernández N, Jerry Bonnell, Song D. Narrativas digitales de la COVID-19 en Twitter: de los datos a la interpretación. *publicaahd* [Internet]. 2020 [acceso 24/06/2022];1:e002. Disponible en: <https://revistas.unlp.edu.ar/publicaahd/article/view/13771>
15. Condor Tinoco EE, Rojas Cusi JA, Zevallos Rodríguez A, Castro Buleje CY. Minería de datos: análisis de sentimiento en Twitter basado en lexicones sobre el uso de dióxido de cloro para el tratamiento del COVID-19. *Actas del III Congreso Internacional de Ingeniería de Sistemas*. 2021 [acceso 25/06/2022]. Disponible en: https://repositorio.ulima.edu.pe/bitstream/handle/20.500.12724/13898/Condor_Rojas_Zevallos_Castro_Miner%3%ADa-de-datos-an%3%A1lisis-de-sentimiento-en-Twitter.pdf?sequence=1&isAllowed=y

16. Sharma C, Whittle S, Haghghi PD, Burstein F, Keen H. Sentiment analysis of social media posts on pharmacotherapy: A scoping review. *Pharmacol Res Perspect.* 2020;8:e00640. DOI: <https://doi.org/10.1002/prp2.640>
17. Méndez-Hernández GM, Fonseca Valido R, Alonso Pérez M. Conversaciones sobre brigadas médicas cubanas en contexto de COVID-19 desde el ambiente info-comunicativo de Twitter. *Rev haban cienc méd.* 2021 [acceso 25/06/2022];20(5):e4173. Disponible en: <http://www.revhabanera.sld.cu/index.php/rhab/article/view/4173>
18. MINSAP. Ministro de Salud: “Hacer juntos es la mejor manera de consolidar el éxito en el enfrentamiento a la epidemia”. 2021 [acceso 25/06/2022]. Disponible en: https://salud.msp.gob.cu/ministro-de-salud-hacer-juntos-es-la-mejor-manera-de-consolidar-el-exito-en-el-enfrentamiento-a-la-epidemia/?doing_wp_cron=1622123296.0725009441375732421875
19. Statcounter GlobalStats. [acceso 25/06/2022] Disponible en: <https://gs.statcounter.com/social-media-stats/all/cuba>
20. Twitter. <https://developer.twitter.com>
21. Tweepy Documentation. Disponible en: <https://docs.tweepy.org>
22. Pandas Documentation. PyData.org. Disponible en: <https://pandas.pydata.org/pandas-docs/stable>
23. Re. Operaciones con expresiones regulares. Disponible en: <https://docs.python.org/es/3/library/re.html>
24. NLTK: Natural Language Toolkit. Disponible en: <https://www.nltk.org>
25. Hunter JD. Matplotlib: A 2D graphics environment. En Matplotlib. Disponible en: <https://matplotlib.org/2.0.2/Matplotlib.pdf>
26. Mullen L. Introduction to the tokenizers Package. [acceso 27/06/2022] Disponible en: <https://cran.r-project.org/web/packages/tokenizers/vignettes/introduction-to-tokenizers.html>
27. Mateo J. Introduction to the Syuzhet Package. 2020 [acceso 27/06/2022]. Disponible en: <https://cran.r-project.org/web/packages/syuzhet/vignettes/syuzhet-vignette.html>

28. Naldi M. A review of sentiment computation methods with R packages. arXivLabs. 2019. DOI: <https://doi.org/10.48550/arXiv.1901.08319>
29. barplot: Bar Plots. [acceso 27/06/2022] Disponible en: <https://www.rdocumentation.org/packages/graphics/versions/3.6.2/topics/barplot>
30. Fellows I. Package ‘wordcloud’. 2018 [acceso 27/06/2022]. Disponible en: <https://cran.r-project.org/web/packages/wordcloud/wordcloud.pdf>
31. Feinerer I. Introduction to the tm Package Text Mining in R. 2020 [acceso 28/06/2022]. Disponible en: <https://cran.r-project.org/web/packages/tm/vignettes/tm.pdf>
32. Silvestre Gómez M. Implementación de asignación jerárquica latente de Dirichlet para modelado de temas (Tesis de maestría). Sevilla: Universidad de Sevilla; 2018 [acceso 28/06/2022]. Disponible en: <https://idus.us.es/bitstream/handle/11441/74394/MAR%c3%8dA%20SILVESTRE%20G%c3%93MEZ.pdf?sequence=1&isAllowed=y>
33. Kozlowski D, Shokida N. Notas de clase del curso de introducción a Data Science. 2019 [acceso 29/06/2022]. Disponible en: https://diegokoz.github.io/intro_ds_bookdown/explicacion-9.html#topic-modelling
34. Melton CA, Olusanya OA, Ammar N, Shaban-Nejad A. Public sentiment analysis and topic modeling regarding COVID-19 vaccines on the Reddit social media platform: A call to action for strengthening vaccine confidence. J Infect Public Health. 2021;14(10):1505-12. DOI: <https://doi.org/10.1016/j.jiph.2021.08.010>
35. Corcoran D. Manipulación de datos e investigación reproducible en R. 2018 [acceso 29/06/2022]. Disponible en: <https://bookdown.org/content/3515/tidydata.html>
36. ggplot2. Disponible en: <https://r-graph-gallery.com/ggplot2-package.html>
37. Rodríguez Sánchez A, Rivero González LM, Soca Losano S. Análisis de sentimientos y tópicos sobre las vacunas cubanas Soberana 02 y Abdala en Twitter [preprint]. 2022 [acceso 29/06/2022]. Disponible en: <https://github.com/luisdev98/sentimentAnalysis>
38. Datareportal. Digital 2022: Cuba. 2022 [acceso 30/06/2022]. Disponible: <https://datareportal.com/reports/digital-2022-cuba>

39. Rodríguez-Orejuela A, Montes-Mora CL, Osorio-Andrade CF. Sentimientos hacia la vacunación contra la COVID-19: panorama colombiano en Twitter. Palabra Clave. 2022;25(1),e2514. DOI: <https://doi.org/10.5294/pacla.2022.25.1.4>
40. Arias García HL, Doria Pérez LC. Análisis de sentimientos sobre la percepción ciudadana de la vacunación del COVID-19 en Colombia. 2021 [acceso 30/06/2022]. Disponible en: <http://repositorio.uan.edu.co/bitstream/123456789/5160/1/2021H%C3%A9ctorLeonardoAriasGarc%C3%ADa.pdf>
41. Vallejo Cabrera PX. Análisis de sentimientos sobre la vacuna para COVID-19 en la red social “twitter” en el contexto ecuatoriano (tesis de Licenciatura). Ecuador: Universidad del Azuay; 2022 [acceso 30/06/2022]. Disponible en: <https://dspace.uazuay.edu.ec/bitstream/datos/11709/1/17238.pdf>
42. Rodríguez K, Haber Guerra Y, Fonseca Valido RA. Análisis discursivo de las vacunas anticovid-19 en Twitter. Question/Cuestión. 2021;70(3). DOI: <https://doi.org/10.24215/16696581e624>
43. Liu S, Liu J. Public attitudes toward COVID-19 vaccines on English-language Twitter: A sentiment analysis. Vaccine. 2021;39(39):5499-505. DOI: <https://doi.org/10.1016/j.vaccine.2021.08.058>
44. Povedano Álvarez D, Portela García-Miguel J, Armas Vega EA. Estudio de la percepción pública de la vacuna contra la COVID-19 mediante técnicas de PLN y de aprendizaje automático [Tesis de Maestría]. España: Facultad de Estadística, Universidad Complutense de Madrid; 2021 [acceso 01/07/2022]. Disponible: <https://eprints.ucm.es/id/eprint/67617/1/daniel-povedano-estudio.pdf>
45. Carrasco-Polaino R, Martín-Cárdaba MÁ, Villar-Cirujano E. Participación ciudadana en Twitter: Polémicas anti-vacunas en tiempos de COVID-19. Comunicar: Revista científica Iberoamericana de comunicación y educación. 2021;(69):21-31. DOI: <https://doi.org/10.3916/C69-2021-02>
46. Roe C, Lowe M, Williams B, Miller C. Public Perception of SARS-CoV-2 Vaccinations on Social Media: Questionnaire and Sentiment Analysis. Int J Environ Res Public Health. 2021 [acceso 01/07/2022];18(24):13028. DOI: <https://doi.org/10.3390/ijerph182413028>

Conflicto de intereses

Los autores declaran que no tienen conflicto de intereses.

Contribuciones de los autores

Conceptualización: Antonio Rodríguez Sánchez, Luis Manuel Rivero González, Sandra Soca Lozano.

Curación de datos: Luis Manuel Rivero González.

Análisis formal: Antonio Rodríguez Sánchez, Luis Manuel Rivero González.

Investigación: Antonio Rodríguez Sánchez, Sandra Soca Lozano.

Metodología: Antonio Rodríguez Sánchez, Luis Manuel Rivero González, Sandra Soca Lozano.

Software: Antonio Rodríguez Sánchez, Luis Manuel Rivero González.

Visualización: Antonio Rodríguez Sánchez, Luis Manuel Rivero González.

Redacción – borrador original: Antonio Rodríguez Sánchez.

Redacción – revisión y edición: Antonio Rodríguez Sánchez, Luis Manuel Rivero González, Sandra Soca Lozano.