



SUMA DE NEGOCIOS



Artículo de investigación

Aislamiento social obligatorio: un análisis de sentimientos mediante *machine learning* [★]



Carlos Alberto Arango Pastrana¹ y Carlos Fernando Osorio Andrade²

¹ Doctor en Organización Industrial y Gestión de Empresas. Profesor Universidad del Valle, Cali, Colombia.
Email: carlos.arango.pastrana@correounivalle.edu.co. ORCID: 0000-0003-0893-1806.

² Magister en Ciencias de la Organización. Profesor Universidad del Valle, Cali, Colombia. (Autor de correspondencia).
Email: carlos.fernando.osorio@correounivalle.edu.co. ORCID: 0000-0002-5095-4991.

INFORMACIÓN DEL ARTÍCULO

Recibido el 5 de noviembre de 2020
Aceptado el 15 de diciembre de 2020
Online 29 de enero de 2021

Códigos JEL:
Z13, B23, H83, I18

Palabras clave:
Aislamiento obligatorio,
redes sociales,
análisis de sentimientos,
machine learning,
COVID-19.

Keywords:
Mandatory isolation,
Social networks,
Sentiment analysis,
Machine learning,
COVID-19.

R E S U M E N

Para reducir la tasa de contagio por COVID-19, el Gobierno colombiano ha adoptado, entre otras medidas, el aislamiento obligatorio. Esta medida ha generado opiniones divididas, pues a pesar de que ayuda a disminuir la propagación del virus, genera problemas mentales y económicos difíciles de sortear. El objetivo de este documento es analizar los sentimientos subyacentes de los comentarios de Twitter relacionados con el aislamiento, identificando los temas y palabras más frecuentemente utilizados en este contexto. Se construyó un algoritmo de *machine learning* para identificar los sentimientos de 72564 publicaciones, y se aplicó un análisis de redes sociales para identificar los temas más frecuentes en los conjuntos de datos. Los resultados sugieren que el algoritmo presenta gran precisión para clasificar sentimientos. Asimismo, a medida que se extiende el aislamiento, los comentarios relacionados con la cuarentena crecen de manera proporcional. Se identificó al miedo como el sentimiento predominante durante todo el periodo de confinamiento en Colombia.

Mandatory social isolation: a sentiment analysis using machine learning

A B S T R A C T

To reduce the rate of contagion by Covid-19, the Colombian government has adopted, among other measures, for mandatory isolation, with divided opinions, because despite helping to reduce the spread of the virus, it generates mental and economic problems that are difficult to overcome. The objective of this document was to analyze the underlying sentiments in the Twitter comments related to isolation, identifying the topics and words most frequently used in this context. A machine learning algorithm was built to identify

★ Premio al mejor artículo del número

sentiments in 72,564 posts and a social network analysis was applied establishing the most frequent topics in the data sets. The results suggest that the algorithm is highly accurate in classifying feelings. Also, as the isolation extends, comments related to the quarantine grow proportionally. Fear was identified as the predominant feeling throughout the period of confinement in Colombia.

Introducción

La llegada de la enfermedad del COVID-19 ha marcado un antes y un después en la sociedad mundial (Crokidakis, 2020). El contagio se produce de persona a persona y las tasas de infección preocupan a los organismos de salud. Por esta razón, los Gobiernos de todo el mundo han adoptado medidas estrictas de salud pública para contener la transmisión del virus y evitar la saturación de los sistemas sanitarios (Steffen et al., 2020). Dentro de las decisiones tomadas para contener la pandemia se encuentran: el uso obligatorio de máscaras, bloqueo de fronteras, cierre de instituciones educativas y aislamiento social obligatorio, también conocido como cuarentena (Weible et al., 2020).

De todas estas medidas, el aislamiento o cuarentena, sin lugar a dudas, ha sido la más polémica, pues ha despertado el debate a nivel mundial frente a sus ventajas y desventajas (Meier et al., 2020). Por un lado, existe abundante evidencia científica que respalda al aislamiento obligatorio como la principal herramienta para aplanar la curva de contagio (Crokidakis, 2020); por otra parte, se conoce que esta medida genera una serie de problemas económicos y sociales, como por ejemplo: recesión económica (Atalan, 2020), efectos psicológicos adversos (Shimizu, 2020), e incluso un detrimento de los derechos fundamentales, especialmente el de la libertad (Botes & Thaldar, 2020).

Este debate científico respecto a las ventajas y desventajas del aislamiento obligatorio se refleja en la opinión pública. En Colombia, a pesar de que la posición formal del Gobierno ha consistido en acatar las recomendaciones médicas y mantener el aislamiento obligatorio (Ministerio de Salud, 2020), opositores, grupos de empresarios y otras asociaciones se han pronunciado en contra, pidiendo mayor libertad para retomar las actividades económicas, conservar la salud mental y propiciar el esparcimiento (Marín Correa, 2020). Por supuesto, este debate se ha extendido a las redes sociales, donde, cada vez más, los usuarios se enfrentan en torno a los beneficios y perjuicios de la cuarentena (Díaz & Gutiérrez, 2002).

Con el propósito, justamente, de identificar los sentimientos subyacentes de los comentarios de las redes sociales relacionados con el aislamiento obligatorio, y de analizar las palabras o temas más frecuentes que usan los usuarios de estos medios, esta investigación efectúa un análisis de sentimientos y redes sociales a 72564 comentarios de Twitter. Para ello, se aplican técnicas de aprendizaje de máquina o *machine learning* (ML), haciendo uso de los paquetes informáticos R, Gephi y Rapidminer. Cabe destacar que ya en investigaciones previas se han aplicado este tipo de técnicas para entender las percepciones y sentimientos de los usuarios de redes sociales frente a diferentes temas. Se han usado para analizar, por ejemplo, las opiniones sobre una

política pública o sobre una noticia internacional. También se han empleado en marketing para analizar las percepciones y opiniones de los usuarios sobre productos y servicios (Martínez et al., 2011; Lin et al., 2017).

Aplicar el análisis de sentimientos a las opiniones de ciudadanos en Colombia representa dos aportes importantes para la literatura académica. En primer lugar, aunque este tipo de análisis de sentimientos (cuyo fin es extraer las emociones predominantes de los comentarios de las redes sociales) se puede encontrar en varios artículos de bases de datos académicas de alto nivel (como Scopus y Web of Science), las investigaciones que adelantan este tipo de modelos en conjuntos de datos distintos al inglés son escasas (Zambrano et al., 2019). De hecho, hasta el momento no se han encontrado investigaciones similares en países latinoamericanos de habla hispana ni en Scopus ni en Web of Science (dos de las bases de datos multidisciplinares más importantes). En otras palabras, este estudio puede ser utilizado por investigaciones futuras que analicen conjuntos de texto en español para identificar los mejores algoritmos de clasificación. Por otro lado, la comprensión de las emociones que profesan los colombianos frente a las medidas de aislamiento obligatorio puede ayudar a entender aspectos psicológicos y sociales potencialmente trascendentales para una futura implementación de acciones efectivas en contextos similares. Es decir, la administración pública puede aprovechar los resultados de esta investigación para entender las percepciones y creencias de los ciudadanos ante las medidas de aislamiento.

Teniendo en cuenta lo anterior, esta investigación se divide en seis apartados, incluida la presente introducción. En el segundo, se detalla la metodología empleada para extraer los comentarios y efectuar el análisis respectivo; en el tercero, se discuten los resultados; en el cuarto, se presenta la discusión frente a otros estudios similares; en el quinto, se exponen las conclusiones más relevantes y, por último, en el sexto apartado, se presentan las limitaciones y las oportunidades de investigaciones futuras.

Marco Referencial

El COVID-19 y las medidas de aislamiento obligatorio

El año 2020 ha traído consigo una nueva enfermedad que azota de manera inclemente a todos los países del mundo: el COVID-19 (Crokidakis, 2020). Desde sus primeros brotes, reportados a finales de 2019, en Wuhan-China, la enfermedad se ha propagado rápidamente por todo el planeta y, debido a esto, fue catalogada como una pandemia en marzo de 2020, por su alta velocidad de transmisión (Williams & Kayaoglu, 2020). Esta enfermedad infecciosa ataca las vías respiratorias y propicia múltiples complicaciones; algunos

individuos pueden, incluso, presentar una neumonía fulminante (Mukherjee, 2020; Organización Mundial de la Salud [OMS], 2020a). Para julio de 2020, el COVID-19 había cobrado la vida de más de 500 mil personas, una tasa de mortalidad considerable, especialmente, en personas mayores o con comorbilidades previas (OMS, 2020b). Por otra parte, las tasas de contagio por COVID-19 son tan altas que han sumido en una tremenda preocupación a los organismos de salud de todo el mundo. Por esto, los Gobiernos han decretado medidas muy estrictas para la contención del virus, que buscan mitigar su transmisión y mantener en funcionamiento los sistemas sanitarios (Steffen et al., 2020). Dentro de estas medidas, se encuentran: el uso obligatorio de máscaras, bloqueo de fronteras, cierre de instituciones educativas y aislamiento social obligatorio o cuarentena (Weible et al., 2020).

De todas las medidas adoptadas para contener el virus la que ha despertado mayor debate ha sido el aislamiento social obligatorio, debido a las ventajas y desventajas que representa (Meier et al., 2020). No se puede negar que la evidencia científica señala que el aislamiento obligatorio es la principal medida para mitigar la curva de contagio de enfermedades respiratorias (Crokidakis, 2020); sin embargo, diversas investigaciones también sugieren que la cuarentena trae consigo una inminente crisis y recesión económica a nivel mundial (Atalan, 2020; Jenson, 2020; Musinguzi & Asamoah, 2020). De acuerdo con Nicola et al., (2020), las cuarentenas de todo el mundo han propiciado estragos productivos y económicos en todos los sectores. Los productos agrícolas, por ejemplo, han visto una caída en sus precios de más del 20%, mientras que, en el sector de servicios, algunas industrias como el turismo reportan que más de 120 millones de empleos están en riesgo (Sigala, 2020). De hecho, las proyecciones establecidas por la Organización Internacional del Trabajo (2020) señalan que se perderán alrededor de 25 millones de empleos a nivel mundial, lo cual propiciará pobreza, caídas en el consumo de productos y servicios, reducción salarial, entre otras consecuencias negativas.

Las razones económicas no son las únicas que evidencian los puntos débiles del aislamiento obligatorio. Los estudios psicológicos y psiquiátricos también han mencionado el costo enorme que genera la cuarentena en varias dimensiones de la vida de las personas, entre ellas, la salud mental (Reddy et al., 2020). Investigaciones sobre estos temas han encontrado que la ansiedad, la depresión, la soledad, el deterioro de la funcionalidad, el insomnio y el trastorno de estrés postraumático, son los problemas mentales más recurrentes en las personas sometidas al aislamiento obligatorio (Shimizu, 2020; Luo et al., 2020; Ramírez et al., 2020). Teniendo en cuenta este escenario, los estudios sugieren que hay que considerar que la salud mental es tan importante como la física y, por tanto, hacen un llamado a no ignorar este factor dentro del debate en torno a la cuarentena (Huang et al., 2019). Finalmente, algunas investigaciones se han propuesto analizar los efectos que tiene el aislamiento obligatorio en los derechos fundamentales, particularmente, en el de la libertad (Botes & Thaldar, 2020).

El debate alcanza a las redes sociales

El debate académico que se ha generado en torno a los beneficios y perjuicios de la cuarentena se refleja también

en la opinión popular. En un país como Colombia, donde la posición del Gobierno se ha centrado en acatar las recomendaciones de los organismos de salud que piden mantener el aislamiento obligatorio (Ministerio de Salud, 2020), han surgido varias voces que se pronuncian en contra de esta medida. Grupos de oposición, gremios de empresarios y otras asociaciones pidieron, en su momento, mayor libertad para reactivar la economía, para cuidar la salud mental y para propiciar el esparcimiento y las actividades culturales (Marín Correa, 2020). De hecho, se hicieron algunas marchas y protestas sociales con el fin de exigirle al Gobierno el levantamiento gradual o total de la cuarentena en el país (El Espectador, 2020). Cabe destacar que esta ola de protestas frente a las medidas de aislamiento no es un hecho aislado que haya ocurrido solo en Colombia. En otros países de la región, como Argentina y Chile, también se han suscitado este tipo de manifestaciones (Ferreya et al., 2020; Nava & Grigera, 2020).

Por supuesto, el debate en torno a la pertinencia de las medidas de aislamiento obligatorio ha alcanzado el escenario de las redes sociales. No hay que olvidar que, en la actualidad, estos medios se constituyen como espacios deliberativos. Un 60% de los adultos utiliza las redes sociales, y alrededor de dos tercios de estas personas son políticamente activos a través de estos medios. Además, las redes sociales se presentan como medios de expresión de la identidad personal y social, lo que ofrece grandes oportunidades para la investigación cualitativa y cuantitativa en ciencias sociales (Smith, 2013; Reyes-Menendez et al., Saura & Thomas, 2020).

Frente a la cuarentena, se han difundido millones de publicaciones en redes sociales como Facebook, Twitter, Instagram, e incluso la recientemente popular Tik Tok (Díaz & Gutiérrez, 2020). Esto quiere decir que las redes sociales se han convertido en un termómetro importante para medir las percepciones de los internautas sobre la aplicación de este tipo de restricciones (Ferreya et al., 2020). Teniendo en cuenta lo anterior, esta investigación, como ya se ha mencionado, se propone identificar los sentimientos y creencias que tienen los usuarios de Twitter con respecto a las medidas de aislamiento promulgadas por el Gobierno de Colombia durante la emergencia del COVID-19.

Metodología de investigación

Análisis de sentimientos en redes sociales

El análisis de sentimientos (AS) es definido por Pang y Lee (2008) como “el tratamiento computacional de opiniones, sentimientos y subjetividad en textos” (p. 6). Por su parte, Cambria, et al. (2012) lo presentan como “el conjunto de técnicas computacionales para la extracción, clasificación, comprensión y evaluación de opiniones expresadas en fuentes publicadas en Internet, comentarios en portales web y en otros contenidos generados por usuarios” (p. 560). Finalmente, Saura et al. (2019) definen el AS como una técnica de investigación que analiza los sentimientos de una muestra determinada de textos que, normalmente, proviene de entornos digitales, como plataformas online o redes sociales. El análisis de sentimientos hace parte del campo del

Procesamiento de Lenguaje Natural (PNL) y, con esta técnica, se busca extraer información de sitios web de manera automática y, además, identificar los sentimientos o emociones que pueden contener los textos. Para realizar el AS pueden aplicarse métodos con diccionarios de emociones preestablecidos, también conocidos como lexicón, o métodos como el *machine learning* (aprendizaje de máquina), que consiste en la construcción de algoritmos que aprenden a clasificar de manera automática grandes conjuntos de datos (Saura, 2020). También existen métodos híbridos que combinan ambas técnicas (Saura et al., 2019; Zambrano et al., 2019).

En la literatura científica se reporta una gran cantidad de publicaciones de alto nivel en bases de datos como Web of Science y Scopus que aplican el AS para analizar comentarios en redes sociales frente a diferentes tópicos. Cano y Arce (2020), por ejemplo, emplearon métodos de *big data* para extraer comentarios de Twitter sobre la vacuna para la gripe en España. Mediante el análisis de sentimientos y el lexicón desarrollado por el National Research Council of Canadá (versión 0.92), en su versión en castellano, determinaron cuáles fueron las emociones más relevantes y, a través de análisis de clúster, hallaron los usuarios o *influencers* más importantes.

Por otro lado, Baker et al., (2020) hicieron uso del AS para detectar la gripe utilizando técnicas de aprendizaje automático en *tweets* de países árabes. Los modelos de clasificación empleados fueron: árboles de decisión, vectores de máquinas de soporte (SMV), Naive Bayes y K vecinos cercanos (K-NN). Wongkar y Angdresey (2019) también realizaron una aplicación de AS para Twitter con el objetivo de determinar las emociones que expresaban los indonesios frente a la campaña presidencial de 2019. Para su clasificación, usaron el algoritmo Naive Bayes. Otro ejemplo de aplicación de análisis de sentimientos se encuentra en el estudio de Saura et al., (2019), quienes analizaron los sentimientos de los usuarios españoles de Twitter respecto a las ofertas del *Black Friday* en 2018.

A pesar de que el AS ha sido ampliamente utilizado en la investigación científica para una gran variedad de temas, la mayoría de las investigaciones solo han considerado el idioma inglés, lo que implica que los avances en otras lenguas son aún escasos (Zambrano et al., 2019). Por tanto, analizar las emociones que profesan los colombianos frente a la cuarentena genera un aporte de valor al campo del AS.

La metodología aplicada para llevar a cabo el análisis de sentimientos y la clasificación de comentarios en Twitter se resume en la Figura 1.

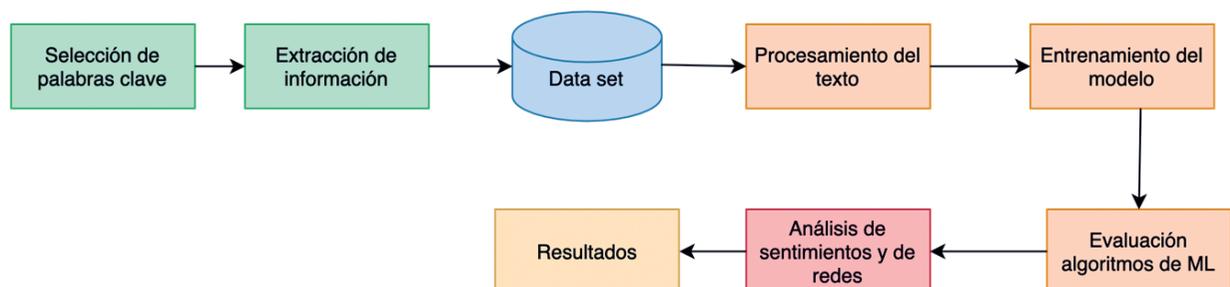


Figura 1. Diagrama de flujo de la metodología
Fuente: elaboración propia.

A continuación, se explica en detalle cada una de las etapas desarrolladas:

Selección de palabras clave (keywords)

Se escogieron las palabras y *hashtags* “aislamiento obligatorio”, “aislamiento”, “confinamiento” y “cuarentena” porque devolvían publicaciones relacionadas con el target de investigación.

Recolección de información

La captura de los comentarios procedentes de Twitter se realizó mediante minería de datos, a través del software estadístico R en su entorno visual Rstudio (Versión 1.2.1335). Se utilizaron las librerías, *Rtweet* (versión 0.7.0) (Kearney y Kearney, 2016), y la conexión con la interfaz de programación de aplicaciones (API) para programadores de Twitter. La búsqueda de comentarios se hizo mediante las palabras clave y se obtuvo un total de 72564 *tweets* desde el 15 de mayo hasta el 31 de julio de 2020, periodo de confinamiento en Colombia.

Cabe resaltar que la API de Twitter tiene la limitación de que solo permite la extracción de comentarios hasta siete días antes de la consulta, por tanto, no se logró obtener comentarios o publicaciones de los primeros días de la cuarentena en Colombia (que empezó en marzo de 2020). Sin embargo, el periodo de tiempo escogido (dos meses y medio) y la cantidad de datos (72564) permitieron un análisis importante de las opiniones y sentimientos de los colombianos frente a las medidas de aislamiento social.

Preprocesamiento de datos

En el preprocesamiento de datos, las publicaciones se someten a una serie de técnicas que permiten limpiar y reducir las características de los textos, facilitando el análisis de sentimientos mediante ML. Para procesar los datos se utilizó el paquete informático *Rapidminer* en su versión 9.6.0. Este programa se escogió porque cuenta con varios algoritmos de clasificación de textos integrados. Además, recientemente ha sido empleado con éxito en otros trabajos de este tipo (Wongkar & Angdresey, 2019; Aballay et al., 2017). Las técnicas aplicadas en este caso fueron cuatro: (1) filtrado de textos, (2) tokenización, (3) *transform cases* y (4) eliminación de *stopwords*, que se explican en detalle a continuación:

- **Filtrado:** se aplica para eliminar palabras innecesarias, como repeticiones, signos de puntuación, emojis, numerales, espacios, entre otros. En este trabajo se realizó la eliminación de preposiciones y conjunciones que se consideran palabras de detención (Allahyari et al., 2017).
- **Tokenización:** tiene como objetivo dividir las oraciones en trozos, ya sea en palabras o frases, y produce piezas más pequeñas denominadas tokens. En este estudio, se dividieron los comentarios por caracteres (palabras) y después se tomaron los tokens para las otras etapas de procesamiento (Demidova & Klyueva, 2017).
- **Transform cases:** se convierten todas las letras a minúsculas para normalizar el texto y evitar redundancias.
- **Eliminación de stop words:** se utiliza para eliminar palabras innecesarias y sin sentido, como conectores y artículos (por ejemplo: y, de, a, etc.). El objetivo de este paso consiste en limpiar los textos y dejar, solamente, palabras relevantes (Pratama et al., 2019).

Entrenamiento del modelo

Para crear un modelo de clasificación de comentarios en función a los sentimientos, existen dos maneras documentadas en la literatura. Por un lado, se pueden utilizar los diccionarios, o lexicón, predefinidos que clasifican las palabras de los textos de acuerdo con el sentimiento que representan (Sauter et al., 2010). Otra manera de hacerlo, y que se considera más apropiada (Manning et al., 2008), es construir un algoritmo de aprendizaje de máquina que aprenda, a partir de conjuntos de datos de entrenamiento y de prueba, a reconocer los sentimientos y emociones en los textos. Para esta investigación se utilizó este último método, y se emplearon los conjuntos de datos de entrenamiento y testeo WASSA-2017 en su versión en español, creados por Mohammad y Bravo (2017). Este corpus de datos es extraído de Twitter con 6755 comentarios (3613 de entrenamiento y 3142 de prueba), cada uno con una etiqueta de sentimiento asignada: alegría, ira, miedo y tristeza. En la Tabla 1 se presenta la distribución de emociones para el corpus empleado.

Tabla 1. Distribución de emociones

Sentimiento	Cantidad datos de entrenamiento	Cantidad datos de prueba
Alegría	823	714
Ira	857	760
Miedo	1147	995
Tristeza	786	673

Fuente: elaboración propia con base en Mohammad y Bravo (2017).

Evaluación de algoritmos

Con el corpus de datos de entrenamiento y de prueba, se construyó el modelo de análisis de sentimientos en Rapidminer. Además, se probaron varios algoritmos de aprendizaje de máquina para clasificar los textos de acuerdo

con las emociones. Se probaron tres algoritmos integrados al software RapidMiner: Naive Bayes, Naive-Bayes Kernel y aprendizaje profundo o Deep Learning (Kotu & Deshpande, 2014). A continuación, se detalla cada uno de ellos:

Algoritmo Naive Bayes (NB)

Este algoritmo es uno de los más utilizados en modelos de análisis de sentimientos. Es un método de clasificación supervisado que se desprende del teorema de Bayes, y precisa la independencia de los atributos en el aprendizaje. Para empezar, la probabilidad de cada palabra está definida, después de eso, el clasificador se construye para agrupar los comentarios según la categoría o el sentimiento al que pertenece.

$$P(H|X) = \frac{P(X|H)(H)}{(X)} \quad (1)$$

Donde: H son los atributos del documento, X la clase y $P(X|H)$ la probabilidad de ocurrencia del atributo en la clase dada. La clase seleccionada por el clasificador será la que maximice $P(X|H)$ (Baker et al., 2020; Han et al., 2011).

Algoritmo Naive Bayes Kernel (NB)

Una variante del Naive Bayes (NB) es el Naive Bayes Kernel (NB-Kernel), que mantiene las ventajas del NB y, además, se puede aplicar en situaciones donde los datos no siguen una distribución normal. Un kernel es una función de peso usada en técnicas de estimación no paramétrica. Los estimadores de tipo kernel se diseñaron para superar las dificultades, pues usan técnicas paramétricas cuando el comportamiento de los datos no sigue una distribución normal. Son los más utilizados en estimación no paramétrica y su propiedad más importante es que no se ven afectados por la función kernel que se elija.

$$f(x|H) = \frac{1}{n} \sum_{i=1}^n K_n(x - x^{(i)}) \quad (2)$$

Donde: H es una matriz $d \times d$, x un vector de d dimensiones, n el número de casos de aprendizaje, i el índice de casos en el conjunto de entrenamiento, y $K_n(\cdot)$ es la función de kernel usada. Es importante observar que X es multivariable y que $f(x|H)$ es una función multivariable. El kernel es determinado por el promedio de n kernel $K_n(\cdot)$ obtenido en cada observación $x^{(i)}$. La función kernel se define a continuación:

$$K_H(x) = |H|^{-1/2} K(H^{-1/2}x) \quad (3)$$

Se asume que K es una función de densidad d-dimensional. Algunos kernel de uso común son: uniforme, triangular y gaussiano (Pérez et al., 2009).

Modelo Deep Learning

Deep Learning (DL), conocido en español como aprendizaje profundo, es una técnica de aprendizaje automático que usa arquitecturas de redes neuronales. Se le atribuye el concepto de "profundo" por el número de capas que posee. Se ocupa de emular el enfoque de aprendizaje que utilizamos

los seres humanos para obtener cierto tipo de conocimiento. Puede considerarse una forma de automatizar el análisis predictivo. Los algoritmos de aprendizaje profundo se apilan en una jerarquía de creciente complejidad y extracción. Cada algoritmo de la jerarquía, en su entrada, aplica una transformación no lineal y utiliza lo que aprende para crear un modelo estadístico como salida. Las iteraciones continúan hasta que la salida ha alcanzado un nivel de precisión aceptable. En el aprendizaje profundo, el modelo aprende por sí mismo y puede, así, descubrir las relaciones entre las variables. En los modelos basados en aprendizaje profundo no interviene el ser humano, ya que el propio modelo es capaz de hacer la selección de características (Sáez, 2019).

Comparación de algoritmos

Existen diversas maneras de evaluar los resultados obtenidos en el análisis de sentimientos mediante el uso de diferentes algoritmos. Las más comunes son la exactitud, la sensibilidad (Recall) y la precisión.

- Exactitud: métrica para evaluar modelos de clasificación. Es la fracción de predicciones que el modelo realiza correctamente.
- La sensibilidad (Recall): número de elementos identificados correctamente como positivos (del total de positivos verdaderos).
- Precisión: número de elementos identificados correctamente como positivos (de un total de elementos identificados como positivos).

Estas métricas se pueden calcular a partir de las siguientes ecuaciones:

$$\text{Accuracy (A)} = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (4)$$

$$\text{Precision (P)} = \frac{TP}{(TP + FP)} \quad (5)$$

$$\text{Recall (R)} = \frac{TP}{(TP + FN)} \quad (6)$$

Donde: TP (True Positive) es verdaderos positivos, FP (False positive) es falsos positivos, TN (True negative) es verdaderos negativos, FN (False negative) es falsos negativos.

Pesado de características

Una vez establecidos los algoritmos de aprendizaje de máquina, se identificó la aparición de los términos relacionados con las emociones de interés: alegría, ira, miedo y tristeza (Teso et al., 2018). Existen diferentes modelos para pesar las características del texto: pesado Booleano o Binario, Frecuencia del Término (TF), Frecuencia Inversa del documento (IDF) y TF-IDF (Montiel et al., 2009). En esta investigación el pesado que arrojó mejores métricas de evaluación fue la frecuencia de términos (TF), propuesto por Luhn (1957). Este pesado otorga mayor relevancia a los términos con mayor frecuencia, evaluando el número de veces que

aparecen en el documento, tal como lo señala la siguiente ecuación:

$$P_i(t_j) = F_{ij} \quad (7)$$

Donde: F_{ij} es la frecuencia del término j en el documento i .

Análisis de redes

Una vez efectuado el análisis de sentimientos, se efectuó un análisis de redes sociales con la intención de determinar los términos más frecuentes en el conjunto de datos estudiados. Mediante este análisis se identificaron los temas de mayor frecuencia, que se presentan en los gráficos como nodos (temas), y se determinaron las conexiones entre ellos mediante la modularidad, que se ve reflejada en el gráfico mediante distintos colores (Bastian et al., 2009). En otras palabras, los términos que aparecen en un mismo color están altamente relacionados entre sí. Para llevar a cabo el análisis de redes se empleó el paquete informático Gephi, en su versión (0.9.2), y también se utilizó R para extraer la nube de palabras y la frecuencia de aparición de los términos más recurrentes.

Resultados

Se presentan los resultados del estudio, empezando por la evaluación de los modelos propuestos. En la Figura 2, se observa el desempeño de los tres algoritmos empleados (NB Kernel, NB y DL).

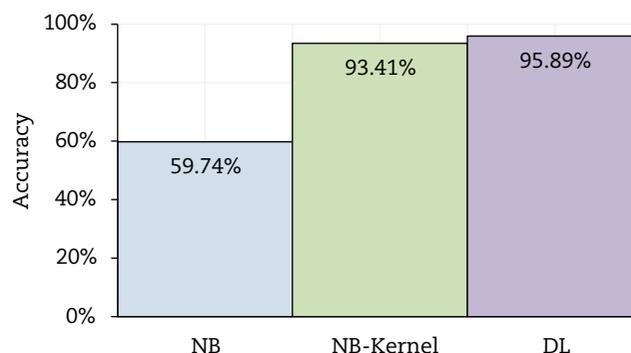


Figura 2. Desempeño de las técnicas de clasificación

Fuente: elaboración propia con base en los datos obtenidos en Twitter (15 de mayo - 31 de julio 2020).

Como puede observarse en la Figura 2, de los tres algoritmos aplicados el que presenta un mejor rendimiento es Deep Learning (DL), con una precisión de clasificación del 95.89%. Por tanto, este fue el algoritmo seleccionado para clasificar los sentimientos en los comentarios extraídos de Twitter. Vale la pena señalar que la precisión de clasificación se encuentra dentro de los rangos apropiados para este tipo de modelos (Pang & Lee, 2008). A continuación, en la Figura 3 se muestra la precisión y la sensibilidad del modelo escogido para cada una de las categorías analizadas.

Como se puede ver, la precisión obtenida en todas las categorías fue mayor al 90%, lo que indica que el modelo realiza una correcta clasificación de los elementos que pertenecen a la categoría correspondiente. Igualmente, en la sensibilidad, se obtuvieron valores mayores al 90%, lo que indica que el modelo presenta un ajuste adecuado.

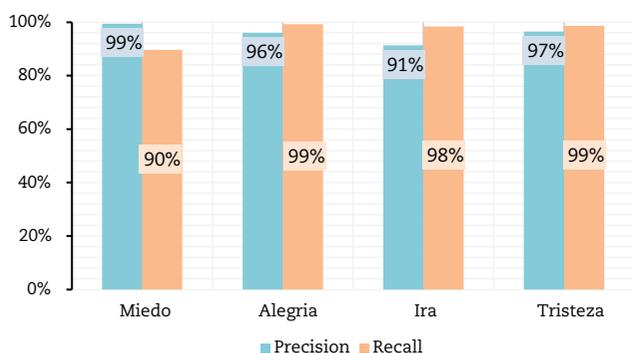


Figura 3. Precisión y sensibilidad de la clasificación

Fuente: elaboración propia con base en los datos obtenidos en Twitter (15 de mayo - 31 de julio 2020).

Volumen de publicaciones asociadas con cuarentena

Como se indicó anteriormente, los comentarios de Twitter recopilados corresponden al periodo del 15 de mayo hasta el 31 de julio de 2020. En promedio, diariamente se observaron 930 tweets referentes al tema, con un mínimo de 158 y un máximo de 4441 en un día, lo que indica una varianza considerable. La Tabla 2 presenta el resumen de los estadísticos descriptivos.

Cabe resaltar que, frente a la veracidad de los tweets publicados, el 94% de las cuentas que participan en este debate no son verificadas; a su vez, el 71% de los tweets son difusiones de otras personas (*retweets*).

Volumen de publicaciones en el tiempo

En la Figura 4 se observa la evolución, a través del tiempo, de los comentarios relacionados con la cuarentena en la red social.

Como puede observarse, los comentarios asociados con la cuarentena se dispararon en las fechas en las que el presidente de la República anunció, en medios de comunicación,

Tabla 2. Estadísticos descriptivos

Estadística	Valores
Media	930.3
Mediana	686.5
Desviación estándar	732.8
Varianza	537001.1
Mínimo	158
Máximo	4441
Suma	72564

Fuente: elaboración propia con base en los datos obtenidos en Twitter (15 de mayo - 31 de julio 2020).

la extensión del aislamiento obligatorio, lo que, hasta agosto de 2020, había ocurrido nueve veces (El Tiempo, 2020a).

Se observan cinco extensiones de la cuarentena: 19 de mayo, 28 de mayo, 23 de junio, 07 de julio y 28 de julio (Dinero, 2020a, 2020b, 2020c, 2020d, 2020e). Es pertinente mencionar que a medida que pasa el tiempo los comentarios relacionados con la cuarentena disminuyen de manera significativa, sobre todo si se contrastan con la semana del 24 al 30 de mayo, donde se observó el mayor volumen de publicaciones asociadas con este tema. Esto se explica porque, para el 31 de mayo, una gran cantidad de ciudadanos creían que se iban a levantar las medidas de aislamiento obligatorio (El Tiempo, 2020b).

Sentimientos asociados con la cuarentena

Ahora, respecto al análisis de sentimientos, a continuación, se presenta la Figura 5, donde se observa la evolución de sentimientos a través del tiempo.

Como puede notarse, durante el mes de mayo el sentimiento predominante fue el miedo; sin embargo, con la

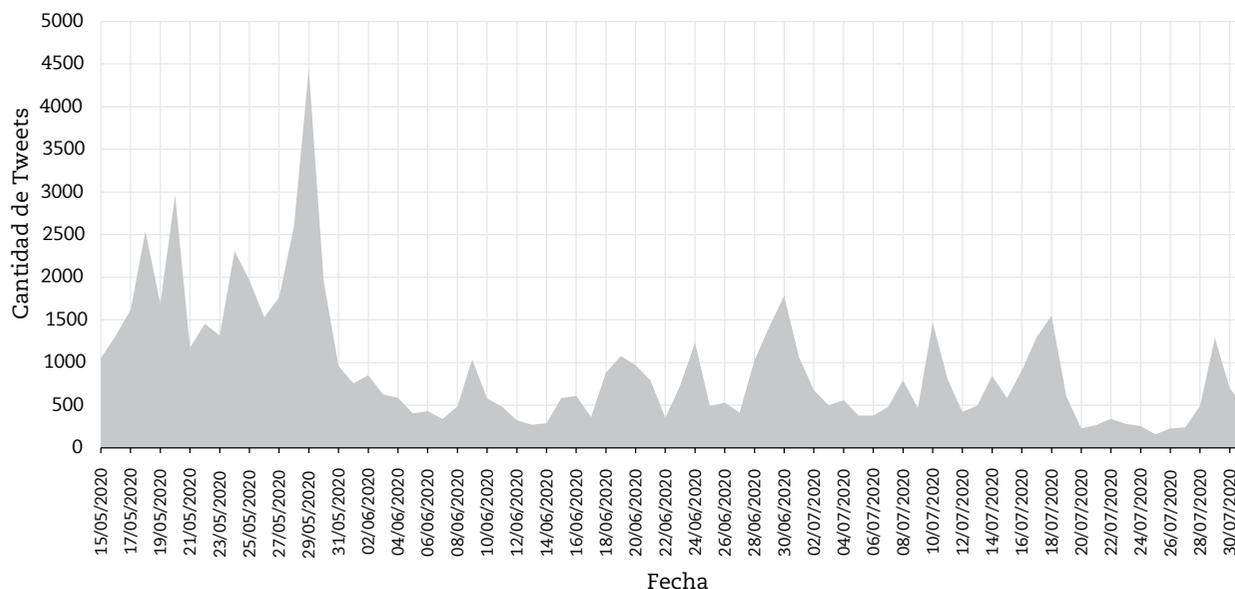


Figura 4. Evolución del volumen de comentarios

Fuente: elaboración propia con base en los datos obtenidos en Twitter (15 de mayo - 31 de julio 2020).

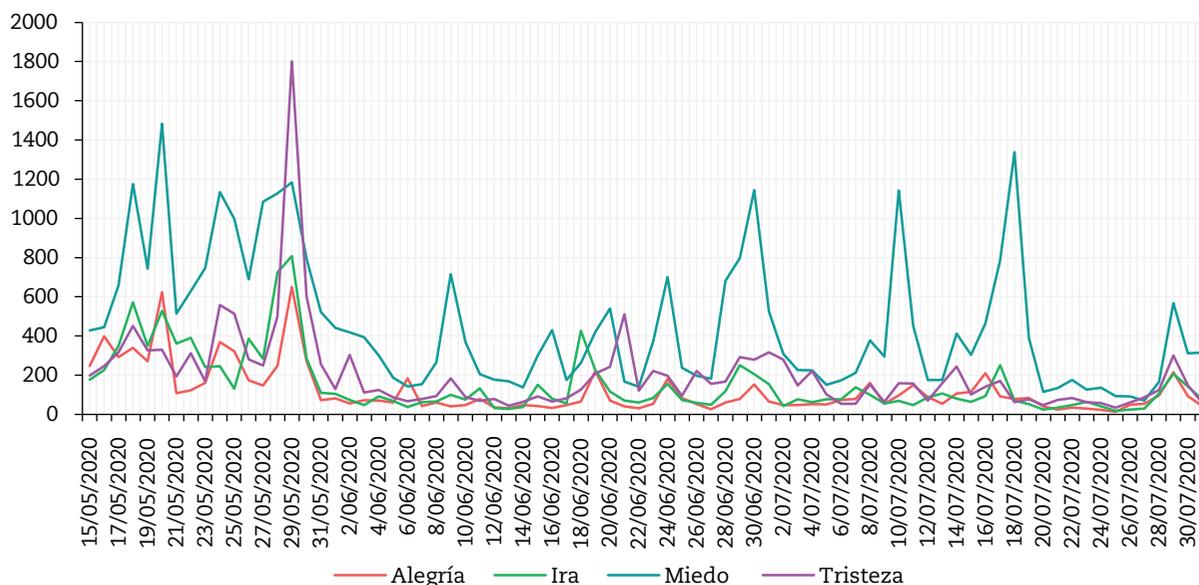


Figura 5. Evolución de las emociones a lo largo del tiempo

Fuente: elaboración propia con base en los datos obtenidos en Twitter (15 de mayo - 31 de julio 2020).

ampliación de la cuarentena en la semana del 24 al 30 de mayo, la tristeza lo superó, brevemente, aunque de manera considerable frente a los demás sentimientos. Durante junio y julio, nuevamente, el miedo fue el sentimiento más importante en los comentarios; no obstante, disminuyó con respecto al mes de mayo. A continuación, en la Figura 6 se observa la distribución general de emociones en los datos analizados.

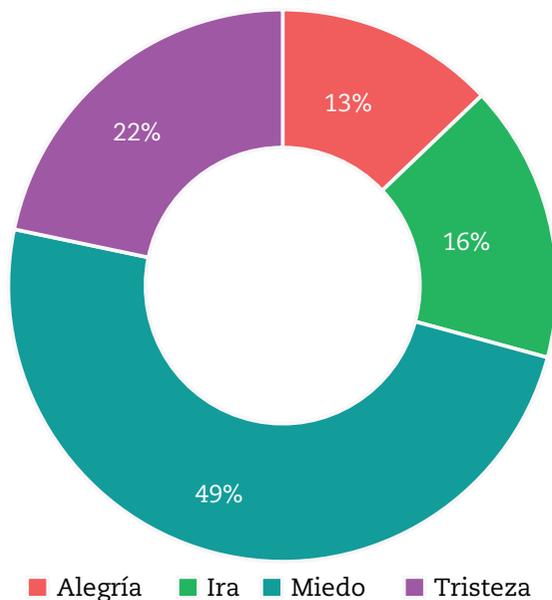


Figura 6. Emociones extraídas de los comentarios de Twitter

Fuente: elaboración propia con base en los datos obtenidos en Twitter (15 de mayo - 31 de julio 2020).

Con el 49%, el sentimiento predominante de los usuarios de Twitter con respecto a la cuarentena fue el miedo, esto quiere decir que, del total de comentarios analizados, 35614 expresan, como sentimiento principal, miedo. En segundo lugar, aparece la tristeza con 22% (15766 comentarios); en

tercer lugar, la ira con 16% (11814 comentarios) y, finalmente, la alegría con solo un 13% (9370 comentarios).

Principales hashtags utilizados en los datos analizados

La centralidad como estadístico de opinión giró en torno a los temas “Colombia” y “cuarentena”; sin embargo, “coronavirus”, “COVID-19”, “quédate en casa” y “pandemia” también están altamente relacionados al confinamiento. En la Tabla 3 se presentan los 10 primeros términos más frecuentemente mencionados con numerales (#) en los comentarios de Twitter.

Tabla 3. Top 10 de los “hashtags” más utilizados

Tema principal	Vector de centralidad
Colombia	1
Cuarentena	0.882143
Covid19	0.302183
Coronavirus	0.299981
Quedateencasa	0.299839
Bogotá	0.236388
Covid_19	0.232682
Medellín	0.17758
Venezuela	0.162774
Pandemia	0.14456
Barranquilla	0.11823
Cali	0.1146
Salud	0.113164

Fuente: elaboración propia con base en los datos obtenidos en Twitter (15 de mayo - 31 de julio 2020).

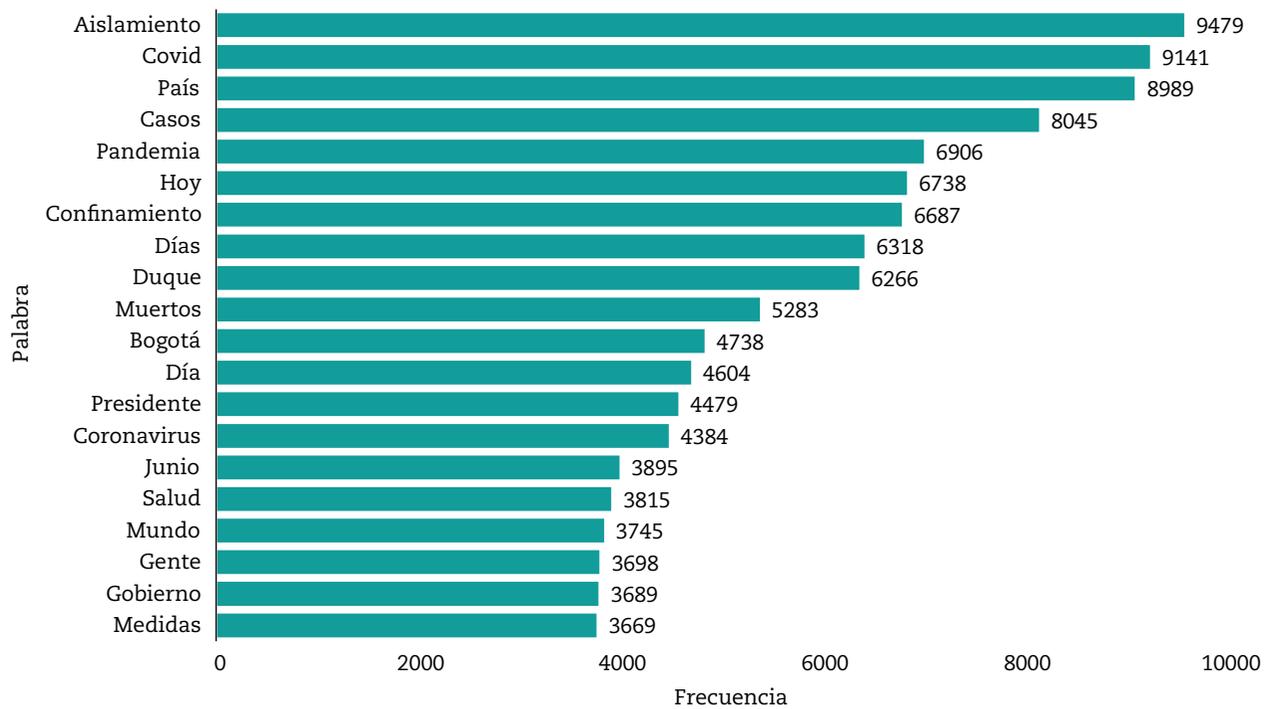


Figura 9. Términos más frecuentes

Fuente: elaboración propia con base en los datos obtenidos en Twitter (15 de mayo - 31 de julio 2020).

precisión bastante considerable, de 95.89%. Cabe resaltar que, hasta el momento, no se registra ni en Scopus ni en Web of Science un estudio que pruebe diferentes algoritmos de ML para el conjunto de datos *Wassa-2017* en su versión en español (que es uno de los conjuntos de entrenamiento más extendidos por la literatura).

En relación con otros trabajos que hayan abordado temas similares, vale la pena destacar el documento de Ferreyra y Nieto (2020), quienes analizan los tópicos más importantes relacionados con la cuarentena en Mar del Plata, Argentina. Si bien no aplican análisis de sentimientos, estos autores emplean la técnica de análisis de redes sociales para determinar grupos o clúster de comentarios relacionados. Ferreyra y Nieto destacan que, gracias al aislamiento obligatorio, también se encontraron grandes grupos de oposición al Gobierno argentino, lo que refuerza la idea de que es una medida polémica, incluso por fuera de Colombia.

El estudio de Xue et al. (2020) también se propone entender el discurso y las reacciones psicológicas de los usuarios de Twitter frente al COVID-19. A diferencia de nuestro estudio, estos autores analizan un conjunto de datos en idioma inglés y emplean el modelo de asociación latente de Dirichlet para identificar los tópicos más recurrentes. Este modelo permite agrupar palabras de acuerdo con grandes temas, haciendo uso de la inteligencia artificial. Para el análisis de sentimientos aplicado, usaron el lexicon NRC del National Research Council, que permite distinguir entre ocho emociones distintas: alegría, tristeza, confianza, disgusto, miedo, ira, sorpresa y anticipación. Encontraron que la emoción más recurrente en el conjunto de datos analizado fue la confianza, lo que contrasta notoriamente con el miedo, que fue la emoción predominante en los usuarios de Twitter colombianos. Sin embargo, hay que destacar que el

estudio de Xue et al. (2020) contempló varios tópicos relacionados con el COVID-19 y no solamente el aislamiento social obligatorio.

Finalmente, otra investigación con la que vale la pena comparar nuestros hallazgos es la de Low et al., (2020). Estos autores aplicaron el análisis de sentimientos a las publicaciones difundidas por usuarios del sitio web Reddit que estaban relacionadas con el COVID-19. Al igual que en nuestra investigación, encontraron que el volumen de publicaciones relacionadas con la enfermedad mantuvo picos en los primeros meses de la pandemia. También, encontraron que sentimientos como la ansiedad y, en general, las polaridades semánticas negativas fueron predominantes en las publicaciones, lo que es coherente con nuestros resultados.

Conclusiones

Del estudio realizado pueden desprenderse conclusiones relevantes para entender el comportamiento y los sentimientos de los ciudadanos frente a las medidas de restricción a la movilidad impuestas por las administraciones. También se identificaron elementos que aportan al campo de estudio del análisis de sentimientos en redes sociales.

Frente al análisis de los comentarios de Twitter relacionados con las medidas de aislamiento obligatorio en Colombia, el volumen de publicaciones disminuye a través del tiempo en la medida en que las restricciones se flexibilizan y, posiblemente, en la medida en que los ciudadanos se adaptan a las nuevas condiciones de aislamiento. Sin embargo, es una constante el hecho de que los comentarios aumenten de manera considerable cuando la Presidencia de la República anuncia las extensiones de la cuarentena. Del mismo modo,

el aumento de comentarios se produce cuando se anuncian estas medidas a través de los medios de comunicación, pero no cuando se emiten los decretos, lo que implica que los colombianos están más pendientes de los medios que de las comunicaciones formales del Gobierno.

En líneas generales, la cuarentena es un tema que genera, ante todo, sentimientos negativos, y el miedo es el más recurrente. Esto es previsible, dado que el aislamiento obligatorio trae consigo desempleo, cierre de negocios y, en general, una profunda crisis económica (Peters 2020) que sumerge a los ciudadanos en un estrés constante, y en un temor considerable, por sus inciertos futuros financieros. Es interesante señalar también que la tristeza, el segundo sentimiento más predominante en la muestra analizada, aumenta de manera considerable cuando el Gobierno nacional extiende la cuarentena. Esto implica que los colombianos se encuentran a la expectativa del momento en el que se levanten por completo las restricciones, pero, si se amplían las medidas de aislamiento, los usuarios de Twitter se sumergen en la pena. Teniendo en cuenta los sentimientos que generan las medidas de aislamiento obligatorio, los Gobiernos de todo el mundo deberían implementar estrategias de comunicación que permitan aminorar estas emociones. Además, esto debería verse reflejado en acciones y políticas públicas que ayuden a sobrellevar las medidas de restricción (Van Bavel et al., 2020).

Los temas que más se abordan en el cuerpo de comentarios analizado se constituyen de corrientes de comentarios alusivos a temáticas específicas, como la preocupación por ciudades concretas, otros países o temas internacionales, pasatiempos y formas de escapar de la rutina del aislamiento, política pública, economía, medidas adoptadas por el gobierno y, finalmente, un grupo que aborda temas asociados con el virus. En líneas generales, entonces, las personas que participan en el debate de Twitter sobre la cuarentena se concentran en tópicos que se relacionan con sus propios intereses.

Limitaciones y futuras líneas de investigación

La principal limitación de este estudio fue no poder conseguir los datos desde el inicio de la cuarentena. Por eso, sería interesante que futuras investigaciones recopilen esta información para analizar la evolución de las emociones durante la totalidad del aislamiento. Por otra parte, como el público objetivo que atiende las redes sociales puede variar de manera considerable en términos psicográficos y conductuales, se abre la posibilidad de analizar los sentimientos en otras redes como Facebook, Instagram o Tik Tok.

Finalmente, estudios similares podrían realizarse en otros países y establecer comparaciones, dado que las respuestas ante las medidas de aislamiento obligatorio pueden diferir.

Financiación

Artículo de investigación producto del proyecto de investigación Sistema de gestión logística local en escenarios de riesgo de pandemia COVID-19 – SIGELO, financiado por el

Ministerio de Ciencia, Tecnología e Innovación de Colombia en el marco de la “Invitación a presentar proyectos que contribuyan a la solución de problemáticas actuales de salud relacionadas con la pandemia de COVID-19”, en la convocatoria 2020. código del proyecto 1106101577569. Proyecto finalizado.

REFERENCIAS

- Aballay, L., Aciar, S. & Reategui, E. (2017). Método para detección de emociones desde foros utilizando Text Mining. *Campus Virtuales*, 6(1), 89-98.
- Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E., Gutiérrez, J. & Kochut, K. (2017). A brief survey of text mining: Classification, clustering and extraction techniques. arXiv preprint arXiv:1707.02919 <https://bit.ly/3jLuZwD>
- Atalan, A. (2020). Is the lockdown important to prevent the COVID-19 pandemic? Effects on psychology, environment and economy-perspective. *Annals of Medicine and Surgery*, 56, 38-42. <https://doi.org/10.1016/j.amsu.2020.06.010>
- Baker, Q., Shatnawi, F., Rawashdeh, S., Al-Smadi, M. & Jararweh, Y. (2020). Detecting epidemic diseases using sentiment analysis of Arabic tweets. *Journal of Universal Computer Science*, 26(1), 50-70. <https://bit.ly/3oPhr6S>
- Bastian, M., Heymann, S. & Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks. *International AAAI Conference on Weblogs and Social Media*, 361-362. <https://bit.ly/2TFXcKO>
- Botes, W. & Thaldar, D. (2020). COVID-19 and quarantine orders: A practical approach. *SAMJ: South African Medical Journal*, 110(6), 1-4. <https://bit.ly/34MwOFx>
- Cambria, E., Grassi, M., Hussain, A. & Havasi, C. (2012). Sentic Computing for social media marketing. *Multimed Tools Appl* 59, 557-577. <https://doi.org/10.1007/s11042-011-0815-0>
- Cano, M. & Arce, S. (2020). Análisis de la comunicación en redes sociales de la campaña de la vacuna de gripe en España. *Revista Española de Salud Pública*, 94(1), 1-10. <https://bit.ly/3eeAG4Y>
- Crokidakis, N. (2020). COVID-19 spreading in Rio de Janeiro, Brazil: Do the policies of social isolation really work? *Chaos, Solitons & Fractals*, 136, 1-6. <https://doi.org/10.1016/j.chaos.2020.109930>
- Demidova, L. & Klyueva, I. (2017). Improving the classification quality of the SVM classifier for the imbalanced datasets on the base of ideas the SMOTE algorithm. *ITM Web of Conferences*, 10, 8-11. <https://doi.org/10.1051/itmconf/20171002002>
- Díaz, L. A. & Gutiérrez, E. (2020). La comunicación gubernamental a través de la red social Facebook en tiempos de coronavirus. Análisis del caso de Bahía Blanca, Argentina. *GIGAPP Estudios Working Papers*, 7(118), 609-626. <http://www.gigapp.org/ewp/index.php/GIGAPP-EWP/article/view/223/237>
- Dinero. (19 de mayo de 2020a). *Duque extiende el aislamiento*. <https://bit.ly/2HWNd0N>
- Dinero. (28 de mayo de 2020b). *Gobierno decretó aislamiento obligatorio hasta el 1 de julio*. <https://bit.ly/2TQOrz7>
- Dinero. (23 de junio de 2020c). *Colombia extiende cuarentena hasta el 15 de julio*. <https://bit.ly/3kPxWgS>
- Dinero. (7 de julio de 2020d). *Cuarentena en Colombia se extiende al 1 de agosto*. <https://bit.ly/3231oZo>
- Dinero. (28 de julio de 2020e). *Gobierno extiende aislamiento hasta el 30 de agosto*. <https://bit.ly/34LRgGd>
- El Espectador. (18 de agosto de 2020). *En imágenes: las protestas de los comerciantes en contra de la cuarentena por localidades, en Bogotá*. <https://www.elespectador.com/noticias/bogota/protesta-de-comerciantes-contracuarentena-por-localidades-en-bogota/>
- El Tiempo. (28 de julio de 2020a). *Cuarentena en Colombia: Duque amplía aislamiento hasta el 30 de agosto*. <https://bit.ly/2JsAiVl>
- El Tiempo. (24 de mayo de 2020b). *Este es el decreto con el que se extiende el aislamiento hasta mayo 31*. <https://bit.ly/3kOfyoO>

- Ferreira, S. G., Nieto, A. A. & Juares, W. I. (2020). Mar del Plata en Twitter: comunidades y tópicos durante la cuarentena. *Enlace Universitario*, 35(7), 15. <https://ri.conicet.gov.ar/handle/11336/116372>
- Han, J., Pei, J. & Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier. <http://myweb.sabanciuniv.edu/rdehkharghani/files/2016/02/The-Morgan-Kaufmann-Series-in-Data-Management-Systems-Jiawei-Han-Micheline-Kamber-Jian-Pei-Data-Mining-Concepts-and-Techniques-3rd-Edition-Morgan-Kaufmann-2011.pdf>
- Huang, Y., Wang, Y., Wang, H., Liu, Z., Yu, X., Yan, J., Yu, Y., Kou, C., Xu, X., Lu, J., Wang, Z., He, S., Xu, Y., He, Y., Li, T., Guo, W., Tian, H., Xu, G., Xu, X., ... Wu, Y. (2019). Prevalence of mental disorders in China: a cross-sectional epidemiological study. *The Lancet Psychiatry*, 6(3), 211-224. [10.1016/s2215-0366\(18\)30511-x](https://doi.org/10.1016/s2215-0366(18)30511-x)
- Jenson, H. (2020). How did “flatten the curve” become “flatten the economy?” A perspective from the United States of America. *Asian Journal of Psychiatry*, 51, 102165. <https://doi.org/10.1016/j.ajp.2020.102165>
- Kearney, M. & Kearney, M. (2016). Package ‘rtweet’. CRAN. <https://bit.ly/3ed6do4>
- Kotu, V. & Deshpande, B. (2014). *Predictive analytics and data mining: concepts and practice with rapidminer*. Morgan Kaufmann.
- Lin, C. Y., Liaw, S. Y., Chen, C. C., Pai, M. Y. & Chen, Y. M. (2017). A computer-based approach for analyzing consumer demands in electronic word-of-mouth. *Electronic Markets*, 27(3), 225-242. <https://doi.org/10.1007/s12525-017-0262-5>
- Luhn, H. (1957). A statistical approach to mechanized encoding and searching of literary information. *IBM Journal of Research and Development*, 1(4), 309-317. <https://doi.org/10.1147/rd.14.0309>
- Luo, X., Estill, J., Oi, W., Meng, L., Yunlan, L., Enmei, L. & Yao-long, C. (2020). The psychological impact of quarantine on coronavirus disease 2019 (COVID-19). *Psychiatry Research*, 291. <https://bit.ly/34MzanF>
- Manning, C., Raghavan, P. & Schütze, H. (2008). *Text Classification and Naive Bayes*. [Diapositiva de PowerPoint]. Web Stanford. <https://stanford.io/2TFZM3k>
- Marín Correa, A. (20 de julio de 2020). *Razones a favor y en contra de la cuarentena total en Bogotá*. El Espectador. <https://bit.ly/2HNAL3z>
- Martínez, E., Matin, M., Perea, J. & Ureña, A. (2011). Técnicas de clasificación de opiniones aplicadas a un corpus en español. *Procesamiento Del Lenguaje Natural*, 47, 163-170.
- Meier, K., Glatz, T., Guijt, M., Piccininni, M., van der Meulen, M., Atmar, K., Jolink, A., Kurth, T., Rohmann, J. & Zamanipour, A. (2020). Public perspectives on protective measures during the COVID-19 pandemic in the Netherlands, Germany and Italy: A survey study. *PloS One*, 15(8), 1-17. <https://doi.org/10.1371/journal.pone.0236917>
- Ministerio de Salud. (6 de marzo de 2020). *Colombia confirma su primer caso de COVID-19*. <https://www.minsalud.gov.co/Paginas/Colombia-confirma-su-primero-caso-de-COVID-19.aspx>
- Mohammad, S. & Bravo, F. (2017). WASSA-2017 shared task on emotion intensity. *8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, 34-49. <https://doi.org/10.18653/v1/w17-5205>
- Montiel, R., García, R., Ledeneva, Y. & Cruz, R. (2009). Comparación de tres modelos de texto para la generación automática de resúmenes. *Procesamiento de Lenguaje Natural*, 43, 303-311. <https://bit.ly/3kNlcGD>
- Mukherjee, R. (2020). Global efforts on vaccines for COVID-19: Since, sooner or later, we all will catch the coronavirus. *Journal of Biosciences*, 45(1). <https://doi.org/10.1007/s12038-020-00040-7>
- Musinguzi, G. & Asamoah, B. (2020). The science of social distancing and total lock down: Does it work? whom does it benefit? *Electronic Journal of General Medicine*, 17(6), 2019-2021. <https://doi.org/10.29333/ejgm/7895>
- Nava, A. & Grigera, J. F. (2020). *Pandemia y protesta social*. Jacobin Press. <https://ri.conicet.gov.ar/handle/11336/116379>
- Nicola, M., Alsafi, Z., Sohrabi, C., Kerwan, A., Al-Jabir, A., Iosifidis, C., Agha, M. & Agha, R. (2020). The socio-economic implications of the coronavirus pandemic (COVID-19): A review. *International Journal of Surgery* (78), 185-193. <https://doi.org/10.1016/j.ijsu.2020.04.018>
- Organización Mundial de la Salud [OMS]. (27 de abril de 2020a). *Coronavirus disease (COVID-19) pandemic*. <https://bit.ly/3oNVLyu>
- Organización Mundial de la Salud [OMS]. (2020b). *WHO Coronavirus Disease (COVID-19) Dashboard*. <https://bit.ly/3kNk1YT>
- Organización Internacional del Trabajo. (18 de marzo de 2020). *El COVID-19 podría cobrarse casi 25 millones de empleos en el mundo*. <https://bit.ly/36jLPQ2>
- Pang, B. & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and trends in Information Retrieval*, 2(1-2), 1-135. <https://doi.org/10.1561/15000000011>
- Pratama, B., Utami, E. & Sunyoto, A. (2019). An optimization of a lexicon based sentiment analysis method on Indonesian app review. *4th International Conference on Information Technology, Information Systems and Electrical Engineering*, 341-346. <https://doi.org/10.1109/ICITISEE48480.2019.9003900>
- Peters, M. (2020). The disorder of things: Quarantine unemployment, the decline of neoliberalism, and the Covid-19 lockdown crash. *Educational Philosophy and Theory*, 1-4. <https://doi.org/10.1080/00131857.2020.1759190>
- Pérez, A., Larrañaga, P. & Inza, I. (2009). Bayesian classifiers based on kernel density estimation: Flexible classifiers. *International Journal of Approximate Reasoning*, 50(2), 341-362. <https://doi.org/10.1016/j.ijar.2008.08.008>
- Ramírez-Ortiz, J., Castro-Quintero, D., Lerma-Córdoba, C., Yella-Ceballos, F. & Escobar-Córdoba, F. (2020). Consecuencias de la pandemia COVID-19 en la salud mental asociadas al aislamiento social. *Colombian Journal of Anesthesiology*, 48(4). <https://doi.org/10.5554/22562087.e930>
- Reddy, P., Selvaraj, S., Muralidharan, K. & Gangadhar, B. (2020). Tele-triaging: The way ahead for tertiary care psychiatry in India post-COVID-19. *Indian Journal of Psychological Medicine*, 42(4), 398-399. <https://doi.org/10.1177/0253717620937974>
- Reyes-Menendez, A., Saura, J. R. & Thomas, S. B. (2020). Exploring key indicators of social identity in the #MeToo era: Using discourse analysis in UGC. *International Journal of Information Management*, 54, 102129.
- Sáez, A. (2019). *Deep learning para el reconocimiento facial de emociones básicas* [Tesis de licenciatura no publicada] Universidad Politécnica de Cataluña.
- Saura, J. R. (2020). Using data sciences in digital marketing: Framework, methods, and performance metrics. *Journal of Innovation & Knowledge*.
- Saura, J. R., Palos-Sanchez, P. & Grilo, A. (2019). Detecting indicators for startup business success: Sentiment analysis using text data mining. *Sustainability*, 11(3), 917.
- Saura, J. R., Reyes-Menendez, A. & Palos-Sanchez, P. (2019). Are black Friday deals worth it? Mining Twitter users' sentiment and behavior response. *Journal of Open Innovation: Technology, Market, and Complexity*, 5(3), 58.
- Sauter, D., Eisner, F., Ekman, P. & Scott, S. (2010). Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proceedings of the National Academy of Sciences of the United States of America*, 107(6), 2408-2412. <https://doi.org/10.1073/pnas.0908239106>
- Shimizu, K. (2020). 2019-nCoV, fake news, and racism. *The Lancet*, 395(10225), 685-686. [https://doi.org/10.1016/S0140-6736\(20\)30357-3](https://doi.org/10.1016/S0140-6736(20)30357-3)
- Sigala, M. (2020). Tourism and COVID-19: Impacts and implications for advancing and resetting industry and research. *Journal of Business Research*, 117, 312-321. <https://doi.org/10.1016/j.jbusres.2020.06.015>
- Steffen, B., Egli, F., Pahle, M. & Schmidt, T. (2020). Navigating the clean energy transition in the COVID-19 crisis. *Joule*, 4(6), 1137-1141. <https://doi.org/10.1016/j.joule.2020.04.011>
- Teso, E., Olmedilla, M., Martínez, M. & Toral, S. (2018). Application of text mining techniques to the analysis of discourse in eWOM communications from a gender perspective. *Technological Forecasting and Social Change*, 129, 131-142. <https://doi.org/10.1016/j.techfore.2017.12.018>
- Van Bavel, J., Baicker, K., Boggio, P., Capraro, V., Cichocka, A., Cikara, M., Crockett, M., Crum, A., Douglas, K., Druckman, J., Drury, J., Dube, O., Ellemers, N., Finkel, E., Fowler, J., Gelfand, M., Han, S., Haslam, A., Jetten, J., ... Willer, R. (2020). Using social and behavioural science to support COVID-19 pandemic response. *Nature Human Behaviour*, 4(5), 460-471. <https://doi.org/10.1038/s41562-020-0884-z>

- Weible, C., Nohrstedt, D., Cairney, P., Carter, D., Crow, D., Durnová, A., Heikkilä, T., Ingold, K., McConnell, A. & Stone, D. (2020). COVID-19 and the policy sciences: initial reactions and perspectives. *Policy Sciences*, 53(2), 225–241. <https://doi.org/10.1007/s11077-020-09381-4>
- Williams, C. & Kayaoglu, A. (2020). COVID-19 and undeclared work: impacts and policy responses in Europe. *The Service Industries Journal*, 40(13-14), 914-931. <https://doi.org/10.1080/02642069.2020.1757073>
- Wongkar, M. & Angdressey, A. (2019). Sentiment analysis using naive bayes algorithm of the data crawler: Twitter. Fourth International Conference on Informatics and Computing (ICIC), 1-5.
- Zambrano, D., Román, D. & Zambrano, M. (2019). Innovation for the analysis of feelings in text, a revision of the current technique applying crowdsourcing methodologies. *Economía y Desarrollo*, 158(2), 138-146. <https://bit.ly/2TPuEyh>