

Medical Supply Management in Ecuador during COVID-19: A novel approach from Twitter

Ester Melo Vargas, Lic.¹, Luis Quijije Cevallos, Lic.²,

Mentor: Christian Vera Alcívar, MSc.³

^{1,2,3}Facultad de Ciencias Sociales y Humanísticas (FCSH), Escuela Superior Politécnica del Litoral (ESPOL), Campus Gustavo Galindo Km. 30.5 Vía Perimetral, P.O. Box 09-01-5863, Guayaquil, Ecuador, emelo@espol.edu.ec, lquijije@espol.edu.ec, cxvera@espol.edu.ec

Abstract— Historically, the management of medical supplies in Ecuador has posed a challenge to timely disease treatment due to their insufficient availability. This situation was exacerbated at the beginning of 2020 due to the coronavirus pandemic. Consequently, individuals resorted to using social media platforms, such as Twitter, to access medicines and medical equipment. Therefore, the present study aims to evaluate the management of medical supplies in Ecuador by the public through the Twitter social network, employing text mining techniques applied to big data, to determine its effectiveness during the Covid-19 pandemic from March 2020 to March 2021. To achieve this, interviews were conducted with physicians to identify the most demanded medical supplies. Subsequently, web scraping was applied to extract relevant posts about the management of supplies on Twitter. Text mining techniques were then employed to identify emotions, classify needs, and validate information. Lastly, effectiveness indicators were established. The results revealed that March and July 2020, as well as January 2021, were critical periods, especially in provinces such as Guayas, Azuay, Imbabura, and Santa Elena, consistent with the information reported by Ecuador's Emergency Operations Committee (COE). It is concluded that this study provides valuable information for the formulation of public policies related to the supply of medical supplies and healthcare provision.

Keywords—Medical supplies, Coronavirus, Twitter, Big Data, Ecuador.

Digital Object Identifier: (only for full papers, inserted by LACCEI).

ISSN, ISBN: (to be inserted by LACCEI).

DO NOT REMOVE

Análisis de la gestión de insumos médicos en Ecuador durante el Covid-19: Un nuevo enfoque desde Twitter

Ester Melo Vargas, Lic.¹, Luis Quijije Cevallos, Lic.²,

Mentor: Christian Vera Alcívar, MSc.³

^{1,2,3}Facultad de Ciencias Sociales y Humanísticas (FCSH), Escuela Superior Politécnica del Litoral (ESPOL), Campus Gustavo Galindo Km. 30.5 Vía Perimetral, P.O. Box 09-01-5863, Guayaquil, Ecuador, emelo@espol.edu.ec, lquijije@espol.edu.ec, cxvera@espol.edu.ec

Resumen— *Históricamente, la gestión de insumos médicos en Ecuador ha representado un obstáculo en el tratamiento oportuno de enfermedades debido a la insuficiencia de estos. Dicha situación se agudizó a principios del 2020 por la pandemia del coronavirus. En consecuencia, las personas optaron por emplear redes sociales, como Twitter, para obtener medicinas y equipo médico. Por ello, el presente estudio tiene como objetivo evaluar la gestión de insumos médicos en Ecuador por parte de la ciudadanía, a través de la red social Twitter empleando técnicas de minería de texto aplicadas a big data, para determinar su grado de efectividad durante la pandemia de la Covid-19 desde marzo del 2020 hasta marzo del 2021. Para lograrlo se realizaron entrevistas a médicos para conocer los insumos más demandados. A partir de ello se aplicó web scraping en Twitter para extraer publicaciones sobre la gestión de insumos. Posteriormente se aplicaron técnicas de minería de texto para identificar emociones, clasificar necesidades y validar la información. Por último, se establecieron indicadores de efectividad. Los resultados revelaron que marzo y julio del 2020 y enero del 2021 fueron tiempos críticos, especialmente en provincias como Guayas, Azuay, Imbabura y Santa Elena; siendo consistentes con la información reportada por el Comité de Operaciones de Emergencia (COE) de Ecuador. Se concluye que este estudio brinda información de interés para la formulación de políticas públicas relacionadas al abastecimiento de insumos y la atención médica.*

Palabras claves— *Insumos médicos, Coronavirus, Twitter, Big Data, Ecuador.*

I. INTRODUCCIÓN

Históricamente, la gestión de insumos médicos en Ecuador ha representado un obstáculo en el tratamiento oportuno de enfermedades o situaciones de emergencia a nivel de salud [1]–[4]. A principios de 2020, dicha falencia se agudizó debido a la pandemia provocada por la Covid-19, la cual ha causado millones de muertes en todo el mundo. Además, durante la pandemia se reportaron casos de corrupción respecto a la distribución de insumos médicos [5], [6]. En consecuencia, las personas vieron la necesidad de emplear canales de comunicación modernos, como las redes sociales, para obtener medicinas y equipo médico que incluso a nivel hospitalario eran insuficientes [7].

Sin embargo, a la fecha se desconocen sobre estudios enfocados en la evaluación de la gestión de este tipo de

recursos en el territorio ecuatoriano y, sobre todo, de indicadores que permitan medir su impacto en diferentes etapas de la pandemia. Adicionalmente, no se cuenta con datos respecto a la identificación de patrones de comportamiento y zonas críticas o de alta demanda de insumos médicos a través de las redes sociales.

La escasez de esta información imposibilita obtener una visión completa de la realidad ecuatoriana en la lucha contra el aumento de casos positivos y muertes por coronavirus. Así mismo, invisibiliza las prácticas de gestión de insumos médicos ejecutadas por la ciudadanía que podrían guiar la toma de decisiones por parte de las autoridades locales.

Por ello, en el presente trabajo de investigación se desarrollará un análisis utilizando técnicas de minería de texto aplicadas a big data a partir de publicaciones extraídas de Twitter relacionadas a la gestión de insumos médicos en el territorio ecuatoriano durante la pandemia de la covid-19. Dicho análisis se basará en fundamentos teóricos y metodológicos para la construcción de indicadores que permitan monitorear la situación sanitaria y las respuestas de la población.

Esto debido a que, dicha gestión debe satisfacer las necesidades de la sociedad en la mayor medida posible y contribuir a la mitigación de casos positivos de coronavirus. En consecuencia, se presentarán los principales hallazgos respecto al impacto que posee Twitter como medio de difusión de la demanda de dichos insumos y la identificación de zonas críticas.

A. La pandemia del coronavirus en Ecuador

La enfermedad por coronavirus fue caracterizada como pandemia el 11 de marzo del 2020 por la Organización Mundial de la Salud (OMS), debido a la extensión de la epidemia por todos los continentes. Para esa fecha, Ecuador ya registraba 17 casos positivos de coronavirus y, de hecho, ocupaba el tercer lugar en Latinoamérica por ello, después de Brasil y Argentina [8].

En consecuencia, el Ministerio de Salud Pública del Ecuador emitió el Acuerdo Ministerial No. 00126-2020 y estableció el Estado de Emergencia Sanitaria. A partir de este acuerdo se estipularon medidas para contener la propagación del coronavirus, entre las cuales se destacan las siguientes:

Digital Object Identifier: (only for full papers, inserted by LACCEI).

ISSN, ISBN: (to be inserted by LACCEI).

DO NOT REMOVE

- Adopción de medidas de prevención en el transporte público por parte de los Gobiernos Autónomos Descentralizados (GAD).
- Cuidado especial a los adultos mayores, personas con discapacidad y con enfermedades crónicas y catastróficas.
- Fortalecimiento de las medidas de bioseguridad al personal de salud.
- Prevención y control en los puntos de ingreso al país.
- Restricciones para eventos masivos y de concentración de gente, considerando la zona y el número de casos.
- Uso de plataformas tecnológicas en telemedicina, educación en línea y teletrabajo para aquellos casos que lo ameriten.
- Otras [9].

Sin embargo, dichas medidas estuvieron acompañadas de una inadecuada gestión, resultando inefectivas y conllevando a la agudización de la situación sanitaria del país [10]. De tal manera que, el 13 de marzo el presidente de turno decretó el Estado de Excepción, dando lugar al Comité de Operaciones de Emergencia (COE Nacional) para la coordinación de la crisis sanitaria [11].

B. Revisión de literatura

García & Berton [12] exploran una amplia cantidad de tweets relacionados al covid-19 en Brasil y Estados Unidos a través de la utilización de la identificación de temas y el análisis de sentimientos; donde concluyen que, la mayoría de los mensajes analizados son negativos dado que son datos sobre la pandemia del coronavirus; sin embargo, al observar las cifras, se evidencia diferencias significativas entre algunos temas en los que fueron clasificados los tweets previamente; ya que por un lado el cuidado de la proliferación fue el tema con mayor porcentaje de mensajes negativos, alrededor de 60% en ambos países; y por otro lado, los tratamientos fueron el tema con mayor cantidad de mensajes positivos, el 33% de los mensajes en total.

Por otro lado, Khan et. al [13] realizaron una investigación con el fin de brindar un análisis de sentimiento sobre la reacción de las personas hacia las decisiones relacionadas al covid-19 tomadas por el gobierno o autoridades locales, mediante un sistema que automatiza el análisis de los tweets y los clasifica como positivos, negativos o neutros. Los resultados demostraron que el número máximo de personas pensó y percibió como positivas dichas decisiones a pesar del aumento de personas infectadas y fallecidas. También se realizó un análisis de 3 meses correspondiente al subcontinente indio, el cual mostró que la variación de los sentimientos positivos, negativos y neutros se mantuvo constante con el número de casos que incrementaban día a día.

II. METODOLOGÍA

De acuerdo con la metodología empleada en otros estudios [7], [12]–[15], en esta investigación se empleó la

entrevista y el web scraping o raspado web como instrumentos de recolección de datos, para luego realizar un análisis aplicando técnicas de minería de texto. El detalle de los instrumentos y procedimientos ejecutados se describe en las siguientes secciones. La metodología general se resume en la figura 1.



Fig. 1 Etapas de la metodología de la investigación. Elaborado por autores.

A. Entrevista

Para la entrevista se optó por un estilo semiestructurado que constó de tres partes: experiencias durante la pandemia, insumos médicos y perspectiva sobre la población; estas cuentan con seis, cinco y tres preguntas, respectivamente. La entrevista fue dirigida a profesionales de la salud, quienes estuvieron en la primera línea de defensa contra el coronavirus, tales como: médicos generales, internistas y especialistas en áreas afines a inmunología, cardiología, virología, entre otros.

La ejecución de entrevistas se llevó a cabo del 10 al 15 de julio del 2021, de forma virtual a través de la plataforma Zoom con una duración aproximada de dos horas por entrevista. Para ello, primero se estableció el contacto con los entrevistados por medio de una invitación formal vía correo electrónico. En esta invitación se proporcionó un enlace de Doodle, en donde se dio a elegir a los entrevistados los horarios de su preferencia para el desarrollo de la entrevista.

Una vez confirmada la participación de los entrevistados, se procedió a enviarles un correo de agradecimiento como respuesta y, adicionalmente, el enlace de acceso a la entrevista a través de Zoom según el horario más votado por el entrevistado en Doodle. Posteriormente, el día de la entrevista, se envió un correo como recordatorio 15 minutos antes de que esta inicie.

B. Recopilación de datos

Para poder extraer información de la red social Twitter, se utilizó la librería Twint de Python, la cual es de código abierto y permite obtener tweets de los usuarios sin la necesidad de tener las limitaciones de la API, cabe destacar que la red social Twitter tiene limitado el raspado de información de aproximadamente 3200 tweets por mes buscado.

Para la extracción de la información mediante Twint se delimitó la zona geográfica en Ecuador con coordenadas de latitud, longitud y un radio en kilómetros, donde se genera una

circunferencia para una búsqueda más exacta. Adicionalmente, se establecieron parámetros de tiempo desde el 1 de marzo del 2020 hasta el 31 de marzo del 2021, siendo así que se genera un archivo csv detallando los siguientes datos:

- **Id_username:** Nombre del usuario del que realiza el tweet.
- **Retweets_count:** Cuántos retweets tuvo la publicación.
- **Likes_count:** Cuántos corazones tuvo la publicación.
- **Hashtags:** Cuáles son los #hashtags en la publicación.
- **Place:** Coordenadas de latitud y longitud del usuario.
- **Date:** Día en el que se ha realizado la publicación.
- **Language:** En qué idioma se encuentra el tweet.

Por otro lado, para el proceso y lectura de la base de datos se utilizó el lenguaje de programación Python junto con su IDE Pycharm con el fin de realizar un mapeo para identificar a qué provincia pertenecen las coordenadas de cada usuario de twitter. Para ello, se hizo uso de un mapa en formato geojson realizado por Marin [16], el cual permite visualizar las 24 provincias del Ecuador y sus zonas no delimitadas.

Como primer paso se extrajo las coordenadas de los usuarios por tweet, con sus respectivos ejes de longitud y latitud, lo cual permite que este sea catalogado como un punto en el Ecuador. Después, mediante la librería Shapely se buscaba a qué provincia pertenece la persona que realizó el tweet, con el objetivo de generar un mapa de colores, en donde se muestre gráficamente los resultados de forma cuantitativa.

C. Extracción de información

La extracción de información consiste en la obtención de categorías del texto, es decir, determinar el tipo de palabra de acuerdo con su función en la oración. Los tipos de palabras identificadas por esta técnica son: sustantivo, pronombre, verbo, adjetivo, adverbio, preposición, conjunción e intersección [17]. El conocimiento de la categoría de la palabra facilita la identificación de patrones y obtención de palabras claves que permitan contar con información relevante, según los objetivos del investigador [18].

En este caso, se requería conocer las características atribuidas a los dos sustantivos que se presentaban con mayor frecuencia en todas las publicaciones, a los cuales se denominó “palabras bases”. Para ello, se optó por extraer palabras frecuentes y una vez obtenido el par de términos más frecuentes, se establecieron reglas o patrones que permitieran asociar dichos sustantivos con uno o varios adjetivos. En la determinación de las reglas se consideraron todas las palabras derivadas, similares o compuestas respecto a los dos términos más frecuentes; por ejemplo: covid, covid-19, covid19, COVID-19, etc. Los patrones establecidos se detallan en la tabla 1.

TABLA I
DESCRIPCIÓN DE PATRONES

Nº	Patrón	Descripción	Ejemplo
1	Palabra base + adjetivo	Obtener una caracterización simple respecto a la palabra base.	Mascarilla efectiva

2	Palabra base + adjetivo + adjetivo o sustantivo complementario	Obtener una caracterización compuesta respecto a la palabra base.	Mascarilla efectiva KN95
---	--	---	--------------------------

Elaborado por autores.

A partir de dichos patrones, se realizó la búsqueda de tweets que cumplieran con estos. La búsqueda y configuración de reglas fueron llevadas a cabo mediante la librería Spacy en Python. Posteriormente, los resultados fueron cuantificados y visualizados para comprender la percepción de la población respecto a los términos bases y conservar la información relevante.

D. Recuperación de información

Seguido de la extracción de la información, fue necesario aplicar la recuperación de la información como última fase de limpieza previo a la aplicación del procesamiento de lenguaje natural. Como se mencionó, esta técnica se basa en un conjunto de palabras claves previamente definidas que permiten recuperar información útil y relevante de la base de datos textuales conformado por los tweets [19]. Por lo tanto, este paso es fundamental para mejorar la precisión de los tweets asociados a los objetivos de la investigación.

Anteriormente twitter restringía las publicaciones hasta 140 caracteres por tweet, por lo que el usuario ocupaba diferentes métodos tales como: abreviaturas, supresión de letras, creación de hilos de tweets, entre otras cosas; lo cual limitaba la obtención de la información importante. Sin embargo, en el año 2017 la red social duplicó la cantidad de caracteres proporcionándole a sus usuarios mayor amplitud para expresar sus sentimientos en un mismo tweet, logrando a su vez fomentar mejores opiniones y, por ende, aumentando la probabilidad de recuperar información valiosa [20].

E. Procesamiento de lenguaje natural: análisis de sentimientos

Para el análisis de sentimiento se ocupó la librería classifier proveniente de SentimentClassifier, el cual realiza un preprocesamiento de la información eliminando lo que no es relevante, como signos de puntuación que hayan sido mal utilizados; además de remover páginas web, eliminar caracteres especiales, y realizar una modificación en las letras que se encuentren en mayúsculas y minúsculas permitiendo tener un mejor procesamiento de los datos.

Sin embargo, no convierte todas las letras en minúsculas, debido a que trabaja con Word Embeddings, el cual es una técnica del procesamiento natural del lenguaje y la modificación completa de la palabra puede cambiar el contexto gramatical para determinar si el tweet es bueno, neutro o malo. Esto genera una probabilidad que se clasifica de la siguiente manera:

- Cuando la probabilidad se acerca a 0 es considerado un tweet malo.
- Cuando la probabilidad se acerca a 0.5 es considerado un tweet neutral.

- Cuando la probabilidad se acerca a 1 es considerado un tweet bueno.

Lo cual permite identificar las emociones de los tweets y clasificarlas de acuerdo con su sentimiento ya sea bueno, malo o neutral.

F. Agrupación o clustering

Como se mencionó previamente, el análisis de sentimientos proporciona información cualitativa de gran valor para comprender de manera profunda lo que la información cuantitativa por sí sola no puede explicar. Sin embargo, al tratarse de registros de todo el Ecuador, se contaba con una perspectiva general que limitaba la interpretación específica de los resultados. En ese sentido, para mejorar la caracterización del problema, se optó por segmentar o agrupar la información mediante el algoritmo K-medias.

El algoritmo K-medias es uno de los más empleados en la literatura por su simplicidad, sobre todo en problemas de aprendizaje no supervisado [21], como en este caso. No obstante, la ejecución de este algoritmo depende un número K que representa la cantidad de grupos a formar. Para obtener este número se pueden emplear diferentes métodos, entre ellos el método del codo, el cual se utilizó en este estudio debido a su bajo costo computacional y el poco espacio de memoria que consume [22].

Una vez obtenido el número K correspondiente, se ejecutó el algoritmo K-medias considerando como variables el número de likes o corazones, el número de retweets y el valor del sentimiento. Posteriormente, los resultados de este análisis fueron segmentados e ilustrados a nivel provincial para lograr una caracterización más precisa.

G. Síntesis de texto

Para poder validar la información recuperada, se utilizó la librería NLTK de Python, la cual es "un conjunto de módulos de programa de código abierto, tutoriales y conjuntos de problemas, que proporciona material didáctico de lingüística computacional" [23]. La importancia de utilizar esta herramienta radica en que reduce la probabilidad de tener datos aberrantes los cuales podrían sesgar los resultados de la investigación; es decir, si no se realiza la validación, se estarían procesando en exceso datos que no proporcionen información relevante, tales como palabras vacías o stopwords.

Para ello, una vez realizado el pre-procesamiento de información se generó un score utilizando la técnica *cosine similarity* con el fin de calcular la importancia de cada vector. El cosine similarity es un modelo que no necesita ser entrenado debido a que trabaja con un algoritmo interno, en donde convierte las oraciones en vectores calculando el ángulo de coseno de acuerdo con la frecuencia de palabras que se encuentre inmerso en cada uno de los vectores. Luego de calcular la importancia, toma los vectores que tengan la

puntuación más alta para realizar un resumen que tenga estructura gramatical de acuerdo con ese orden.

H. Indicadores de efectividad

En congruencia con lo reportado por estudios similares, los indicadores de efectividad corresponden a medidas de carácter cuantitativo y cualitativo que se pueden extraer de los tweets a fin de conseguir información sobre la interacción y capacidad de respuestas de los usuarios. En este caso, con los resultados de las técnicas aplicadas y los reportes emitidos por el Comité de Operaciones de Emergencia (COE) nacional e instituciones afines, se establecieron dos indicadores: tiempos y zonas críticas.

- Tiempos críticos

Para definir los tiempos críticos, se graficó una serie de tiempo de todos los tweets obtenidos, identificando los picos o puntos destacados que indicaban un comportamiento diferente o significativo. Luego, se compararon los resultados con los tiempos críticos declarados por el COE Nacional, a fin de identificar semejanzas o diferencias y validar los resultados del análisis ejecutado.

- Zonas críticas

Las zonas críticas se definieron a partir los grupos identificados mediante la técnica de agrupación, los cuales fueron ilustrados en el mapa de Ecuador. Los resultados se compararon con la información reportada por el COE Nacional, respecto a las zonas más afectadas durante la pandemia, considerando el período de estudio.

III. RESULTADOS

A. Análisis de entrevistas

De la entrevista con el primer médico se destacó que, tanto a nivel público como privado, se llevaron a cabo las medidas necesarias para afrontar la emergencia sanitaria y que, en todo tiempo, se priorizó el bienestar y la seguridad del personal de salud. Los insumos necesarios como: mascarillas, alcohol, gel antibacterial, entre otros.

Sin embargo, a pesar de que los hospitales cuentan con bombas de oxígeno, este fue cada vez más escaso a medida que avanzaba la pandemia y debido a la poca colaboración de la población al seguir los cuidados recomendados. Según el doctor, esta situación fue más crítica en la segunda ola de la pandemia, es decir, de abril a junio de 2020.

La entrevista con la segunda médica brindó una perspectiva distinta de la obtenida con el primero, ya que, por el contrario, el lugar donde desempeñó sus labores padeció de una gran escasez de mascarillas, específicamente, las N-95. A partir de esta escasez se tuvo que establecer un protocolo a fin de extender la vida útil de este insumo y este consistió en reutilizar las mascarillas tres guardias seguidas, es decir, a la semana solo empleaban dos mascarillas de ese tipo.

A partir de las entrevistas realizadas, se determinaron los insumos más mencionados o comunes y estos fueron: Mascarilla, N-95, Nitazoxanida, Paracetamol, Oxígeno,

Azitromicina, Amoxicilina, Ivermectina, Jengibre, Canela, Eucalipto, Árbol de neem, Colufase, Alcohol y Gel antibacterial.

Con este conjunto de palabras, se procedió a realizar la extracción de datos y a aplicar las técnicas de minería de texto establecidas en la sección anterior.

B. Recopilación de datos

En la búsqueda del archivo se ocuparon las palabras claves que se obtuvieron directamente de las entrevistas, las cuales son: mascarilla, N-95, nitazoxadina, paracetamol, oxígeno, axitromicina, amoxicilina, ivermectina, jengibre, canela, eucalipto, árbol de neem, colufase, alcohol, gel antibacterial; lo cual dio como resultado una base de datos pura de 131098 tweets en la que estaban incluidos los tweets de todo el país. Además, se obtuvieron las características de los tweets, tal como retweets y likes o corazones. La cantidad de retweets fue de 556151 y la cantidad de likes fue de 1276567, estos posteriormente fueron de utilidad para aplicar las técnicas de minería de texto, cuyos resultados se describen en las siguientes secciones.

C. Identificación de patrones

Previo a la identificación de patrones, se obtuvo un top 10 de palabras frecuentes para seleccionar el par de términos que servirían como palabras bases para la determinación de patrones. Como se esperaba, en congruencia con el estudio en cuestión, los resultados indicaron que los dos sustantivos más frecuentes entre las publicaciones de Twitter eran “covid” y “coronavirus”, siendo mencionados más de 30000 y 10000 veces, respectivamente. Seguido de estos términos, se encontraron otras palabras comunes en el contexto pandémico, las cuales se muestran en la figura 2.

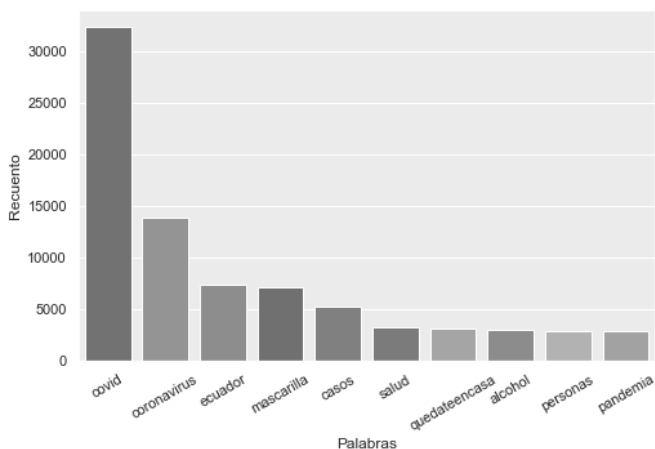


Fig. 2 Top 10 de palabras frecuentes. Elaborado por autores.

Una vez obtenidas las palabras bases “covid” y “coronavirus”, se procedió a establecer los patrones descritos en la metodología, por cada uno de los términos mencionados y, luego, se realizó la búsqueda de tweets que cumplieran con las reglas. De la palabra base “covid” se obtuvieron 314

patrones, de los cuales 296 patrones correspondían al primer patrón y 18, al segundo. Mientras que del término “coronavirus” se obtuvieron 399 patrones, de los cuales 371 correspondían al primer patrón y 28, al segundo.

Dada la cantidad de patrones identificados, mostrar cada uno de ellos resultaría complejo. Por ello, a continuación, se presenta una muestra de los patrones identificados por cada una de las palabras bases

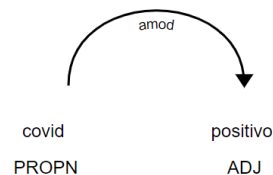


Fig. 3 Muestra del primer patrón asociado a la palabra "covid". Elaborado por autores.

Como se observa en la figura 3, para la palabra “covid”, el primer patrón identificado asocia esta con el adjetivo “positivo”. En primera instancia se denota que, en efecto, se cumple adecuadamente la regla y se conserva el sentido de este resultado con el contexto estudiado, es decir, la pandemia. Esta relación identificada hace referencia a los casos positivos del virus que, hasta ahora, continúan siendo controversiales.

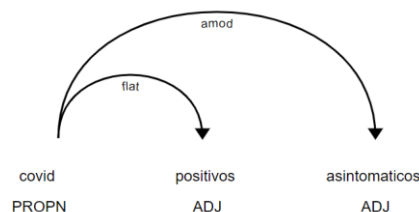


Fig. 4 Muestra del segundo patrón asociado a la palabra "covid". Elaborado por autores.

Tal como se presenta en la figura 4, para el segundo patrón de la palabra “covid”, relaciona esta con los adjetivos “positivos” y “asintomáticos”, haciendo referencia los casos positivos del virus y una de las características de este. Nuevamente, se comprueba que el tweet asociado al patrón cumple adecuadamente con sus especificaciones y mantiene una relación coherente entre estas.

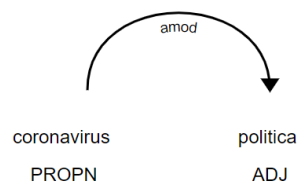


Fig. 5 Muestra del primer patrón asociado a la palabra "coronavirus". Elaborado por autores.

La figura 5 representa la relación identificada entre la palabra base “coronavirus” y el adjetivo “política” en

respuesta al primer patrón. Dicho adjetivo indica que la relación debe interpretarse desde una perspectiva política, lo cual es congruente con uno de los sucesos que intensificó el impacto negativo de la pandemia en Ecuador durante el periodo de estudio y esta situación fue la gestión de la emergencia sanitaria por parte de las autoridades mediante políticas y normas.

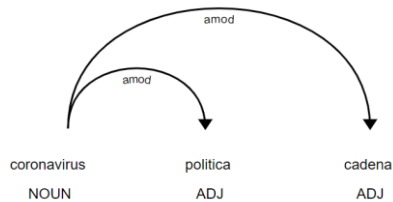


Fig. 6 Muestra del segundo patrón asociado a la palabra "coronavirus".
Elaborado por autores.

El segundo patrón se muestra en la figura 6 y complementa la comprensión del primer patrón, haciendo referencia a las cadenas políticas que se intensificaron por ese tiempo debido a algunos casos de corrupción asociados a la distribución de mascarillas KN95 y otros insumos. De manera que el patrón es consistente con el tema en cuestión.

Por lo tanto, este procedimiento permitió verificar que los tweets sean relevantes y así conservar aquellos que realmente aportaban información a las campañas de insumos médicos. Los tweets verificados y conservados con este procedimiento fueron alrededor de 131000, confirmando la pertinencia de los tweets recopilados mediante web scraping.

D. Recuperación de información

De la base de 131098 tweets obtenidos, se eliminaron datos atípicos que no pertenecían al territorio ecuatoriano empleando la técnica de recuperación de información. Para ello, se utilizó el código del tweet junto a una condición que consistía en que, si el tweet pertenecía a una provincia del Ecuador, este era considerado un dato válido, caso contrario era eliminado. Con esto se obtuvo una base de datos completos y validos de 112578 tweets. Por último, a partir de las palabras claves se comprobó que cada uno de los tweets estaban relacionados a la gestión de insumos médicos, conservando información relevante.

E. Análisis de sentimientos

Los resultados que ha otorgado el análisis de sentimiento determinan que se han analizado un total de 112578 tweets, de los cuales 8825 son positivos y corresponden el 7.84% del total de la muestra, 6297 son neutrales y corresponden el 5.59% de la muestra para lo cual se consideraron los valores que oscilan con un margen ± 0.05 , y finalmente 97456 son negativos y representan el 86.57%.

Por otro lado, en la Tabla 2 se puede apreciar la cantidad total por provincia de los tweets considerados como negativos, positivos y neutrales. La mayor cantidad de tweets negativos

proviene de Guayas (21%) seguidos de Imbabura (17%), Azuay (16%), Santa Elena (12%) y Cotopaxi (9,4%), lo cual está relacionado con el número total de tweets, ya que justamente son las provincias de las cuales se registró un mayor número de tweets durante el tiempo considerado en el estudio.

TABLA II
RESULTADO DEL ANÁLISIS DE SENTIMIENTO

Provincias	#Tweets	# Positivos	#Neutrales	#Negativos
Azuay	17513	1386	961	15166
Bolivar	676	49	37	590
Cañar	2171	107	86	1978
Carchi	12	1	1	10
Cotopaxi	10459	716	558	9185
Chimborazo	434	25	20	389
El oro	2213	189	109	1915
Esmeralda	1376	117	62	1197
Guayas	24349	2056	1495	20798
Imbabura	19072	1319	1025	16728
Loja	994	76	40	878
Los Ríos	2328	187	140	2001
Manabí	1856	174	116	1566
Morona Santiago	330	30	18	282
Napo	9571	662	516	8393
Pastaza	154	35	7	112
Tungurahua	1100	109	50	941
Pichincha	2470	189	144	2137
Zamora Chinchipe	1359	108	69	1182
Galapagos	133	14	7	112
Sucumbios	291	23	12	256
Orellana	151	37	6	108
Santo Domingo de los Tsachilas	117	8	10	99
Santa Elena	13403	1203	800	11400
Zona no delimitada	46	5	8	33

Elaborado por autores.

De igual manera, se muestra que Guayas (23%), Imbabura (16%), Azuay (15%), Santa Elena (14%) y Cotopaxi (8%) encabezan las listas de las provincias que más emitieron tweets catalogados como positivos durante el período de estudio del proyecto, lo cual también es congruente con la cantidad de tweets que provienen de cada una de las provincias mencionadas.

A través de este análisis descriptivo se determina que la mayor cantidad de tweets fueron negativos y que las provincias que más emitieron tweets tanto positivos como negativos son Guayas, Imbabura, Azuay, Santa Elena y Cotopaxi, las cual conforman aproximadamente el 75% de la población estudiada.

F. Identificación de la efectividad de las campañas de insumos médicos

Como se mencionó, mediante esta técnica se relacionaron tres variables o características: los sentimientos o emociones, los likes o corazones y la cantidad de retweets. Estas variables, en otras palabras, representan indicadores que en conjunto posibilitan la comprensión de la efectividad de las campañas de insumos [24]. La dispersión de los tweets con respecto a dichas variables se muestra en la figura 7.

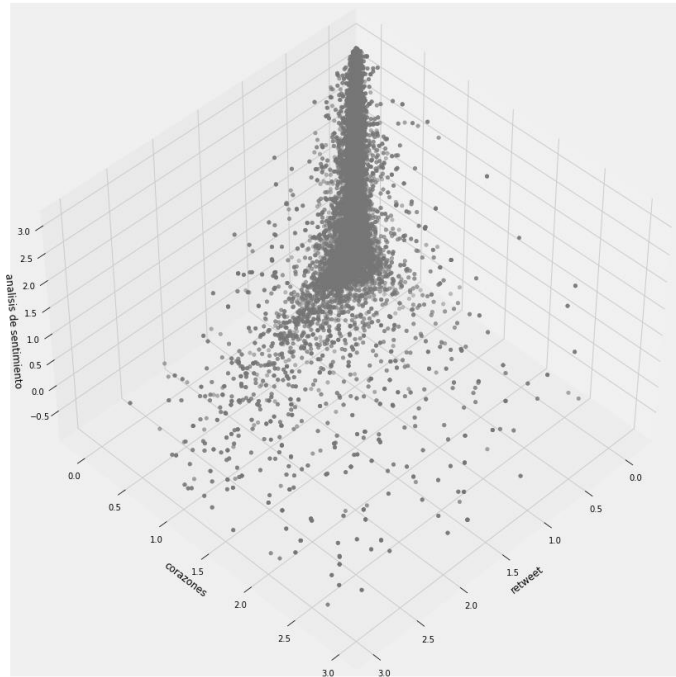


Fig. 7 Dispersión de los tweets con base al análisis de sentimiento, corazones y retweets. Elaborado por autores.

Los resultados del método del codo aplicado para obtener el número K apropiado de grupos a formar, mostraron que este era 2. Como se observa en la figura 8, la curva se aplanaba significativamente en el cluster 2, lo que se traduce en el número de grupos [25].

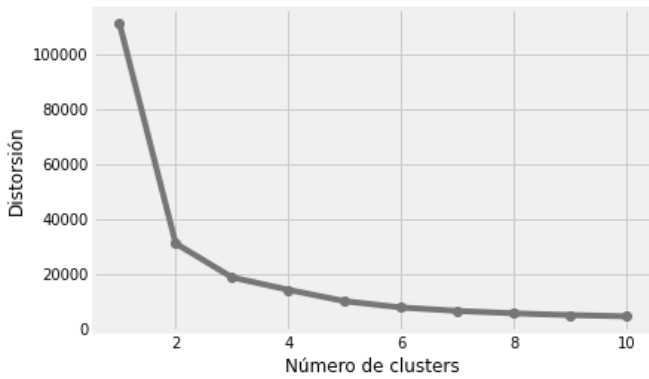


Fig. 8 Número de clusters mediante el método del codo. Elaborado por autores.

Como resultado del método del codo, se identificaron 2 grupos claramente diferenciados, como se observa en la figura 9. La composición de los grupos obtenidos a partir del K-medias muestra que el primer cluster contenía en promedio los sentimientos negativos, ya que el valor es cercano a cero. No obstante, su capacidad de respuesta era mayor, dado que el valor de los likes o corazones y los retweets era superior al segundo grupo. Mientras que el segundo grupo se componía

de sentimientos neutrales, debido a que su valor era cercano a 0.5 y poseía una menor capacidad de respuesta, reflejada en su bajo promedio en likes y retweets.

En otras palabras, el primer grupo representaba las “campañas efectivas”, dado que contenían los sentimientos más negativos y la mayor capacidad de respuesta por parte de la audiencia; lo cual es congruente considerando que los tweets que exponían necesidades de insumos poseían mayor atención. A diferencia del segundo grupo cuyos tweets referían a normas o resoluciones por parte de las autoridades y, por ende, tenían menor atención, ya que no se relacionaban con la necesidad de insumos. Por ello, el segundo grupo representaba las “campañas no efectivas”. En la tabla 3 se detalla la composición por grupo.

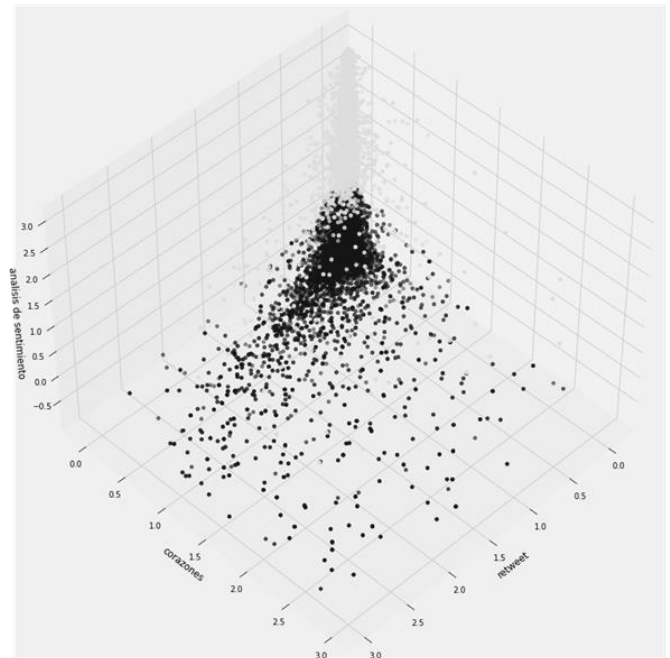


Fig. 9 Identificación de campañas de insumo efectivas y no efectivas. Elaborado por autores.

TABLA III
COMPOSICIÓN DE LOS GRUPOS IDENTIFICADOS

Cluster	Tipo	Retweet	Corazones o Likes	Análisis de Sentimiento
1	Efectivo	2.437176	5.254753	0.072025
2	No efectivo	1.917113	5.160046	0.485975

Elaborado por autores.

A. Validación

De acuerdo con la segmentación de información mediante el clustering de los datos catalogados como relevantes, y a su vez, agrupados por palabras se obtuvo una síntesis de texto o “text summarization”, la cual permitió validar la información condensada, es decir verificar que los tweets mantenían el mismo sentimiento.

Por un lado, se tiene que de los 112578 tweets recopilados la mayor parte fueron los negativos, los cuales corresponden a un valor total de 97456, mostrando así que son los más relevantes y que cuya información no se pierde, sino que más bien reflejaba emociones negativas. Mientras que los tweets positivos y neutrales no fueron tan significativos ya que del total a penas 8825 y 6297 respectivamente, corresponden a dichos sentimientos.

G. Identificación de patrones

1) Tiempos críticos

Los resultados de la serie de tiempo sobre la evolución de los tweets catalogados como negativos muestran que en el mes de marzo se registró la cifra más elevada, lo cual coincide con la declaración de emergencia sanitaria del 12 de marzo por parte del gobierno ecuatoriano. A partir de esa fecha los tweets negativos fueron decreciendo hasta llegar a julio, en donde se evidencia un punto crítico o de crecimiento; dado que en ese tiempo también se establecieron restricciones a la movilidad. Así mismo, el sentimiento negativo reflejado en los tweets muestra un crecimiento para diciembre, alcanzando un máximo relativo en enero de 2021 y decreciendo nuevamente para los siguientes meses. Cabe mencionar que el gobierno anunció nuevamente estado de excepción junto con toque de queda a finales de diciembre con una duración de 30 días, lo cual coincide con la evolución de los tweets negativos en esos meses.

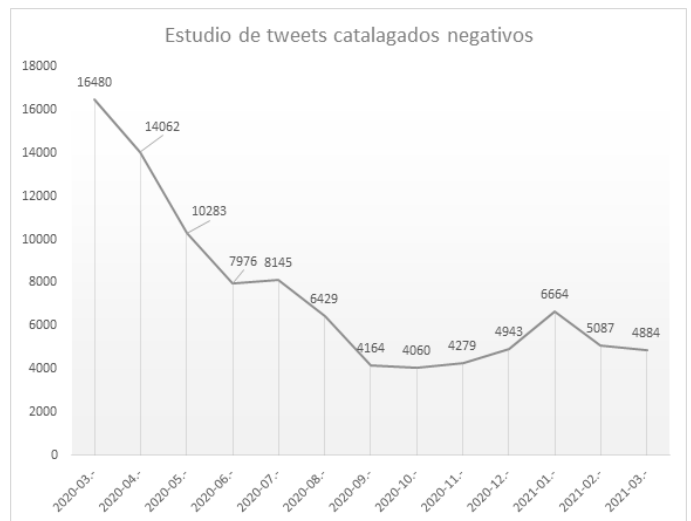


Fig. 10 Serie en el tiempo de tweets catalogados negativos. Elaborado por autores.

Para validar los resultados, estos se compararon con los casos positivos de coronavirus en el país registrados semanalmente desde el 13 de marzo del 2020 hasta el 31 de marzo del 2021, los cuales se ilustran en la figura 11. Como se observa, desde la semana 11 a la 15 del 2020, se reportó un incremento significativo de los casos positivos de covid-19, correspondiendo a los meses de marzo y abril. Así mismo, de la semana 26 a la 31 del mismo año, se mantuvo una curva de crecimiento de los casos positivos de coronavirus, correspondiendo a los meses de julio y agosto.

Por último, durante las tres primeras semanas del 2021, se evidenció un crecimiento constante de coronavirus, correspondientes a enero de ese año. Por lo tanto, se confirma la consistencia de los tiempos críticos obtenidos a partir de los tweets negativos que coinciden con los casos de coronavirus reportados por el COE nacional [26].

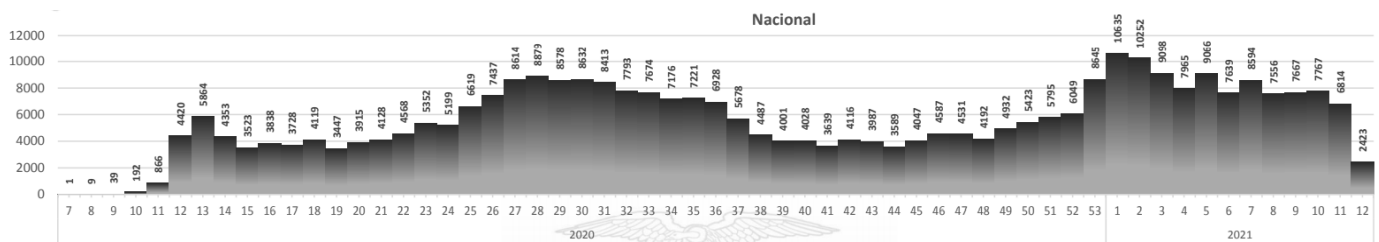


Fig. 11 Curva epidemiológica de casos positivos de covid-19 acumulados por semana epidemiológica. Autor: COE Nacional [26].

2) Zonas críticas

Los resultados de la efectividad de las campañas de insumos en el país por provincia se ilustran en la figura 12. Como se puede observar, las provincias con campañas no efectivas están representadas con un color morado oscuro el cual va tornándose más claro a medida que aumenta la cantidad de campañas efectivas.

Se evidencia entonces que las provincias en donde se concentraron las campañas efectivas fueron Guayas, Imbabura, Azuay y Santa Elena. Es decir, en estas zonas hubo una mayor urgencia o necesidad de insumos médicos, por lo que la ciudadanía optó por gestionar tales recursos a través de twitter a fin de conseguirlos. Dada la crisis que se reflejaba en los tweets, estos alcanzaron una mayor capacidad de respuesta y, por ende, resultaron efectivos.

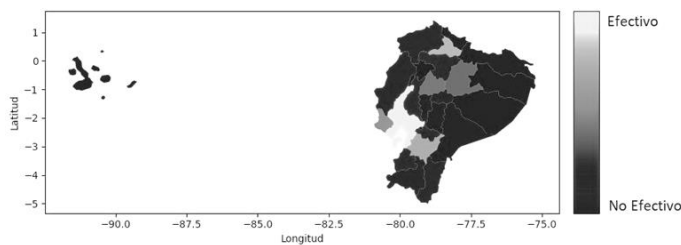


Fig. 12 Identificación de la efectividad de las campañas de insumos en Ecuador por provincia. Elaborado por autores.

Los resultados fueron validados con la información de casos positivos de coronavirus por provincia reportados por el COE nacional [26], los cuales se ilustran en la figura 13. De esto se evidencia una importante coincidencia entre los resultados de las campañas efectivas y las zonas críticas reportadas por el COE, siendo Guayas, Azuay, Imbabura y Santa Elena las provincias que reportaron más del 1.2% de casos confirmados de coronavirus acumulados por provincia referente al total nacional. Es decir, debido a que estas zonas eran las más críticas, hubo mayor movilidad por parte de las personas para realizar campañas de insumos a través de Twitter.

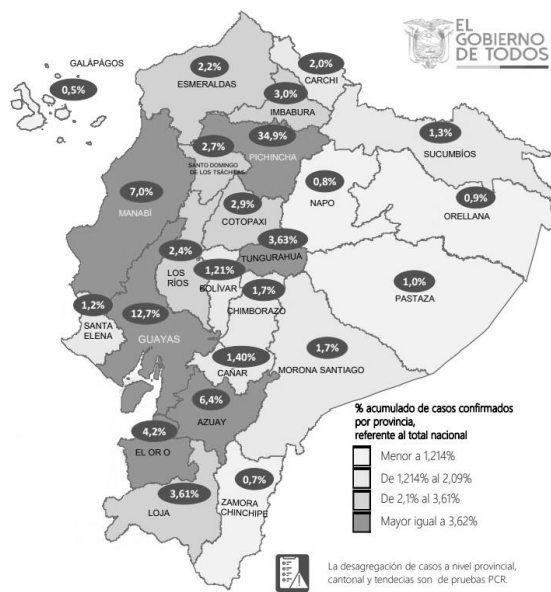


Fig. 13 Casos confirmados de coronavirus en Ecuador por provincia. Autor: COE Nacional [26].

IV. DISCUSIÓN Y CONCLUSIONES

En el presente estudio se recopiló información sobre el tratamiento de la covid-19, mediante entrevistas a expertos, para identificar los insumos médicos más requeridos. A partir de ello se obtuvo una base de 15 palabras, entre las cuales se destacaron recursos como: mascarilla, nitaxozanida, paracetamol, oxígeno, entre otras.

La identificación de insumos más demandados sirvió de base para la aplicación de web scraping en Twitter, obteniendo como resultado alrededor de 131000 tweets, 556151 retweets y 1276567 likes o corazones, desde el 11 de marzo del 2020 hasta el 31 de marzo del 2021. Con esta información se creó una base de datos que facilitó la ejecución de técnicas de preprocesamiento y minería de texto.

Como resultado de la aplicación de técnicas de preprocesamiento y minería de texto se obtuvieron 112578 tweets completos y válidos, que conservaban información relevante referente a la gestión de insumos llevada a cabo por la ciudadanía a través de esta plataforma. Así mismo, se pudo identificar las emociones de los ciudadanos, siendo mayormente negativas debido a la emergencia sanitaria y a la necesidad de conseguir recursos médicos. A su vez, se evidenciaron dos grupos de campañas claramente diferentes y estas correspondían a las campañas efectivas y no efectivas.

En resumen, se identificó una importante coincidencia con los informes proporcionados por el COE nacional, siendo marzo y julio del 2020 y enero del 2021 los meses más críticos, especialmente en zonas como Guayas, Imbabura, Azuay y Santa Elena. Estos resultados representan información de interés para la formulación de políticas públicas relacionadas al abastecimiento de insumos y atención médica.

Por lo tanto, esta investigación presenta un marco metodológico de utilidad para el manejo de la información a través de plataformas modernas de comunicación, como Twitter; el cual puede ser aplicado tanto en situaciones de emergencia sanitaria como en otro tipo de circunstancias de riesgo, a corto y largo plazo.

Cabe destacar que este trabajo es el primer estudio en Ecuador enfocado en la evaluación de la gestión de insumos médicos a través de Twitter, posibilitando una visión completa de la realidad ecuatoriana en la lucha contra el coronavirus. A su vez, sus hallazgos mantienen una relación directa con la mejora del sistema de salud, contribuyendo así la salud en general, siendo esta el tercer objetivo de desarrollo sostenible para la agenda 2030.

IV. TRABAJOS FUTUROS

Considerando las limitaciones de este estudio asociadas al tiempo y disponibilidad de recursos, se recomienda ampliar la muestra de entrevistados enfatizando en su diversidad geográfica. Es decir, contar con la experiencia de profesionales de la salud provenientes de distintas partes del país, podría mejorar la caracterización de las zonas a fin de poseer una perspectiva más completa de la realidad de estas durante la pandemia.

Así mismo, como trabajo futuro se espera ampliar el estudio a nivel latinoamericano a fin de comprender las similitudes o diferencias en las medidas de acción llevadas a cabo por los habitantes de cada país de esta región, frente a la afectación en la gestión de insumos médicos y la atención hospitalaria. Además de comprender las ventajas y desventajas

de las decisiones establecidas por las autoridades locales e internacionales como respuesta a la pandemia del coronavirus.

Por último, se espera definir un modelo que permita establecer los factores socioeconómicos que inciden en la efectividad de la gestión de insumos médicos; ya que se ha demostrado que incluso en provincias altamente urbanizadas como el Guayas, se evidenció una escasez de insumos. Es decir, una zona con facilidades de acceso a insumos en condiciones normales no implica una adecuada gestión de estos en momentos de crisis. Conocer esta relación a través de un modelo, podría mejorar el establecimiento de políticas de gestión.

REFERENCIAS

- [1] D. F. López-Cevallos, C. Chi, y F. Ortega, "Consideraciones para la transformación del sistema de salud del Ecuador desde una perspectiva de equidad", *Rev. Salud Pública*, vol. 16, núm. 3, pp. 346–359, jun. 2014, doi: 10.15446/rsap.v16n3.34610.
- [2] G. P. Mena-Ribadeneira, R. Cañizares-Fuentes, y G. Barquet-Abi-hanna, "Análisis del Sistema de Salud del Ecuador", *Medicina (Mex.)*, vol. 19, núm. 4, pp. 193–204, feb. 2019, doi: 10.23878/medicina.v19i4.1080.
- [3] C. Campos y C. J., "Evolución del sistema de salud de Ecuador: Buenas prácticas y desafíos en su construcción en la última década 2005-2014", *An. Fac. Med.*, vol. 78, núm. 4, pp. 452–460, oct. 2017, doi: 10.15381/anales.v78i4.14270.
- [4] D. Borja y V. Buitrón, "Sí, La Normalidad Es El Problema: Inequidad, Exclusión Y Fuerza Estatal En La Crisis De La Covid-19 En Guayaquil", *J. Lat. Am. Geogr.*, vol. 19, núm. 3, pp. 224–233, 2020, doi: 10.1353/lag.2020.0082.
- [5] D. Herrera, C. Troya Altamirano, y D. Gaus, "COVID-19 in Ecuador: Imported Control Strategies without Context in a Challenged Healthcare System", *Am. J. Trop. Med. Hyg.*, vol. 104, núm. 2, pp. 414–415, feb. 2021, doi: 10.4269/ajtmh.20-1347.
- [6] E. Ortiz-Prado, R. Fernandez-Naranjo, Y. Torres-Berru, R. Lowe, y I. Torres, "Exceptional Prices of Medical and Other Supplies during the COVID-19 Pandemic in Ecuador", *Am. J. Trop. Med. Hyg.*, vol. 105, núm. 1, pp. 81–87, jul. 2021, doi: 10.4269/ajtmh.21-0221.
- [7] A. Wahbeh, T. Nasrallah, M. Al-Ramahi, y O. El-Gayar, "Mining Physicians' Opinions on Social Media to Obtain Insights Into COVID-19: Mixed Methods Analysis", *JMIR Public Health Surveill.*, vol. 6, núm. 2, p. e19276, jun. 2020, doi: 10.2196/19276.
- [8] N. Pérez, "Ecuador, tercero en Latinoamérica en casos positivos de coronavirus", *El Universo*, el 11 de marzo de 2020. <https://www.eluniverso.com/noticias/2020/03/11/nota/7776414/ecuador-tercero-latinoamerica-casos-positivos-coronavirus> (consultado el 10 de junio de 2022).
- [9] C. Andramuño, "Acuerdo N° 00126 - 2020", Ministerio de Salud Pública, Quito, 2020. [En línea]. Disponible en: https://www.salud.gob.ec/wp-content/uploads/2020/03/SRO160_2020_03_12.pdf
- [10] A. Acosta, "El coronavirus en los tiempos del Ecuador", *Análisis Carol.*, abr. 2020, doi: 10.33960/AC_23.2020.
- [11] Organización de las Naciones Unidas, "Plan de Respuesta Humanitaria COVID-19 Ecuador", Equipo Humanitario de País. ONU., 2020. [En línea]. Disponible en: <https://reliefweb.int/sites/reliefweb.int/files/resources/20200430-EHP-ECUADOR-COVID-19.pdf>
- [12] K. Garcia y L. Berton, "Topic detection and sentiment analysis in Twitter content related to COVID-19 from Brazil and the USA", *Appl. Soft Comput.*, vol. 101, p. 107057, mar. 2021, doi: 10.1016/j.asoc.2020.107057.
- [13] R. Khan, P. Shrivastava, A. Kapoor, y A. Mittal, "SOCIAL MEDIA ANALYSIS WITH AI: SENTIMENT ANALYSIS TECHNIQUES FOR THE ANALYSIS OF TWITTER COVID-19 DATA", *J. Crit. Rev.*, vol. 7, p. 2020, ago. 2020.
- [14] M. O. Lwin *et al.*, "Global Sentiments Surrounding the COVID-19 Pandemic on Twitter: Analysis of Twitter Trends", *JMIR Public Health Surveill.*, vol. 6, núm. 2, p. e19447, may 2020, doi: 10.2196/19447.
- [15] T. K. Mackey *et al.*, "Big Data, Natural Language Processing, and Deep Learning to Detect and Characterize Illicit COVID-19 Product Sales: Infection Study on Twitter and Instagram", *JMIR Public Health Surveill.*, vol. 6, núm. 3, p. e20794, ago. 2020, doi: 10.2196/20794.
- [16] J. Marín, "Ecuador Geo-collection". el 27 de julio de 2021. Consultado: el 10 de junio de 2022. [En línea]. Disponible en: https://github.com/jpmarindiaz/geo-collection/blob/ea8b34178317899bdf3be523324347c2817521c/ecu/ecuador.geojson?short_path=9b67de6
- [17] D. Jurafsky y J. Martin, "Information Extraction", en *Speech and Language Processing*, 3rd ed. New Jersey: Prentice Hall. [En línea]. Disponible en: <https://web.stanford.edu/~jurafsky/slp3/17.pdf>
- [18] S. Ahmad y R. Varma, "Information extraction from text messages using data mining techniques", *Malaya J. Mat.*, vol. S, pp. 26–29, ene. 2018, doi: 10.26637/MJMOS01/05.
- [19] IBM, "What is Text Mining?", 2020. <https://www.ibm.com/cloud/learn/text-mining> (consultado el 10 de junio de 2022).
- [20] A. Rosen, "Hoy Twitrear es más fácil", 2017. https://blog.twitter.com/es_la/topics/product/2017/Hoy-es-mas-facil-Twitrear (consultado el 10 de junio de 2022).
- [21] K. S. Adewole, T. Han, W. Wu, H. Song, y A. K. Sangaiah, "Twitter spam account detection based on clustering and classification methods", *J. Supercomput.*, vol. 76, núm. 7, pp. 4802–4837, jul. 2020, doi: 10.1007/s11227-018-2641-x.
- [22] S. Ahuja y G. Dubey, "Clustering and sentiment analysis on Twitter data", en *2017 2nd International Conference on Telecommunication and Networks (TEL-NET)*, ago. 2017, pp. 1–5. doi: 10.1109/TEL-NET.2017.8343568.
- [23] E. Loper y S. Bird, "NLTK: The Natural Language Toolkit", University of Pennsylvania, Pennsylvania, 2002.
- [24] S. Agarwal y M. Damle, "SENTIMENT ANALYSIS TO EVALUATE INFLUENCER MARKETING: EXPLORING TO IDENTIFY THE PARAMETERS OF INFLUENCE", *PalArchs J. Archaeol. Egypt Egyptol.*, vol. 17, núm. 6, Art. núm. 6, dic. 2020.
- [25] R. Tibshirani, G. Walther, y T. Hastie, "Estimating the number of clusters in a data set via the gap statistic", *J. R. Stat. Soc. Ser. B Stat. Methodol.*, vol. 63, núm. 2, pp. 411–423, 2001, doi: 10.1111/1467-9868.00293.
- [26] Sistema Nacional Descentralizado de Gestión de Riesgos, "Informes de Situación del Coronavirus en Ecuador. Servicio Nacional de Gestión de Riesgos y Emergencias (SNGRE)", Sistema Nacional Descentralizado de Gestión de Riesgos, Quito, 2020. [En línea]. Disponible en: <https://www.gestionderiesgos.gob.ec/wp-content/uploads/2021/01/INFOGRAFIA-NACIONALCOVID19-COE-NACIONAL-Y-ANEXO-08h00-31012021.pdf>