



**UNIVERSIDAD POLITÉCNICA SALESIANA**

**SEDE QUITO**

**CARRERA DE COMPUTACIÓN**

**ANÁLISIS DE SENTIMIENTOS EN CONTRASTE CON DATOS  
OFICIALES DURANTE LA PANDEMIA DE COVID-19 EN ECUADOR**

**Trabajo de titulación previo a la obtención del  
Título de Ingeniero en Ciencias de la Computación**

**AUTORES:**

**ALISSON KAROLINA MENDOZA TORRES**

**JOEL NICOLÁS CARRASCO PILCO**

**TUTOR:**

**DIEGO FERNANDO VALLEJO HUANGA**

Quito - Ecuador

2022

# CERTIFICADO DE RESPONSABILIDAD Y AUTORÍA DEL TRABAJO DE TITULACIÓN

Nosotros, Alisson Karolina Mendoza Torres con documento de identificación N°1718574385 y Joel Nicolás Carrasco Pilco con documento de identificación N°1753497757; manifestamos que:

Somos los autores y responsables del presente trabajo; y, autorizamos a que sin fines de lucro la Universidad Politécnica Salesiana pueda usar, difundir, reproducir o publicar de manera total o parcial el presente trabajo de titulación.

Quito, 14 de marzo del año 2022

Atentamente,



.....  
Alisson Karolina Mendoza Torres  
1718574385



.....  
Joel Nicolás Carrasco Pilco  
1753497757

## **CERTIFICADO DE CESIÓN DE DERECHOS DE AUTOR DEL TRABAJO DE TITULACIÓN A LA UNIVERSIDAD POLITÉCNICA SALESIANA**

Nosotros, Alisson Karolina Mendoza Torres con documento de identificación N°1718574385 y Joel Nicolás Carrasco Pilco con documento de identificación N°1753497757, expresamos nuestra voluntad y por medio del presente documento cedemos a la Universidad Politécnica Salesiana la titularidad sobre los derechos patrimoniales en virtud de que somos autores del Artículo Académico: "Análisis de Sentimientos en Contraste con Datos Oficiales durante la Pandemia del COVID-19 en Ecuador", el cual ha sido desarrollado para optar por el título de: Ingeniero en Ciencias de la Computación, en la Universidad Politécnica Salesiana, quedando la Universidad facultada para ejercer plenamente los derechos cedidos anteriormente.

En concordancia con lo manifestado, suscribimos este documento en el momento que hacemos la entrega del trabajo final en formato digital a la Biblioteca de la Universidad Politécnica Salesiana.

Quito, 14 de marzo del año 2022

Atentamente,



.....  
Alisson Karolina Mendoza Torres  
1718574385



.....  
Joel Nicolás Carrasco Pilco  
1753497757

## CERTIFICADO DE DIRECCIÓN DEL TRABAJO DE TITULACIÓN

Yo, Diego Fernando Vallejo Huanga con documento de identificación N°1720162708, docente de la Universidad Politécnica Salesiana, declaro que bajo mi tutoría fue desarrollado el trabajo de titulación: ANÁLISIS DE SENTIMIENTOS EN CONTRASTE CON DATOS OFICIALES DURANTE LA PANDEMIA DE COVID-19 EN ECUADOR, realizado por Alisson Karolina Mendoza Torres con documento de identificación N°1718574385 y por Joel Nicolás Carrasco Pilco con documento de identificación N°1753497757, obteniendo como resultado final el trabajo de titulación bajo la opción Artículo Académico que cumple con todos los requisitos determinados por la Universidad Politécnica Salesiana.

Quito, 14 de marzo del año 2022

Atentamente,



.....  
Ing. Diego Fernando Vallejo Huanga, MSc  
1720162708

# Análisis de Sentimientos en Contraste con Datos Oficiales durante la Pandemia de COVID-19 en Ecuador

1<sup>st</sup> Alisson Karolina Mendoza Torres amendozat@est.ups.edu.ec 2<sup>nd</sup> Joel Nicolas Carrasco Pilco jcarrascopl@est.ups.edu.ec 3<sup>st</sup> Diego Fernando Vallejo Huanga dvallejoh@est.ups.edu.ec

**Resumen**—Ecuador fue uno de los primeros países latinoamericanos en tener un caso comprobado del nuevo coronavirus SARS-CoV-2. Las redes sociales fueron los medios más utilizados por la ciudadanía para replicar noticias sobre la pandemia, y emitir comentarios sobre el manejo de la crisis sanitaria. Este artículo tiene como objetivo presentar una herramienta web para el análisis de sentimientos en Twitter con tres formas diferentes de analizar el corpus y las polaridades: un modelo basado en diccionario de palabras, un modelo de aprendizaje automático supervisado entrenado a medida y una biblioteca de código abierto para procesar datos textuales y permite obtener una métrica de polaridad a partir de un tweet. Luego, para definir la polaridad final de cada tweet, se utiliza un modelo de aprendizaje automático conjunto para combinar las predicciones de las tres técnicas a través de una votación por mayoría. El sistema web fue desarrollado con herramientas de software libre y se acompaña de visualizaciones y gráficos estadísticos.

**Palabras Clave**—Twitter, Procesamiento del lenguaje natural, TextBlob, SARS-CoV-2, Análisis textual.

**Abstract**—Ecuador was one of the first Latin American countries to have a proven case of the new coronavirus SARS-CoV-2. The social networks were the media most used by citizens to replicate news about the pandemic, and issue comments about the handling of the health crisis. This article aims to present a web tool for sentiment analysis on Twitter with three different ways to analyze the corpus and polarities: a word-dictionary-based model, a custom trained supervised machine learning model, and an open-source library to process textual data and allows obtaining a polarity metric from a tweet. Then, to define the final polarity of each tweet, an ensemble machine learning model is used for combining the predictions from the three techniques through a hard majority voting ensemble. The web system was developed with free software tools and is accompanied by visualizations and statistical graphics.

**Keywords**—Twitter, Natural Language Processing, TextBlob, SARS-CoV-2, Textual Analytics.

## I. INTRODUCCIÓN

En marzo de 2020, la Organización Mundial de la Salud (OMS) declaró oficialmente una pandemia por el virus SARS-CoV-2 que causa la enfermedad Covid-19 [1]. El virus que se originó en Wuhan en la provincia de Hubei en China, hasta mediados de abril de 2021, ha causado alrededor de 136 millones de infecciones y 2.94 millones de muertes en todo el mundo [2]. Debido a la rápida propagación de la enfermedad y su alta tasa de contagio, muchos países cerraron sus fronteras para prevenir la propagación de la enfermedad y pusieron en cuarentena a los ciudadanos, limitando la vida diaria de

las personas. La nueva enfermedad viral ha provocado un cambio de paradigma en todos los dominios y disciplinas, afectando a millones de personas en todo el mundo, directa o indirectamente, y la epidemia se ha convertido en el evento de salud pública más grave que afecta a la humanidad en el siglo XXI.

Ecuador fue uno de los primeros países latinoamericanos en tener un caso positivo de la nueva variante del coronavirus [3], el 26 de febrero de 2020, en la ciudad de Guayaquil. El Gobierno ecuatoriano cerró sus fronteras terrestres, aéreas y marítimas y se implementaron varias restricciones al tránsito vehicular y peatonal. Sin embargo, a pesar de estas medidas, la cantidad de casos colapsó la infraestructura hospitalaria y desbordó las instituciones de salud.

En los primeros días de la pandemia, las agencias gubernamentales alertaron sobre la peligrosidad del nuevo virus y se puso de manifiesto el desconocimiento del gobierno ecuatoriano en el tratamiento de salud de la nueva enfermedad. Las instancias oficiales publicaron cifras de personas infectadas, recuperadas y fallecidas por el virus SARS-CoV-2 que se alejaban de la realidad en varias ciudades del país. Algunos medios internacionales se hicieron eco de las cifras divergentes [3].

En este contexto, muchos medios independientes y la ciudadanía utilizaron las redes sociales, especialmente Twitter, para poder difundir imágenes catastróficas de la realidad de varias ciudades del país. Las redes sociales fueron uno de los medios de comunicación más utilizados por los ciudadanos para expresar sus sentimientos, replicar noticias, emitir comentarios sobre el manejo de la crisis de salud en Ecuador. El Instituto Nacional de Estadísticas y Censos del Ecuador (INEC) junto con el Registro Civil del Ecuador, publicaron datos sobre el número de nuevas muertes en el país, lo que mostró que hubo un subregistro en el número de personas que fallecieron por la pandemia [4] y esto provocó que muchos ciudadanos expresaran su descontento a través de las redes sociales.

Por otro lado, el análisis de sentimientos, también conocido como minería de opiniones o análisis de emociones, es un conjunto de técnicas computacionales que permiten analizar la polaridad y sentimientos de una opinión, actitud o comentario expuesto en una red social, foro, blog, etc. Actualmente, las plataformas de redes sociales, como el *microblogging* Twitter,

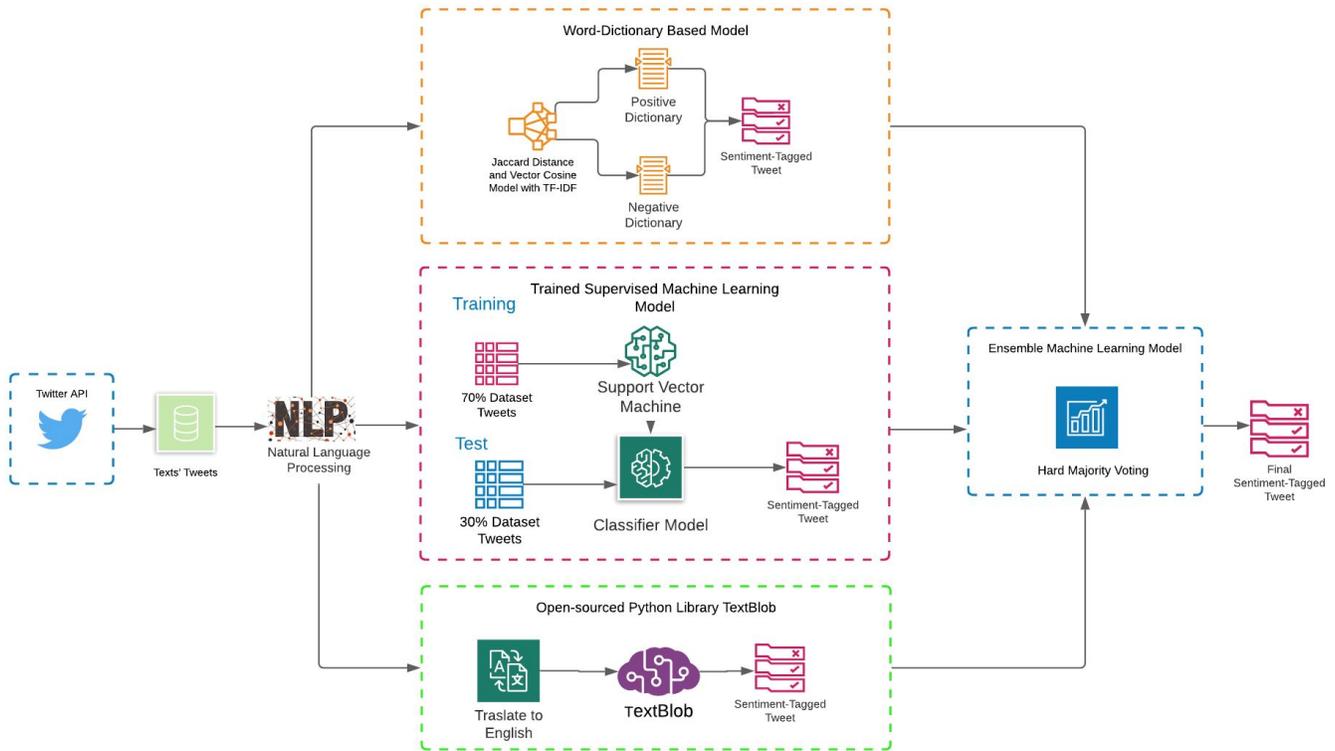


Fig. 1: Diagrama de bloques de la metodología para el despliegue de la herramienta web.

por ejemplo, con más de 300 millones de usuarios mensuales, son de inmensa importancia para la vida cotidiana de las personas debido a la amplitud de temas tratados. Twitter es una opción a la que recurren los investigadores para el análisis de sentimientos, debido a sus políticas de privacidad, la versatilidad de uso de su Interfaz de Programación de Aplicaciones (API) y los costos asociados con la explotación de datos [5].

En todo el mundo, y debido a la facilidad de explotar los datos de las redes sociales a través de una conexión con la API o técnicas de web *scraping*, se han desarrollado algunas investigaciones con análisis de sentimientos en el contexto de la pandemia causada por el COVID-19 [6] [7] [8].

Existen varios trabajos relacionados con análisis de sentimientos en el contexto ecuatoriano. Algunos de ellos han sido desarrollados como herramientas de propósito general para el análisis de sentimientos [9], otros han analizado la polaridad del sentimiento en un contexto específico como la educación [10], la gestión política [11] y el desarrollo de eventos masivos [12]. Después de una revisión sistemática de la literatura científica, no se ha encontrado ninguna herramienta web que permita comparar los datos oficiales de la pandemia en contraste con un análisis de sentimientos.

Este artículo presenta una herramienta web, desarrollada para contrastar los datos oficiales de la pandemia: personas infectadas, recuperadas y muertas frente a los comentarios relacionados con COVID-19, en la red social Twitter en una

misma línea de tiempo. La herramienta analiza la polaridad de las opiniones y comentarios de los *tweets* emitidos, en un período de tiempo, por múltiples usuarios de la red social. El sistema de recuperación de información se limita a la búsqueda de *tweets* escritos en español, ya que este es el idioma oficial del país.

## II. MATERIALES Y MÉTODOS

### A. Esquema metodológico

La Figura 1 muestra un diagrama general de bloques de la metodología utilizada en el desarrollo de la herramienta web.

En primera instancia, la aplicación web se conecta a la API de Twitter y descarga los textos de *tweets* relacionados con COVID-19 (o palabras similares) en Ecuador. De forma pre-determinada, la aplicación solo puede acceder a la información pública de Twitter, utilizando los códigos de acceso: clave del consumidor, clave secreta del consumidor, token de acceso y token secreto de acceso. La herramienta utiliza las potencialidades gratuitas de la API que ofrece para los desarrolladores de software. La colección total de *tweets* recuperados  $T_m$ , con longitud  $m$ , depende de las restricciones en la solicitud que el usuario ha seleccionado en la interfaz gráfica, es decir, el intervalo de fechas o el número de *tweets* a recuperar. Cada *tweet*  $t_i (i = 1, 2, \dots, m)$  tiene un corpus textual de  $n$  palabras de longitud que depende del usuario que redactó y envió el mensaje. El  $t_i$  es tratado como una entidad independiente,

donde su polaridad será analizada individualmente, a través del estudio de su corpus.

En el pre-procesamiento lingüístico de cada entidad  $t_i$ , se realizan operaciones de limpieza y normalización, para unificar los criterios en el procesamiento del texto contenido en cada *tweet*. En primera instancia, se llevó a cabo la eliminación de caracteres especiales, signos de puntuación, emojis y emoticones, reduciendo la longitud de las palabras de  $n$  a  $k$ . Luego, se ejecutó la tokenización del *tweet* y la conversión a letras minúsculas de los  $k$  tokens extraídos ( $t_{i_k}$ ) de  $t_i$ . Además, en cada *tweet* tokenizado se eliminan las palabras vacías del idioma español y finalmente se ejecuta un proceso de *stemming* en los tokens, utilizando el algoritmo de Porter, para eliminar los sufijos y mejorar el proceso de NLP.

Para el análisis de sentimientos, la herramienta propuesta utiliza tres diferentes formas para analizar el corpus y las polaridades de los *tweets*. El primero utiliza un modelo basado en un diccionario de palabras, donde se recogió un conjunto de aproximadamente 4000 tokens en un diccionario  $D$ , que incluye palabras con polaridad positiva  $d_{j(+)}$  y polaridad negativa  $d_{j(-)}$ , en español y adaptadas al contexto ecuatoriano. De esta forma, al recuperar un nuevo *tweet*, se realiza el análisis de sentimientos, en base al corpus del *tweet* y su similitud con el diccionario de sentimientos. En esta primera metodología, se utilizaron dos métricas de similitud para definir la etiqueta de polaridad del *tweet*  $L_{t_i}$ , que puede ser positiva  $L_{t_i(+)}$ , negativa  $L_{t_i(-)}$  o neutra  $L_{t_i(\pm)}$ . Las métricas de similitud son el coeficiente de Jaccard y el coseno vectorial con TF-IDF.

El coeficiente de Jaccard mide la similitud normalizada entre dos conjuntos, donde los valores cercanos a 0 indican menos similitud entre el token del *tweet*  $t_{i_k}$  y el token del diccionario  $d_j$ . Por otro lado, los valores cercanos a 1 indican una mayor similitud, como se muestra en la Ecuación 1. En este sentido, si una entidad  $t_i$  tiene un mayor grado de similitud con las palabras positivas  $d_{j(+)}$ , la etiqueta de polaridad del *tweet* es positivo  $L_{t_i(+)}$ , caso contrario el  $t_i$  se etiqueta como negativo  $L_{t_i(-)}$ . Si un token no está contenido dentro de  $D$ , se etiquetará como neutral.

$$J(t_i, D) = \frac{|t_i \cap D|}{|t_i \cup D|}; J \in [0, 1] \quad (1)$$

Luego, la etiqueta que define la polaridad del *tweet*,  $L_{t_i}$ , está determinada por la Ecuación 2. Cuando el número de tokens positivos es igual al número de tokens negativos, el  $L_{t_i}$  se etiqueta como neutro  $L_{t_i(\pm)}$ .

$$L_{t_i} = \begin{cases} \text{if } \sum t_{i_k(+)} > \sum t_{i_k(-)} \rightarrow L_{t_i(+)} \\ \text{if } \sum t_{i_k(+)} < \sum t_{i_k(-)} \rightarrow L_{t_i(-)} \\ \text{if } \sum t_{i_k(+)} = \sum t_{i_k(-)} \rightarrow L_{t_i(\pm)} \end{cases} \quad (2)$$

El método de similitud de coseno vectorial con TF-IDF se procesa a través de una bolsa de palabras formada por el diccionario  $D$  más los tokens recuperados de todos los *tweet*  $t_{i_k}$  de la colección  $T_m$ . En la bolsa de palabras, se aplica un proceso de ponderación TF-IDF y se calcula la divergencia con la similitud de coseno entre el vector de palabras positivas y el vector que representa el *tweet*  $C_{i(+)}$ . De la misma forma,

la divergencia también se mide con el vector de palabras negativas  $C_{i(-)}$ . Cuando,  $C_{i(+)} > C_{i(-)} \rightarrow L_{t_i(+)}$ , si  $C_{i(+)} < C_{i(-)} \rightarrow L_{t_i(-)}$  y si  $C_{i(+)} = C_{i(-)} \rightarrow L_{t_i(\pm)}$ . El proceso de etiquetado se ejecuta para todos los vectores de *tweets*.

La segunda técnica de análisis de sentimientos utiliza un modelo de aprendizaje automático supervisado. Para el entrenamiento del algoritmo, se creó un conjunto de datos específico para esta tarea. Se descargaron alrededor de 1500 *tweets* desde el 23 de mayo de 2020 hasta el 3 de junio de 2020 y los sentimientos expresados en el corpus se etiquetaron manualmente como positivos, negativos y neutros. El sistema utiliza el conjunto de datos con 500 *tweets* para cada clase y un clasificador multiclase mediante el algoritmo de Máquina de Soporte Vectorial [13]. El 70% de los datos se utilizaron para el entrenamiento y el 30% para la prueba, de forma aleatoria. El rendimiento del modelo obtuvo un *accuracy* de 0.744, un *precision* de 0.742, un *recall* de 0.751 y un *F1-measure* de 0.741. Una vez validado el modelo, se lo implementó en el sistema web para que pueda ejecutar la tarea de clasificación de sentimientos cuando el usuario ejecuta una nueva búsqueda y recuperación de una  $C_m$  colección de *tweets*.

El tercer método para el análisis de polaridad de sentimientos utiliza una biblioteca de código abierto de Python, llamada *TextBlob* [14], para procesar datos textuales y obtener una métrica de polaridad. La biblioteca está diseñada para trabajar con textos en inglés, por lo que nuestra herramienta realiza ciertas adaptaciones para procesar texto en español.

Finalmente, para definir la polaridad de cada *tweet*, se utiliza un método *ensemble* de aprendizaje automático que combina las predicciones de las tres técnicas a través de una votación por mayoría. Ya que no existe evidencia empírica de qué técnica funciona mejor, todos los métodos agregan el mismo peso a la decisión final de polaridad  $L_{pol}$ . Adicionalmente, aunque no se utiliza para la decisión de la polaridad del *tweet*, la herramienta web ejecuta una *Latent Dirichlet Allocation* (LDA) [15] con una visualización de las polaridades de los sentimientos.

Los conjuntos de datos y el código fuente de la aplicación se pueden consultar en [https://github.com/dievalhu/Sentiment\\_Analysis\\_COVID19](https://github.com/dievalhu/Sentiment_Analysis_COVID19) y la versión web de la herramienta está disponible en <https://analisis-sentimiento.herokuapp.com/>.

## B. Detalles de implementación y arquitectura

La metodología de Ingeniería Web basada en UML (UWE) [16] se utiliza para el desarrollo de la aplicación web y la herramienta está diseñada bajo la arquitectura de software Modelo Vista Controlador (MVC) [17] que separa los datos de la aplicación, la interfaz de usuario y la lógica de control en tres componentes diferentes como se muestra en la Figura 2. De esta manera, el *back-end* controla el área lógica de la herramienta web, mientras que el *front-end* es la sección destinada al usuario que incluye la línea de diseño y los elementos gráficos de la web.

El componente Modelo contiene una representación de los datos manejados por el sistema, es decir, el conjunto de datos

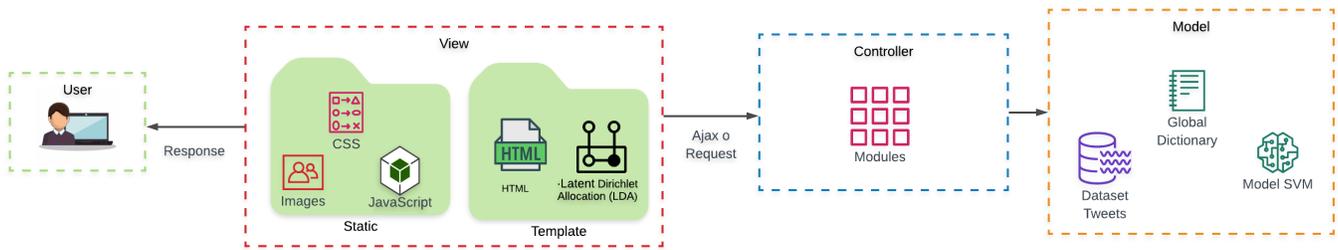


Fig. 2: Patrón de arquitectura MVC de la herramienta web de análisis de sentimientos.

de los *tweets* con etiquetado manual, el diccionario de tokens para el modelo basado en diccionario de palabras y el modelo SVM entrenado que se almacena en un archivo SVM.pkl. La Vista, o interfaz de usuario, contiene la información que se envía al cliente y sus mecanismos de interacción a través de dos secciones. La primera es una sección *Estática* que contiene todos los estilos, imágenes y archivos CSS y JavaScript que componen la herramienta web. La segunda sección, *Plantillas*, almacena los archivos HTML de la web y la sección que muestra los gráficos generados por LDA.

El Controlador es un intermediario entre el Modelo y la Vista, que gestiona el flujo de información y las transformaciones para adaptar los datos. Está dividido por los módulos: conexión API COVID-19, coeficiente de Jaccard, cálculo de similitud coseno con TF-IDF, escritura y lectura de archivos de texto plano (.csv y .txt), algoritmo SVM, librería Textblob, modelado de temas, recuperación de texto de los *tweets*, algoritmo *ensemble* de votación mayoritaria, preprocesamiento de los textos con NLP y el principal llamado *Procesos* que controla todo el módulo anterior y los ejecuta uno por uno.

Nuestra aplicación web utiliza Python 3.8, en la arquitectura MVC, para que los algoritmos de aprendizaje automático del Controlador se ejecuten junto con los datos extraídos del Modelo y envíen los resultados a la Vista. La aplicación web se desarrolló con el *framework* Flask.

### III. DESCRIPCIÓN Y EVALUACIÓN DE LA HERRAMIENTA WEB

El *front-end* de nuestra herramienta web está diseñado para que el flujo de información y solicitudes a los servicios de *back-end* se ejecuten de forma transparente para el usuario. Así, la herramienta inicia su proceso haciendo clic en *Buscar*, donde se realiza la solicitud al servidor a través de la biblioteca Ajax, que procesa la solicitud al servidor en segundo plano. El servidor procesa la solicitud y devuelve los *tweets* extraídos de la API, con sus metadatos. Una vez que se han recibido los datos, se utiliza la biblioteca de código abierto *JQuery JavaScript* para que el Modelo de objetos de documento (DOM) inyecte datos donde fuese necesario. El DOM funciona tanto para los módulos que se encuentran en el controlador como para su representación gráfica interactiva que emplea la biblioteca de *JavaScript Highcharts*.

Se utilizó el *framework* Bootstrap que crea las interfaces web con CSS y JavaScript y utiliza la técnica de diseño y de-

sarrollo, web *responsive*. La interfaz gráfica de la herramienta web tiene tres secciones. La primera sección configura los parámetros temporales y cuantitativos para la conexión con la API de Twitter, entre estos *# Tweets*: número de *tweets* a recuperar de la API; *Fecha de Inicio*: fecha inicial de recuperación de *tweets* y *Fecha de Finalización*: fecha final de recuperación de *tweets*.

La segunda sección ejecuta los modelos de Inteligencia Artificial para encontrar la polaridad de los *tweets* junto con gráficos estadísticos de los resultados. Esta sección está compuesta por tres pestañas: *Análisis de Sentimientos*: tabla resumen con los porcentajes de *tweets* positivos, negativos y neutros para cada método de análisis de sentimientos; *Gráficos Estadísticos*: gráficos porcentuales de los resultados de los algoritmos de análisis de sentimientos; *Tweets*: tabla con el corpus de *tweets* recuperados, cada uno con su etiqueta de fecha y polaridad de sentimiento.

La tercera sección, denominada visualización, contiene las pestañas: *Nube de palabras*: representación de las palabras más recurrentes en los *tweets* extraídos de la API en forma de nube de palabras; *LDA*: representación visual interactiva del modelo LDA; *Covid-19 vs. Sentimientos*: gráficos del análisis de sentimientos con cada método (ver Figura 3), donde la línea de tiempo se ubica en el eje de las abscisas y la polaridad de sentimiento en el eje de las ordenadas: positivo (1), negativo (-1) o neutro (0).

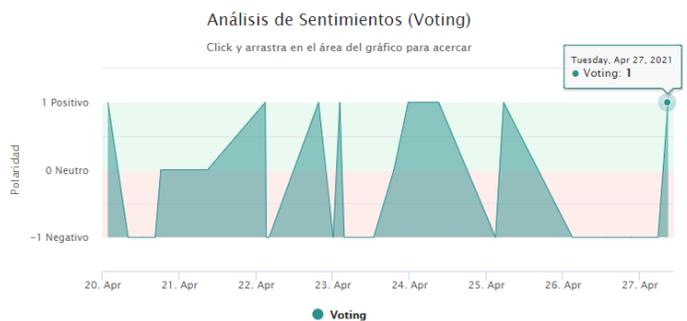


Fig. 3: Análisis de sentimientos con método *ensemble* de votación mayoritaria desde el 20 de abril de 2021 hasta el 27 de abril de 2021.

Además, en la pestaña *Covid-19 Vs. Sentimientos*, en la misma línea de tiempo, los datos recopilados de [2] y extraídos

desde el comienzo de la pandemia sobre personas muertas, recuperadas, confirmadas y activas por COVID-19 se grafican frente al resultado general del análisis de sentimientos. Así, la herramienta permite monitorear el progreso de la pandemia de COVID-19 en Ecuador y contrarrestar estos datos con los sentimientos generalizados de la población en Twitter, como se muestra en la Figura 4.

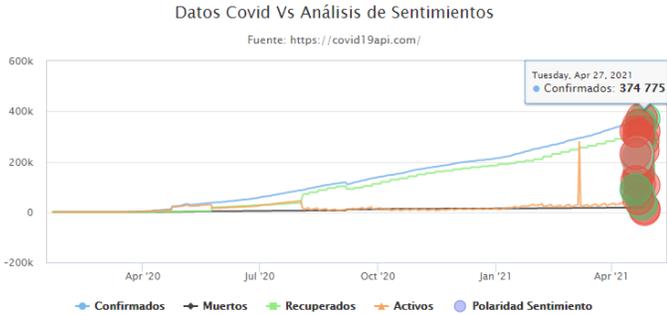


Fig. 4: Cronología de los datos oficiales de la pandemia (personas confirmadas, muertas, recuperadas y activas) y el resultado del análisis de sentimientos de abril de 2020 a abril de 2021, en Ecuador.

Finalmente, se han realizado pruebas de estrés y carga de la herramienta web, para conocer su comportamiento con diferentes números de usuarios, los límites del proveedor de *hosting*, la tasa de incremento de los servicios de la web hasta su punto de ruptura y posibles fallos de la herramienta web. El desempeño fue monitoreado con la herramienta de código abierto JMeter, desarrollada en Java, que permite realizar pruebas de comportamiento funcional y medición del desempeño de la herramienta web alojada en Heroku [18]. La Tabla I resume los resultados para 10 y 100 muestras, los valores medios de tiempo y desviación estándar, el rendimiento definido como el número de solicitudes por tiempo total y el porcentaje de error.

URL	Número de muestras		$\mu + \sigma$ (seg)		Rendimiento (muestras/seg)		Error (%)	
	Carga	Estrés	Carga	Estrés	Carga	Estrés	Carga	Estrés
/	10	100	2.61±0.075	5.13±1.705	2.8	8.2	0	0
/lit1	10	100	8.73±3.111	28.07±10.080	0.68	2.3	0	49
/topic	10	100	3.37±0.175	6.58±1.680	2.4	8.8	0	0
/api	10	100	4.36±0.592	18.89±5.545	1.8	3.6	20	70

TABLE I: Resultados de rendimiento de las pruebas de estrés y carga de la herramienta web.

#### IV. CONCLUSIONES Y TRABAJOS FUTUROS

En este artículo científico, se ha discutido tres formas diferentes de desarrollar un modelo de análisis de sentimientos en una herramienta web. En trabajos futuros, se pretende mejorar los resultados mediante el uso de conjuntos de datos más grandes e incluyendo el almacenamiento en una base de datos en la nube. En este artículo se utilizaron algoritmos y técnicas para el análisis de sentimientos de un *tweet*, sin embargo, el enfoque se puede ampliar para comparar cada resultado con el *ground truth* y descubrir qué algoritmo tiene mejor rendimiento. Además, en un enfoque de clasificación

multiclase, la cantidad de categorías se puede aumentar para capturar los sentimientos de los *tweets* relacionados con COVID-19 con mayor precisión.

#### REFERENCES

- [1] F. Di Gennaro, D. Pizzol, C. Marotta, M. Antunes, V. Racalbutto, N. Veronese, and L. Smith, "Coronavirus diseases (covid-19) current status and future perspectives: a narrative review," *International journal of environmental research and public health*, vol. 17, no. 8, p. 2690, 2020.
- [2] "Covid 19 api: A free api for data on the coronavirus." [Online]. Available: <https://covid19api.com/>
- [3] A. Luque, F. Maniglio, F. Casado, and J. García Guerrero, "Transmedia context and twitter as conditioning the ecuadorian government's action: the case of the "guayaquil emergency" during the covid-19 pandemic," *Tripodos. Facultat de Comunicació i Relacions Internacionals Blanquerna-URL*, vol. 2, no. 47, pp. 47–68, 2020.
- [4] I. Torres and F. Sacoto, "Localising an asset based covid-19 response in ecuador," *The Lancet*, vol. 395, no. 10233, p. 1339, 2020.
- [5] E. Kouloumpis, T. W., and J. M., "Twitter sentiment analysis: The good the bad and the omg!" *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 5, no. 1, 2011.
- [6] G. Barkur and G. Vibha, "Sentiment analysis of nationwide lockdown due to covid 19 outbreak: Evidence from india," *Asian journal of psychiatry*, vol. 51, p. 102089, 2020.
- [7] J. Samuel, G. Ali, M. Rahman, E. Esawi, and Y. Samuel, "Covid-19 public sentiment insights and machine learning for tweets classification," *Information*, vol. 11, no. 6, p. 314, 2020.
- [8] T. Wang, K. Lu, K. Chow, and Q. Zhu, "Covid-19 sensing: Negative sentiment analysis on social media in china via bert model," *Ieee Access*, vol. 8, pp. 138 162 – 138 169, 2020.
- [9] I. Utitiáj, P. Morillo, and D. Vallejo Huanga, "Sentiment analysis tool for spanish tweets in the ecuadorian context," *3rd International Conference on Algorithms, Computing and Artificial Intelligence*, pp. 1 – 6, 2020, association for Computing Machinery, New York.
- [10] R. Pazmiño, F. Badillo, M. González, and F. García Peñalvo, "Ecuadorian higher education in covid-19: A sentiment analysis," *Information*, pp. 758 – 764, 2020.
- [11] L. Terán and J. Mancera, "Dynamic profiles using sentiment analysis and twitter data for voting advice applications. government information quarterly," vol. 36, no. 3, pp. 520 – 535, 2019.
- [12] R. Rivera-Guamán, R. López-Lapo, and L. Neyra-Romero, "Sentiment analysis related of international festival of living arts loja-ecuador employing knowledge discovery in tex," *Applied Technologies: Second International Conference, ICAT 2020*, vol. 1388, p. 327, 2020, springer Nature.
- [13] C. Cortes and V. Vapnik, "Support-vector networks. machine learning," vol. 20, no. 3, pp. 273 – 297, 1995.
- [14] S. Loria, "textblob documentation. release 0.16.0," 2020.
- [15] D. Blei, A. NG, and M. Jordan, "Latent dirichlet allocation. journal of machine learning research," vol. 3, pp. 993 – 1022, 2003.
- [16] N. De Koch, "Software engineering for adaptive hypermedia systems," 2001, ph.D Thesis, Verlag Uni-Druck, Munich.
- [17] T. Reenskaug, "The original mvc reports," 1979.
- [18] A. Hanjura, "Heroku cloud application development," 2014, packt Publishing Ltd.