

Minería de datos: análisis de sentimiento en Twitter basado en lexicones sobre el uso de dióxido de cloro para el tratamiento del COVID-19

Enrique Edgardo Condor-Tinoco

enricoti@unajma.edu.pe / Universidad Nacional José María Arguedas

José Antonio Rojas-Cusi

joseantoniorojas999@gmail.com / Universidad Nacional José María Arguedas

Anthony Zevallos-Rodríguez

anthony.elzr7@gmail.com / Universidad Nacional José María Arguedas

Carlos Yinmel Castro-Buleje

ccastro@unajma.edu.pe / Universidad Nacional José María Arguedas

Recepción: 1/9/2020 Aceptación: 23/10/2020

RESUMEN. El COVID-19, según la Organización Mundial de Salud (OMS), es la “enfermedad infecciosa causada por el coronavirus que se ha descubierto más recientemente”. Para el tratamiento del COVID-19 se han informado por diferentes medios diversos productos, entre ellos se tiene al dióxido de cloro; se dice que puede curar y prevenir el COVID-19 porque incrementa los niveles de oxígeno en la sangre; también expertos en la salud indicaron que es falso y, por el contrario, esta sustancia puede oxidar la hemoglobina y desencadenar más problemas de salud. Se planteó el objetivo de desarrollar un modelo de minería de datos sobre el análisis de sentimiento en Twitter basado en lexicones sobre el uso de dióxido de cloro para el tratamiento del COVID-19, para su desarrollo se usó la metodología CRISP-DM y la herramienta Orange Canvas. En los resultados se obtuvo que la palabra *dióxido de cloro* está asociada a las palabras *sí*, *tomar* y *COVID*; por otro lado, la opinión de las personas en relación al dióxido de cloro, con tendencia a negatividad y neutralidad, dan como resultado calificaciones compuestas, lo que significa que 4500 tuits de los usuarios de Twitter tienen opiniones de neutralidad y rechazo en distintas proporciones sobre el tratamiento del dióxido de cloro para el COVID-19.

PALABRAS CLAVE: análisis de sentimiento / dióxido de cloro / Twitter / COVID-19

Data Mining: Lexicon-Based Sentiment Analysis on Twitter Concerning the Use of Chlorine Dioxide for COVID-19 Treatment

ABSTRACT. According to the World Health Organization (WHO), COVID-19 “is an infectious disease caused by a newly discovered coronavirus.” Different media have recommended diverse treatments for COVID-19 including chlorine dioxide, which is said to cure and prevent COVID-19 because it increases oxygen levels in the blood. Health experts declared that such information was false and, on the contrary, it could cause hemoglobin oxidation and trigger more health problems. This research aimed to develop a data mining model on a lexicon-based sentiment analysis on Twitter concerning the use of chlorine dioxide for COVID-19 treatment, using the CRISP-DM methodology and the Orange Canvas tool. The results showed that the word chlorine dioxide is associated with the words yes, take and COVID. On the other hand, people’s opinion regarding chlorine dioxide tended to be negative and neutral, and resulted in mixed ratings. This means that the 4,500 tweets analyzed in the research either accepted or rejected chlorine dioxide for COVID-19 treatment in different proportions.

KEYWORDS: sentiment analysis / chlorine dioxide / Twitter / COVID-19

1. INTRODUCCIÓN

Twitter fue fundado por Jack Dorsey; es un servicio de microblogueo y de red social, que se encuentra continuamente alimentado por millones de personas desde diferentes partes del mundo. Es una plataforma que limita sus mensajes a 280 caracteres. Es una plataforma en línea para publicar y leer mensajes de texto de hasta 140 caracteres (Orihuela, 2011). A diferencia de las redes sociales de Facebook y LinkedIn, las relaciones entre los extremos de los usuarios no se ponen en contacto directo; en el caso de Twitter se tiene a seguidores (*followers*) y los seguidos (*followed*), de esta manera es que se tienen dos listas en esta red social: la lista de los que te siguen y la lista de a quienes sigues. Si sigues a alguien, eso quiere decir que verás sus tuits (mensajes que escribe en Twitter) en tu “cronología” (o *timeline*, conjunto ordenado en el tiempo de los mensajes que te llegan de la gente a quienes sigues) (Rubín, 2020).

La presente investigación trata del tema del uso del dióxido de cloro para el tratamiento del COVID-19. El COVID-19, según la Organización Mundial de Salud (OMS), es la “enfermedad infecciosa causada por el coronavirus que se ha descubierto más recientemente; este nuevo virus, así como la enfermedad que provoca, eran desconocidos antes de que estallara el brote en Wuhan (China) en diciembre del 2019” (Organización Mundial de Salud, 2020); asimismo, la OMS ya lo declaró como una pandemia que afecta a muchos países de todo el mundo. Para el tratamiento del COVID-19, debido a la información que se ha publicado en diferentes medios, hay gente que está consumiendo el dióxido de cloro para tratarse; este compuesto químico es un gas de color amarillo-rojizo que se descompone muy rápido en el aire, se usa como blanqueador en las fábricas que producen papel y en las plantas de tratamiento de agua que producen agua potable. El dióxido de cloro forma iones de clorito, que son sustancias muy reactivas; por lo tanto, le permite matar bacterias y otros microorganismos en el agua (Agencia para Sustancias Tóxicas y el Registro de Enfermedades, 2004).

Se ha advertido, en diferentes medios, sobre el uso del dióxido de cloro, ya que se dice que puede curar y prevenir el COVID-19, porque incrementa los niveles de oxígeno en la sangre; del mismo modo, expertos en la salud indicaron que esta información es falsa y, por el contrario, que este puede oxidar la hemoglobina y desencadenar más problemas de salud. En la página web AS Perú se publicó que el Ministerio de Salud del Perú (Minsa) prohíbe la venta de dióxido de cloro para tratar el COVID-19 señalando que “el consumo de dióxido de cloro o clorito de sodio puede causar graves daños a la salud de las personas, pues no se trata de un medicamento, sino de un poderoso agente blanqueador, por lo tanto, no cura ninguna enfermedad” (AS Perú, 2020). En la web de la BBC News se publicó que este producto conocido como “suplemento mineral milagroso” lleva ya varios años publicitándose como un remedio para muchas enfermedades, incluso el cáncer. Sin embargo, ninguna institución sanitaria lo reconoce como medicamento y en la actualidad, con la pandemia del COVID-19 que azota al mundo, vuelve a aparecer como algo capaz de prevenir los efectos del virus (Orgaz, 2020). En la página web de la CNN en Español se publicó que la Cámara de Diputados de Bolivia aprobó

un proyecto de ley que regula de manera excepcional la elaboración, comercialización y uso consentido del dióxido de cloro, esto de manera preventiva y como tratamiento para pacientes con COVID-19 (Trucco, 2020). En la investigación denominada *Eficacia y seguridad del dióxido de cloro para el tratamiento de COVID-19*, los autores llegaron a las siguientes conclusiones: a) En la búsqueda de la literatura, no encontraron ningún estudio que muestre la seguridad y eficacia de esta sustancia para el tratamiento del COVID-19. b) Diversas agencias gubernamentales de nuestra región (Chile, Argentina, Uruguay, Paraguay y Ecuador) recomiendan no consumir esta sustancia para el tratamiento del COVID-19, debido a que no se ha demostrado ningún beneficio terapéutico; por el contrario, ha mostrado eventos adversos graves (Cabezas Sánchez, Solari Zerpa y Caballero Ñopo, 2020).

El tema del uso del dióxido de cloro para el tratamiento del COVID-19 se ha difundido de manera masiva en Twitter y las opiniones son diversas, justamente a causa de las diferentes noticias que informan sobre este producto, tal como se detalla en el párrafo anterior. Con las herramientas que posee el *software* Orange Canvas 3.26 y los datos que se encuentran en esta red social sobre el uso del dióxido de cloro se puede realizar minería de datos (MD), específicamente en el campo del análisis de sentimiento. La minería de datos es un intento de buscarle sentido a la explosión de información que actualmente puede ser almacenada (Mitra y Acharya, 2005).

En función de todo el análisis que se ha realizado en los párrafos anteriores, se plantea el objetivo de desarrollar un modelo de minería de datos sobre el análisis de sentimiento basado en lexicones sobre el uso del dióxido de cloro para el tratamiento del COVID-19, haciendo uso de la metodología CRISP-DM, modelo estándar abierto del proceso, en inglés *Cross Industry Standard Process for Data Mining*, que divide el proceso de minería de datos en seis fases principales, las que son explicadas en la parte de metodología de la presente investigación.

Para cumplir con el objetivo planteado y obtener los resultados, en la introducción se ha descrito el problema, los objetivos y las herramientas y metodologías que se usarán para desarrollar la investigación. En la parte de la metodología, siguiendo los pasos de CRISP-DM se desarrolla el modelo de minería de datos sobre el análisis de sentimiento en Twitter basado en lexicones sobre el uso del dióxido de cloro para el tratamiento del COVID-19; después se presentan los resultados obtenidos, para luego finalizar con las conclusiones de la investigación.

2. METODOLOGÍA

El desarrollo de la investigación se centra en los pasos de la metodología CRISP-DM (comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación y despliegue). La secuencia se sigue según lo descrito en la figura 1, la herramienta usada para el procesamiento de los datos fue Orange Canvas, los datos fueron tomados de la red social Twitter a través del *widget* Twitter que el *software* posee, el tema es sobre el análisis

del sentimiento basado en lexicones sobre el dióxido de cloro como posible tratamiento para el COVID-19.

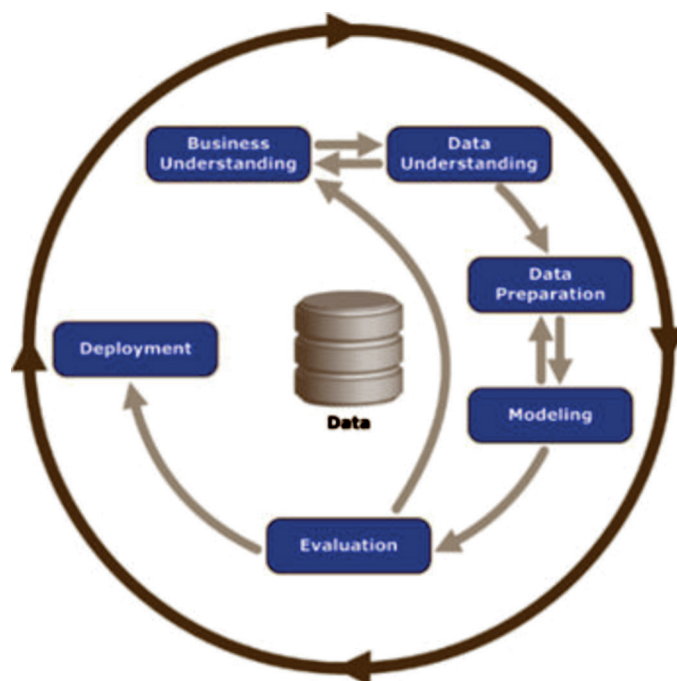










Figura 1. Pasos de la metodología CRISP-DM
Fuente: Vallalta Ru (s. f.)

2.1 Comprensión del negocio

Se comenta en diferentes medios sobre el uso del dióxido de cloro, ya que se dice que puede curar y prevenir el COVID-19, porque incrementa los niveles de oxígeno en la sangre; sin embargo, expertos en la salud indicaron que esta información es falsa. El tema del uso del dióxido de cloro para el tratamiento del COVID-19 se ha difundido de manera masiva en Twitter y las opiniones son diversas, justamente a causa de las diferentes noticias que se ha dicho sobre este producto. Es preciso analizar el sentimiento de los comentarios realizados e identificar lo que piensa la población sobre este producto.

2.2 Comprensión de los datos

Los *widgets* de Orange Canvas usados en el tratamiento de los datos fueron:

- Twitter: permite consultar tuits a través de la API de Twitter. Se puede consultar por contenido y autor o ambos y acumular resultados.  Twitter
- Data Table: recibe uno o más conjuntos de datos en su entrada y los presenta como una hoja de cálculo.  Data Table
- Preprocess Text: divide el texto en unidades más pequeñas (*tokens*), las filtra, ejecuta la normalización (derivación, lematización).  Preprocess Text
- Word Cloud: muestra *tokens* en el corpus, su tamaño indica la frecuencia de la palabra en el corpus o el recuento promedio de una bolsa de palabras.  Word Cloud
- Sentiment Analysis: predice el sentimiento para cada documento en un corpus. Utiliza los módulos de sentimiento de Liu Hu y Vader de NLTK.  Sentiment Analysis
- Corpus Viewer: está diseñado para ver archivos de texto, siempre generará una instancia de corpus.  Corpus Viewer
- Select Columns: el usuario decide qué atributos se utilizarán. Orange distingue entre atributos ordinarios, atributos de clase (opcionales) y metaatributos.  Select Columns
- FreeViz: ayuda a visualizar los resultados: los puntos de la misma clase se atraen entre sí, los de diferente clase se repelen.  FreeViz

2.3 Extracción de los datos

Los datos fueron recopilados directamente de Twitter haciendo uso de su API que está incorporada en el mismo *widget* Twitter. Una vez obtenidos los tuits se seleccionaron las columnas importantes para nuestro análisis, quedando las siguientes con las cuales se trabajó: autor, tuit, descripción del autor y fecha y hora. Para empezar la consulta se deben tener las credenciales de la API de Twitter; de ser el caso que no cuente con estos datos, se debe hacer una solicitud de acceso como desarrollador desde la internet. Para la selección de los tuits se usó la palabra clave *dióxido de cloro* con el fin de obtener mejores resultados en tiempo real, ya que el *widget* Twitter nos permite realizar este proceso. Se usó este criterio para obtener consultas más precisas, ya que esta palabra se encuentra en tendencia en estos tiempos de pandemia debido a su polémica como tratamiento para el COVID-19. De este modo, se estableció analizar 4500 tuits.

title	Author	Content	Date	Language	Location
1	@algotagniaaa	RT @IdeasCongresoCL: @floricatamotuda El clorito de sodio, al activarse con un ácido ligero, se transforma en dióxido de clor...	2020-10-11 ...	es	?
2	@NOVOLV5NLU...	@javierlanari TODAVIA ESTAMOS ESPERANDO SOLIDARIDAD CON LA FAMILIA DEL NIÑO MUERTO POR TOMAR EL DIOXIDO ...	2020-10-11 ...	es	?
3	@EfectoCecuyo	@CecuyoChequea No hay ninguna evidencia médico-científica que demuestre que tomar dióxido de cloro puede curar ...	2020-10-11 ...	es	?
4	@miñidalgo	RT @alfrepared: Las películas norteamericanas de Netflix son un cóctel de inclusión y corrección política tan fuertes que sab...	2020-10-11 ...	es	?
5	@CarilucCarina	@javierlanari TAMBIEN FALLARON LOS MEDIOS POR SU FALTA DE COMPROMISO PORQUE UNA PERSONA AL AIRE TOMO ...	2020-10-11 ...	es	?
6	@wada_3M	@Angelismo2 Allí no va Kilgates, además saben utilizar el Dióxido de Cloro. (Malaria)	2020-10-11 ...	es	?
7	@yymigue	No aguanto más venga el dióxido de cloro https://t.co/ePPKiz5Wv	2020-10-11 ...	es	?
8	@introdagor	@HenryChinask67 Que tomen dióxido de cloro para aclarar la mente	2020-10-11 ...	es	?
9	@ESAArgentina	Coronavirus en Argentina: diputada de Jc recomendó usar dióxido de cloro y generó fuerte repudio Cámara de Diputados, ...	2020-10-11 ...	es	?
10	@cáaveritabike	@pasotti_ En el mio hablan de las maravillas de tomar dióxido de cloro	2020-10-11 ...	es	?
11	@lonaramirezX	RT @anticio2: @bbimbi Desinforma bastante y promueve el dióxido de cloro ...	2020-10-11 ...	es	?
12	@MIMabadosnik...	@eccc_ es Se puede curar cáncer con dióxido de cloro, peróxido de hidrógeno, vitaminas naturales(no sintéticos) y alimentos...	2020-10-11 ...	es	?
13	@EnricBermada...	@carmenvivomivid @crisis_historia @A_Corales_1020 @gemmopp @Random_Dragon @SalvaGarc4 @iPodMusica ...	2020-10-11 ...	es	?
14	@cescept	RT @anticio2: Este es el documento publicado por USA Today donde dan los argumentos por los que pidieron la detención d...	2020-10-11 ...	es	?
15	@Avisiv	Mi papá chingándose dióxido de cloro cada día y le demostré (con fuentes confiables) que te servía para pura verga y que só...	2020-10-11 ...	es	?
16	@ani72_jos	@fflorentantoni No es la única porque también Santiago Cúneo y Lilia Lemoine hacen apología al dióxido de cloro, o sea ha...	2020-10-11 ...	es	?
17	@Aldeadeletras	RT @hambroid: Entrevista en Imagen Querétaro sobre el Dióxido de cloro. Parte 1...	2020-10-11 ...	es	?
18	@caro_wy6i	PANDEMIA...	2020-10-11 ...	es	?
19	@FerFrias	RT @anticio2: Este es el documento publicado por USA Today donde dan los argumentos por los que pidieron la detención d...	2020-10-11 ...	es	?
20	@beta_foly	@cesaralva58 Algo raro. Con Ivermectina, hidrocicloroquina, dióxido de cloro, y mas yerbas, yo he visto vivir gente. Que ...	2020-10-11 ...	es	?
21	@anticio2	@FernandoCarvera escuche el podcast donde hablaban de dióxido de cloro, buen trabajo	2020-10-11 ...	es	?
22	@joselan_mx	@almadelm33: @vivicanosok Vivi querida! Te informo, no se si sabias que la diputada Frade, pidió en una sesion de la ...	2020-10-11 ...	es	?
23	@BetliaAreta	RT @ttnn_podemos: Dr. Alfredo Valencia: "Tomando el dióxido de cloro noté mejoría muy muy pronto, a los 6 o 7 días ya ...	2020-10-11 ...	es	?
24	@joselan_mx	RT @lamber_2009: LA DIPUTADA ARGENTINA MONICA FRADE DEFIENDE EL DIOXIDO DE CLORO https://t.co/Ve9t3F4E7 a ...	2020-10-11 ...	es	?
25	@joselan_mx	RT @JStraneo: El dióxido de cloro está revolucionando a la medicina https://t.co/XG1q7vWVV a través de @metrolibrety	2020-10-11 ...	es	?
26	@RagnarDanne...	RT @JuanGnusmas: Encuesta...	2020-10-11 ...	es	?

Figura 2. Muestra de las columnas obtenidas en la consulta
Elaboración propia en Orange Canvas

3. PREPARACIÓN DE LOS DATOS

En la extracción de los datos, se observa que hay caracteres extraños, los cuales no servirían para el procesamiento y análisis. Para la limpieza de los datos se usa el *widget* Preprocess Text, con el cual se remueven URL, acentos, códigos HTML, convertir a minúsculas, se tokeniza, etcétera, así como también se restringirá los caracteres extraños, con el objetivo de garantizar el trabajo con datos limpios. En un bloc de notas se listan los caracteres o palabras que se desea excluir y deben ser cargados al Preprocess Text como se observa en figura 3.

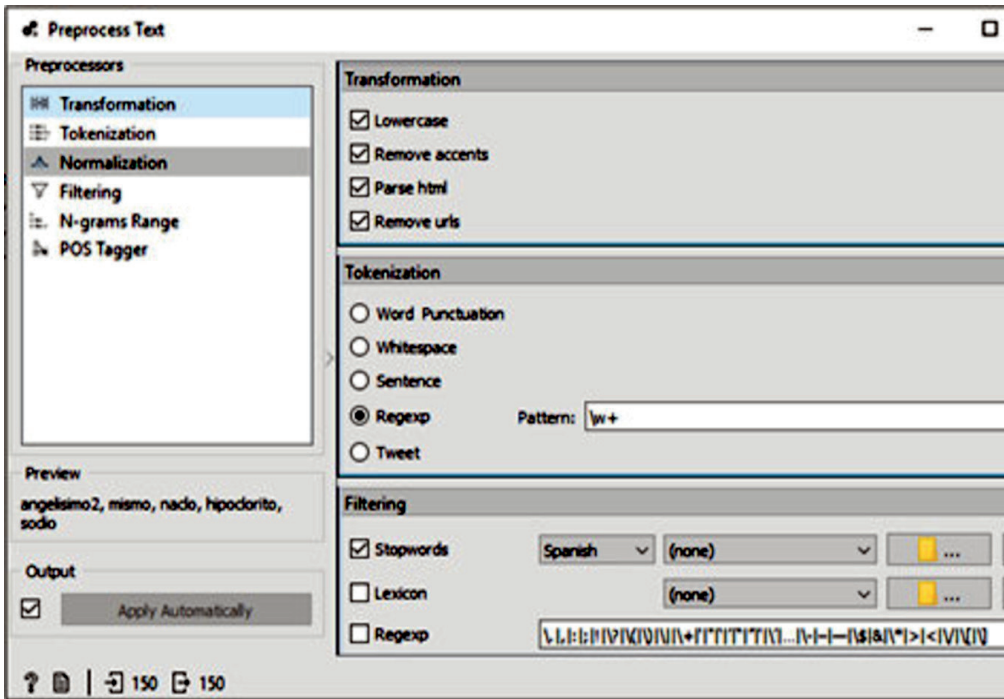


Figura 3. Interfaz de inicio del Preprocess Text
Elaboración propia en Orange Canvas

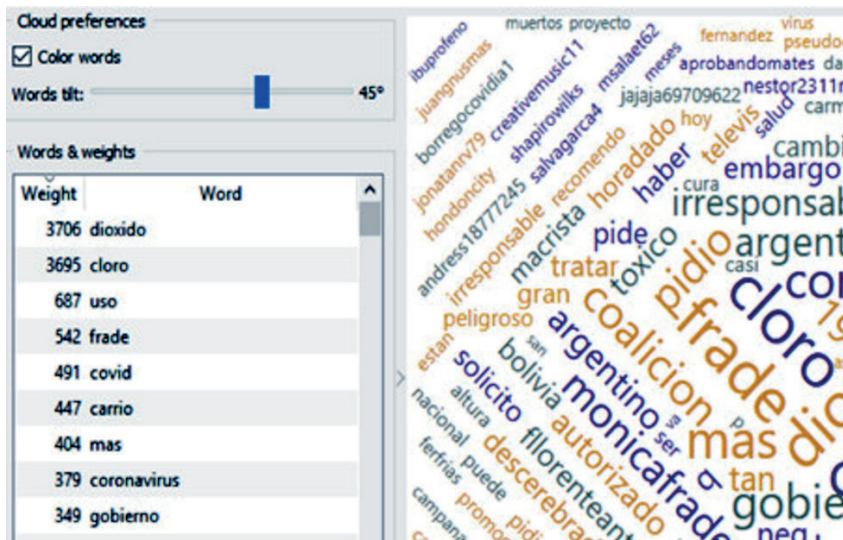


Figura 4. Word Cloud después de la limpieza de palabras
Elaboración propia en Orange Canvas

Después, usando el Word Cloud, a simple vista se pueden observar palabras más grandes que las otras, lo que indica que estas están con más frecuencia; asimismo, se observa en la parte derecha (figura 4) la cantidad de veces que estas palabras se repiten como es el caso de *cloro* (3695 veces), *dióxido* (3706 veces), *uso* (687 veces), *COVID* (497 veces), etcétera.

4. MODELADO

El modelo de minería de datos sobre el análisis de sentimiento basado en el lexicon sobre el uso del dióxido de cloro para el tratamiento del COVID-19 se presenta en la figura 5.

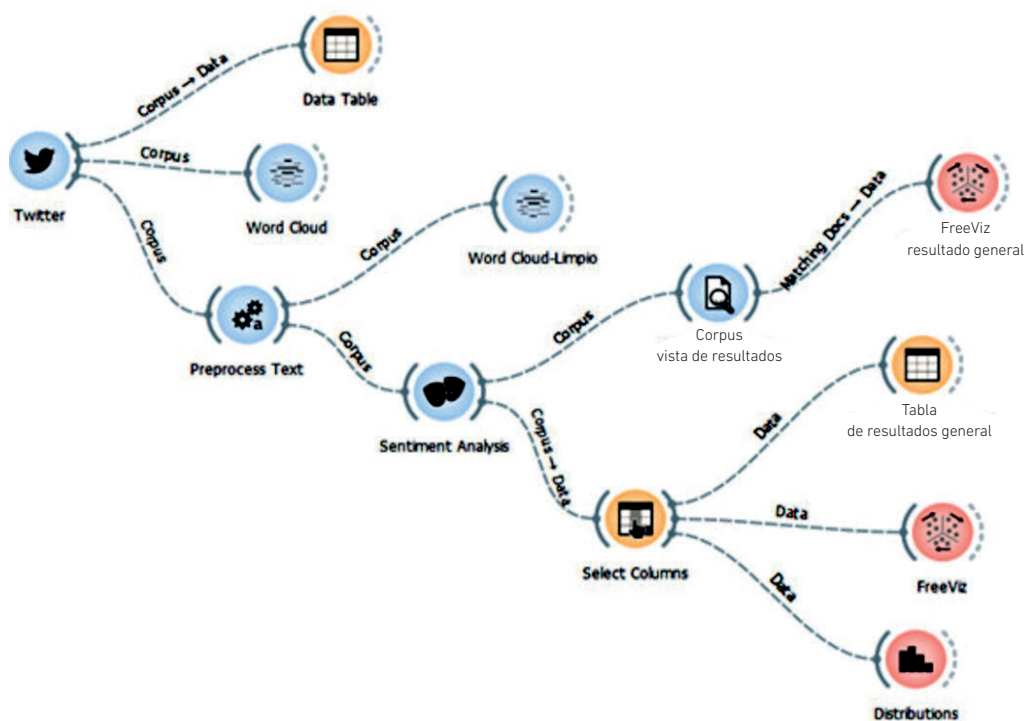


Figura 5. Diseño del modelo de minería de datos sobre el análisis de sentimiento en Twitter, basado en lexicones sobre el uso de dióxido de cloro para el tratamiento del COVID-19

Elaboración propia en Orange Canvas

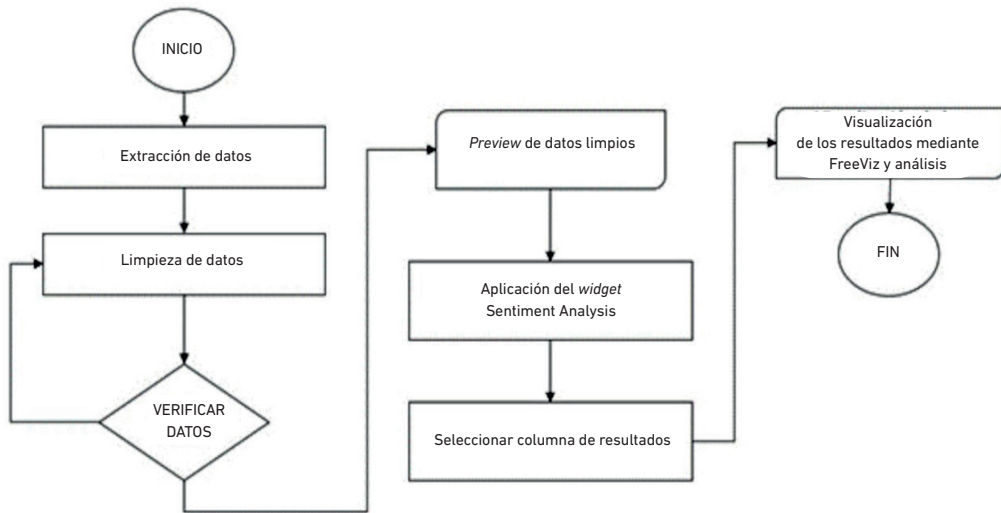


Figura 6. Modelo de minería de datos sobre el análisis de sentimiento en Twitter basado en lexicones sobre el uso de dióxido de cloro para el tratamiento del COVID-19

Elaboración propia

La evaluación y despliegue de la metodología CRISP-DM se realiza en la parte de resultados, según el diseño del modelo que se presenta en la figura 5.

5. RESULTADOS

El *widget* Análisis de sentimientos (Sentiment Analysis) predice el sentimiento para cada documento en un corpus. Utiliza los métodos de sentimiento Liu Hu y VADER de NLTK. Ambos están basados en léxico. Para esta investigación se usó el método VADER análisis de sentimientos basado en léxico y reglas; este módulo nos proporciona cuatro tipos de resultados (positivo, neutro, negativo, compuesto) para cada documento o tuit. Una puntuación compuesta representa el sentimiento total de un tuit, donde -1 es el más negativo y 1 el más positivo.

La puntuación compuesta se calcula sumando las puntuaciones de valencia de cada palabra en el léxico, se ajusta de acuerdo a las reglas y luego se normaliza para estar entre -1 (más extremo negativo) y $+1$ (más extremo positivo). Los valores umbrales típicos son:

- sentimiento positivo: compuesto $\geq 0,05$
- sentimiento neutral: compuesto $> -0,05$ y compuesto $< 0,05$
- sentimiento negativo: compuesto $\leq -0,05$

En cambio, los puntajes positivo, negativo y neutro representan las proporciones de texto que caen en cada categoría, por lo que todos deben sumar 1 o cerca de él con la operación flotante (EGilbert, 2014).

Para la visualización de los resultados mediante un gráfico y la clasificación de estos, utilizamos el *widget* FreeViz.

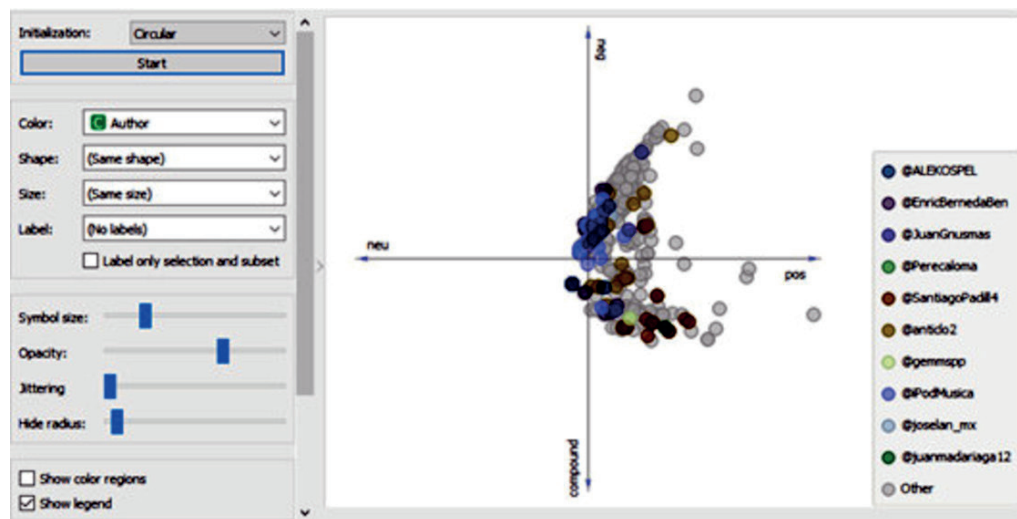


Figura 7. Vista de resultados mediante el widget FreeViz
Elaboración propia en Orange Canvas

Como se podrá observar en la figura 7, se muestra el resultado de todo el análisis de sentimiento en función del autor de cada tuit; de esta manera, se observa de forma general dichos resultados. Para una mejor interpretación y análisis de esta información se analizarán estos resultados en función de los valores (intensidad del tuit) obtenidos de cada tuit como negativos, positivos, neutros y compuestos.

Para ello se procedió a utilizar el *widget* Data Sampler con el cual para la interpretación de resultados se utilizó el total de datos equivalente a 4500 tuits.

5.1 Análisis de tuits con puntuación hacia positivo

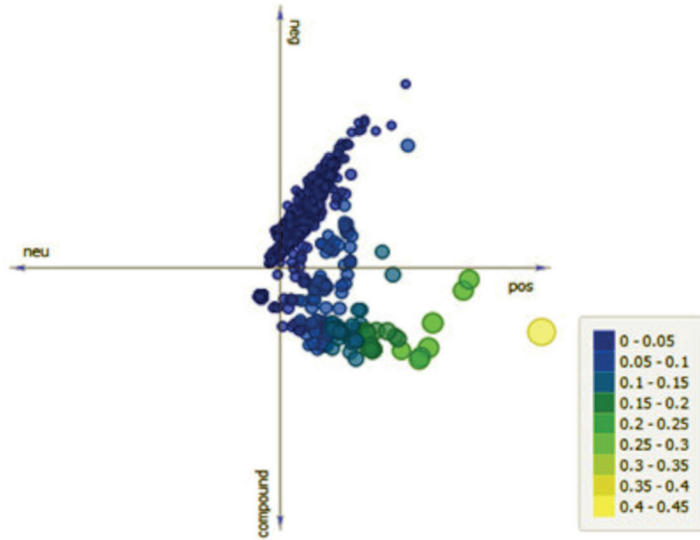


Figura 8. Resultados con puntuación de positividad con FreeViz
Elaboración propia en Orange Canvas

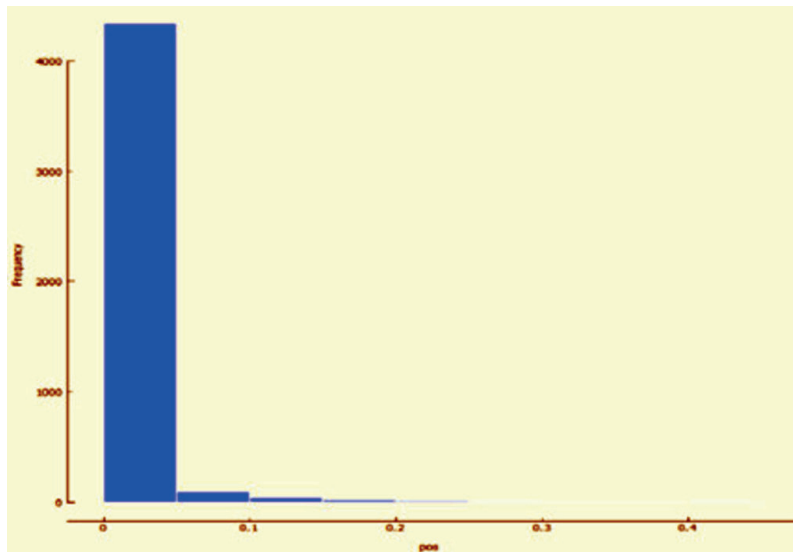


Figura 9. Gráfico de barras de resultados con puntuación hacia positivo
Elaboración propia en Orange Canvas

Como se puede observar en la figura 9, se obtuvo en la primera barra un puntaje entre 0 y 0,05 para 4341 tuits que representan al 96,47 %; de igual manera la segunda barra con un valor entre 0,05 y 0,1 para 91 tuits que representan al 2,02 %; en la tercera barra con un valor entre 0,1 y 0,15 para 40 tuits que representan al 0,89 %; en la cuarta barra con un valor entre 0,15 y 0,2 para 19 tuits que representan al 0,42 %; en la última barra con un valor entre 0,2 y 0,25 para 7 tuits que representan al 0,16 %; del mismo modo, en la figura 8 se pueden observar los resultados clasificados de los tuits antes mencionados con mayor intensidad de positividad con colores más claros como el amarillo y verde. Para este caso, el resultado más intenso a positivo tiene un valor de 0,403. De este primer análisis podemos mencionar que, de los tuits analizados, se obtuvo una intensidad positiva baja considerando que 4341 tuits de los 4500 tienen una puntuación entre los valores 0 y 0,05 que representa el 96,47 % de tuits.

5.2 Análisis de tuits con puntuación hacia neutro

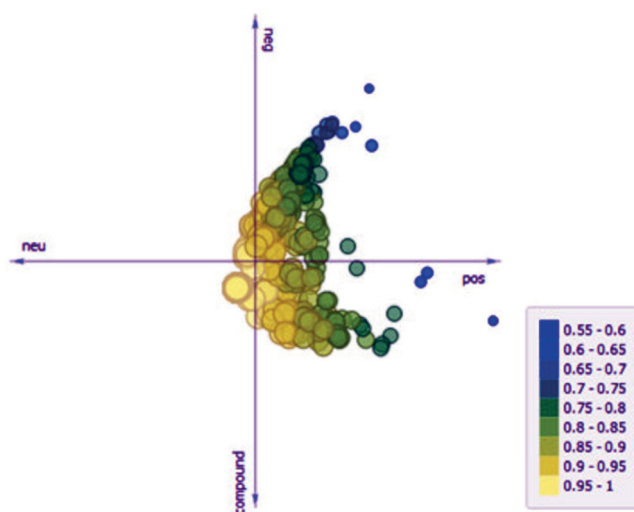


Figura 10. Resultados con puntuación de neutralidad con FreeViz
Elaboración propia en Orange Canvas

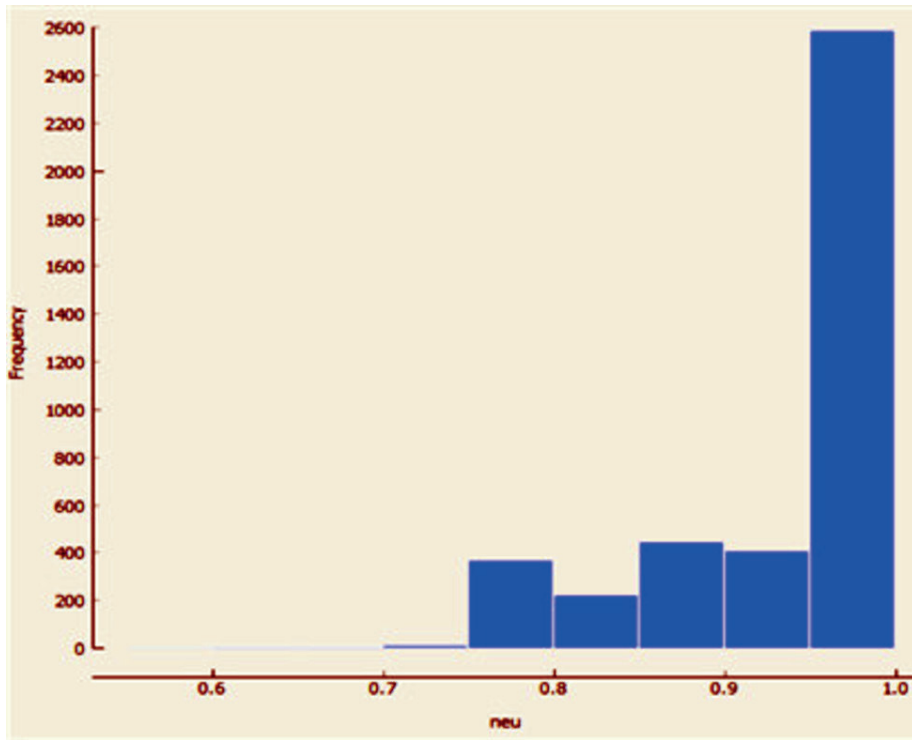


Figura 11. Gráfico de barras de resultados con puntuación hacia neutro
Elaboración propia en Orange Canvas

En la figura 11 se muestra el análisis de la intensidad a neutralidad de los 4500 tuits; se observó que en 15 tuits se tienen valores de entre 0,7 y 0,75, que representan al 0,33 %, así como también 397 tuits con un valor entre 0,75 y 0,8, que representan 8,82 %; asimismo, 252 tuits han sido clasificados entre 0,8 y 0,85, con un porcentaje del 5,60 %, 480 tuits entre 0,85 y 0,9 con un porcentaje de 10,67 %, y en la penúltima barra podemos observar 446 tuits con un porcentaje de 9,91 % con una intensidad de 0,9 y 0,95, y la última barra con 2902 tuits con un valor mayor o igual a 0,95 con un porcentaje de 64,49 %. De la misma manera, en la figura 10 se pueden observar los resultados clasificados de los tuits antes mencionados como neutros, con colores más claros y de color amarillo y verde.

Para nuestro análisis el resultado neutro con mayor intensidad es de 0,95. En este segundo análisis se puede mencionar que se obtuvieron 2902 tuits de los 4500 con una intensidad de sentimiento neutro alto, considerando que 64,49 % se encuentran entre los valores de mayor o igual a 0,95.

5.3. Análisis de tuits con puntuación hacia negatividad

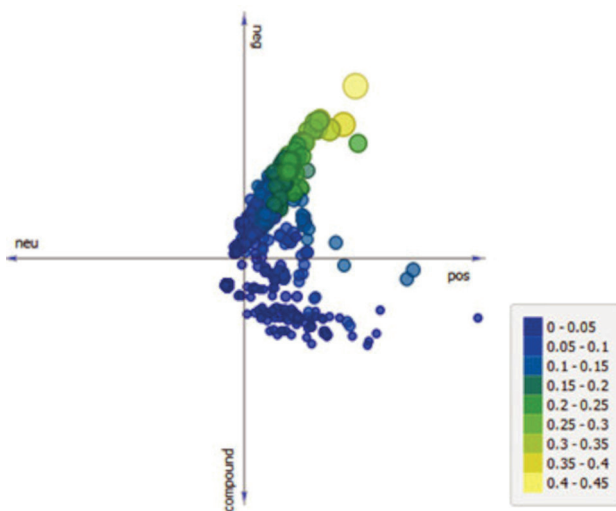


Figura 12. Resultados con puntuación de negatividad con FreeViz
Elaboración propia en Orange Canvas

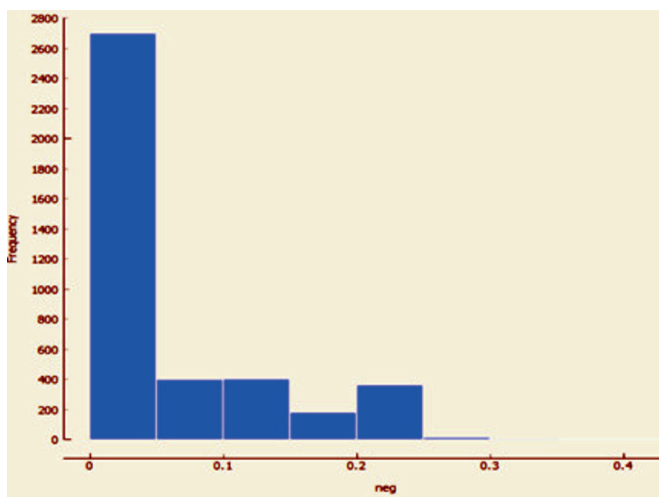


Figura 13. Gráfico de barras de resultados con puntuación hacia negatividad
Elaboración propia en Orange Canvas

De acuerdo con los resultados obtenidos, en la figura 13 se muestra el análisis de la intensidad a negatividad de los 4500 tuits; la primera barra agrupa 3018 tuits con un valor menor a 0,05 y un porcentaje de 67,07 %; la segunda barra agrupa 435 tuits con un valor de 0,05 y 0,1 que representa un 9,67 %; la tercera barra reúne 436 tuits con valores de entre 0,1 y 0,15, lo que representa el 9,69 %; y 205 tuits con valores de entre 0,15 y 0,2, con un porcentaje de 4,56 %; y 389 tuits con un valor de entre 0,2 y 0,25, con un porcentaje de 8,64 %. Por último, 13 tuits con una intensidad de entre 0,25 y 0,3, con un porcentaje de 0,29 %. De acuerdo con estos valores podemos corroborar lo mostrado en la figura 12 donde se muestra el tuit con una intensidad mayor negativa de 0,423. Se tienen 2700 tuits de los 4500 cuyos valores de puntuación son menores a 0,05 y que representan el 67,07 %, lo que nos indica que existe una intensidad de sentimiento a -1 .

5.4 Análisis de tuits con puntuación compuesta

La puntuación compuesta muestra la puntuación total de un comentario. Cuanto más cercano a 1, el nivel de puntuación tiende a ser más positivo; de lo contrario, se acerca a -1 y tiende a ser más negativo. En la figura 14 se muestran las puntuaciones compuestas, por lo que se ajustó el gráfico de barras a los valores umbrales típicos para determinar la puntuación total de los comentarios de manera unidimensional basándonos en los valores umbrales.

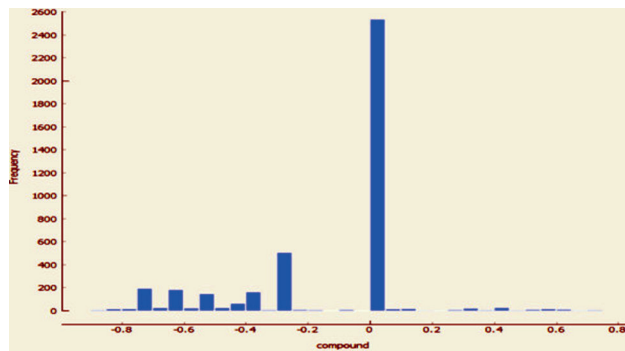


Figura 14. Gráfico de barras de valores umbrales
Elaboración propia en Orange Canvas

En el primer rango de la figura 14 se muestra un umbral en una puntuación entre -1 y $-0,05$ con 1524 tuits (33,9 %); este rango representa un sentimiento negativo en los tuits.

En los siguientes umbrales se muestra que están en el rango de $-0,05$ y $0,05$ con 2837 tuits (63,04 %); este rango representa un sentimiento neutro en los tuits. Como tercer umbral tenemos los valores de entre $0,05$ y 1 con 139 tuits (3,09 %); este rango representa un sentimiento positivo en los tuits.

6. CONCLUSIONES

En la red social Twitter, de los 4500 tuits que fueron analizados, los comentarios realizados por los usuarios acerca del dióxido de cloro como tratamiento del COVID-19 se asocian con los siguientes términos: *dióxido* (3706 veces), *cloro* (3695 veces), *uso* (687 veces), *COVID* (491 veces) y *coronavirus* (379 veces). Los datos se obtuvieron con la expresión *dióxido de cloro* y esta, a su vez, vinculó palabras como *uso*, *COVID* y *coronavirus*.

Los tuits de los usuarios, tomados los días 8, 9, 10 y 11 de octubre del año 2020, muestran sentimiento neutral y negativo sobre el dióxido de cloro como tratamiento para el COVID-19 y, de acuerdo con la investigación, 1524 tuits obtuvieron una puntuación de negatividad en relación a este tema, con un porcentaje de 33,9 %; del mismo modo, 2837 tuits obtuvieron puntuación de neutralidad ante el dióxido de cloro con un porcentaje de 63,04 %, y 139 tuits obtuvieron una puntuación de positividad sobre el dióxido de cloro con un porcentaje de 3,09 %.

REFERENCIAS

- Agencia para Sustancias Tóxicas y el Registro de Enfermedades. (2004). *Dióxido de Cloro y Clorito*. Edición del Autor.
- AS Perú. (26 de julio del 2020). *Qué es y para qué sirve el dióxido de cloro y qué dice el Minsa sobre su uso*. https://peru.as.com/peru/2020/07/26/actualidad/1595791000_910783.html
- Cabezas Sánchez, C., Solari Zerpa, L., y Caballero Ñopo, P. (2020). Eficacia y seguridad del dióxido de cloro para el tratamiento de COVID-19. *Instituto Nacional de Salud*, 2-10.
- EGilbert, H. (2014). *VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text*. Conference on Weblogs and Social Media (ICWSM-14).
- González Figueroa, M. (2018). *Búsqueda, descubrimiento y análisis de sentimientos en Twitter de la institución* [Tesis de maestría, Universidad Técnica Federico Santa María]. https://www.mti.cl/wp-content/uploads/2019/01/Tesina_2018_Gonz%C3%Allez-Marcela.pdf
- Mitra, S., y Acharya, T. (2005). *Data Mining: Multimedia, Soft Computing and Bioinformatics*. Wiley.
- Organización Mundial de la Salud. (2020). *Preguntas y respuestas sobre la enfermedad por coronavirus (COVID-19)*. https://www.who.int/es/emergencies/diseases/novel-coronavirus-2019/advice-for-public/q-a-coronaviruses?gclid=EAIAIQobChMIi-T4iYK46wIVeP7jBx1czw11EAAYASABEgKmh_D_BwE

- Orgaz, C. J. (15 de abril del 2020). *Coronavirus | Dióxido de cloro, el peligroso químico que se promociona como cura para el COVID-19 y sobre el que advierten los expertos*. <https://www.bbc.com/mundo/noticias-52303363>
- Orihuela, J. L. (2011). *Mundo Twitter*. Alienta.
- Rubín, R. (22 de mayo del 2020). *Qué es Twitter, cómo funciona y qué te puede aportar esta red social*. *Ciudadano 2.0*. <https://www.ciudadano2cero.com/twitter-que-es-como-funciona/>
- Trucco, F. (6 de agosto del 2020). *Bolivia aprueba dióxido de cloro como tratamiento para coronavirus*. <https://cnnespanol.cnn.com/2020/08/06/camara-de-diputados-de-bolivia-aprueba-dioxido-de-cloro-como-tratamiento-para-coronavirus/>
- Vallalta Rueda, J. F. (s. f.). *CRISP-DM: una metodología para minería de datos en salud*. Health Data Miner. <https://healthdataminer.com/data-mining/crisp-dm-una-metodologia-para-mineria-de-datos-en-salud/>