



Tipo de artículo: Artículos originales
Temática: Inteligencia artificial
Recibido: 08/03/2023 | Aceptado: 15/06/2023 | Publicado: 30/09/2023

Identificadores persistentes:
DOI: [10.48168/innosoft.s12.a89](https://doi.org/10.48168/innosoft.s12.a89)
ARK: [ark:/42411/s12/a89](https://nbn-resolving.org/urn:nbn:org:ark:42411/s12/a89)
PURL: [42411/s12/a89](https://purl.org/urn:nbn:org:ark:42411/s12/a89)

Modelo de clasificación de depresión en Tweets usando BERT

Depression classification model on Twitter using BERT

Guillermo José Aleman-Zambrano ¹[0000-0001-5471-4226], Marvik Irzovic Del Carpio-Lazo ²[0000-0002-0019-2458], Daniel Gustavo Mendiguri-Chávez ³[0000-0002-0588-6520], Daniela Carolina Vilchez-Silva ⁴[0000-0002-7896-8228], Franco Tejada-Toledo ⁵[0000-0002-1675-7097]

¹ Universidad La Salle, Arequipa, Perú. galemanz@ulasalle.edu.pe

² Universidad La Salle, Arequipa, Perú. mdelcarpiol@ulasalle.edu.pe

³ Universidad La Salle, Arequipa, Perú. dmendiguric@ulasalle.edu.pe

⁴ Universidad La Salle, Arequipa, Perú. dvilchezs@ulasalle.edu.pe

⁵ Universidad La Salle, Arequipa, Perú. ftejadat@ulasalle.edu.pe

* Autor para correspondencia: dmendiguric@ulasalle.edu.pe

Resumen

Hoy en día existen muchos indicios de depresión, así como muchos intentos de suicidio causados por este trastorno emocional, esto se ve reflejado mayormente en redes sociales principalmente en Twitter. Por ello, es importante que los especialistas y organizaciones que busquen salvaguardar la vida de las personas, utilicen herramientas de software que permitan abordar este problema. Para ello, en este trabajo se propone una herramienta web llamada “UBDevs-Depression-Classifer” que permite clasificar y obtener tweets de forma automática por algún tema específico. Se puso un mayor énfasis a tweets relacionados con el COVID-19 debido a que en los años 2020-2021 en el mundo se vivió una pandemia que incrementó los casos de depresión en muchos lugares. Esta propuesta de investigación se centra en la utilización en un modelo basado en NLP (Natural Language Processing) para la clasificación de Tweets con el fin de encontrar aquellos que inciten a la depresión o den a entender que los usuarios se encuentren en un mal estado de ánimo, todo ello con el fin de mantener la salud mental y física de los usuarios de esta plataforma. Existen varios modelos usados como base para proyectos de NLP, sin embargo, en la actualidad BERT ha demostrado ser uno de los más eficientes por ello lo seleccionamos para el desarrollo de nuestra propuesta. Para evaluar la eficiencia del proyecto aplicamos la métrica F1 obteniendo un valor de 0.8806, resultado bastante aceptable respecto a una clasificación textual.

Palabras clave: Clasificación de depresión, clasificación de texto, procesamiento de lenguaje natural, BERT, redes sociales.

Abstract

Today there are many signs of depression, as well as many suicide attempts caused by this emotional disorder, and this is reflected mostly on social networks, mainly on Twitter. For this reason, it is important for specialists and organizations seeking to safeguard people's lives to use software tools to address this problem. For this, in this work a web tool called "UBDevs-Depression-Classifler" is proposed, that allows you to automatically obtain and classify tweets for a specific topic. A greater emphasis was placed on tweets related to COVID-19 in the years 2020-2021 the world experienced a pandemic that increased cases of depression in many places. This research proposal focuses on the use of a model based on NLP (Natural Language Processing) for the classification of Tweets in order to find those that incite depression or imply that users are in a bad mood, all this in order to maintain the mental and physical health of the users of this platform. There are several models that are used as a basis for NLP projects, however, at present BERT has proven to be one of the most efficient, so we selected it for the development of our proposal. To evaluate the efficiency of the project we applied the F1 metric obtaining a value of 0.8806, a quite acceptable result with respect to a textual classification.

Keywords: *Depression classification, text classification, natural language processing, BERT, social networks.*

Introducción

En la actualidad, debido al contexto social aislado, el estrés laboral y académico, la falta de recursos y estabilidad, entre otros factores han contribuido al aumento de depresión y sentimientos de soledad [1][2][3][4], forzando a muchos al suicidio y con mayor incidencia en jóvenes alrededor de los veinte años [5].

Por ello, es importante prevenir este tipo de casos a través de la detección temprana de personas con problemas de depresión, para que puedan ser asistidas por las autoridades competentes y salvar su vida, antes de formar parte del 60% al 70% de suicidios causados por depresión [6].

Analizar los sentimientos de la población a través del entendimiento de sus expresiones hace que este tema no sólo sea interesante, sino necesario en la lucha contra el problema global que significa una población depresiva, y el riesgo latente sobre la vida de aquellos que no encuentran ayuda a tiempo [7].

Bajo el contexto de la pandemia, el aislamiento social ha contribuido a que el uso de redes sociales incremente, donde se sabe que los jóvenes son quienes más lo utilizan, sobre todo con el fin de expresar pensamientos, comentarios, sentimientos y contenido [8].

Twitter es una de las redes sociales con mayor número de usuarios a nivel mundial [9], estos mayormente son jóvenes, los cuales encuentran en esta red social un lugar donde expresar ideas y sentimientos, que pueden ser tanto de alegría, enojo, tristeza e incluso depresión. Con tal fin se desea contribuir con la detección de personas emocionalmente

vulnerables por problemas emocionales, especialmente con tendencias depresivas quienes se expresan por Twitter. Precisamente esta red social será utilizada para extraer tweets con su respectiva información y posteriormente clasificarlos.

Existen muchos modelos para realizar el procesamiento de lenguaje natural, uno de los más prometedores actualmente es el modelo propuesto por Google llamado BERT el cual analiza el texto de forma bidireccional lo que permite que sea muy eficiente [19]. Una característica importante es que para lograr que este modelo logre un objetivo específico primero tiene que pasar por una fase de entrenamiento con un dataset. Así mismo para que BERT pueda entender el texto primero debemos realizar una tokenización y codificación de las palabras [14].

Con el fin de que la propuesta de este trabajo tenga buenos resultados se realizó una limpieza previa de datos irrelevantes en los tweets antes del proceso de tokenización y codificación.

Los resultados obtenidos podrán ser visualizados mediante una herramienta web la cual presentará gráficos estadísticos de tipo Pie Chart, mostrando que cantidad de tweets han sido clasificados como depresivos y no depresivos. Estos gráficos lograrán que los usuarios comprendan mejor los valores estadísticos obtenidos.

Trabajos relacionados

La información generada en internet día a día es inconmensurable por la cantidad de usuarios que registran datos de diferentes temas, en distintas plataformas alrededor del mundo. Kauffman habla de cómo la analítica del big data se está convirtiendo de forma acelerada en una práctica de tendencia adoptada por muchas organizaciones con el objetivo de obtener información valiosa [11].

En tal sentido, hablamos del poder de la data existente en internet y continuamente renovada por los usuarios, para que organizaciones privadas o de interés público puedan tomar decisiones sobre la población que le compete o sobre la que es responsable.

Un gran inconveniente que tuvo Kauffman es que al trabajar con distintos tipos de comentarios y calificaciones enfocados a analizar algo en concreto, obtuvo también otros que están fuera de lo que se desea analizar. Por ello Kauffman determinó que a la calificación se pueden sumar varios aspectos que pueden llegar a ser importantes para su investigación.

La propuesta de Back puede extraer información sobre los sentimientos en grandes cantidades de datos generados por usuarios (big data no estructurado de SNS) utilizando Naive Bayes y NLP, en su estudio realizan múltiples pruebas donde cada uno de estos métodos muestra su desempeño; llegan a la conclusión que NLP tiene un menor rango de error

pero tiende a demorar 5 veces más que el algoritmo de Bayes pero la precisión de Bayes es 63.5% , comparando con la precisión de NLP que es 72.28% entonces concluyen que si el objetivo es un sistema que procese una gran cantidad de datos la mejor opción es Bayes, pero si nuestro objetivo es tener un análisis confiable lo mejor es utilizar NLP [12]. Ya que este documento intenta enfocarse en la clasificación de texto mediante el uso de NLP es la mejor opción para poder obtener la mayor fiabilidad sobre la data estudiada.

Sobrino explica dos clases de métodos para la clasificación de textos enfocados en NLP:

A. Métodos basados en diccionarios

Para este método se hace uso de listados de palabras y frases ya etiquetadas con la polaridad del sentimiento que expresan, así como la intensidad o fuerza de dicho sentimiento.

Luego los textos a ser clasificados se dividen y se busca en los diccionarios para obtener su respectivo resultado. Ejemplo de este se da en el estudio de Taboada, donde haciendo uso de un diccionario previamente clasificado con polaridades y sus respectivas puntuaciones, luego el sistema extrae las palabras que tienen información de sentimiento mediante un análisis de adverbios, verbos y adjetivos, después se busca sus respectivas puntuaciones para así tener calcular el sentimiento global [10].

Algunas desventajas de este método es que una misma palabra puede contener diferentes polaridades lo cual puede llevar a resultados incorrectos ya que el contexto de la palabra dentro de la oración no está presente en este tipo de calificaciones. Otra desventaja es que se debe tener un gran diccionario con un inmenso contenido de palabras con calificaciones, el cual puede ser muchas veces difícil de construir y consultar. [10].

B. Métodos basados en relaciones lingüísticas

En este método se busca opiniones y sentimientos con mayor probabilidad para ser usados luego en la categorización del texto global, obteniendo también la categoría gramatical de las palabras, siendo esta la característica más importante [10].

El estudio de Hatzivassiloglou [23] quién hace uso de este método, afirma que dependiendo de los conectores usados entre dos palabras, dice si habrá polaridades opuestas o iguales.

Por ejemplo, si dos palabras tienen el conector “y”, esta tendrá la misma polaridad, por otro lado, con “pero” ocurrirá lo contrario, para ahorrar así análisis extra, luego con estos resultados se calculará el sentimiento global [10].

Luego de presentar los métodos explicados por Sobrino, se buscó enfocar el documento hacia el análisis de sentimientos. El estudio de Leis habla sobre la identificación de características lingüísticas en tweets en español y patrones de comportamiento de los usuarios de Twitter que podría sugerir signos de depresión, lo cual firma un precedente sobre el análisis de tweets en español [13].

En [13] indican que los trastornos mentales son una preocupación importante en la salud pública y son una de las principales causas de morbilidad en el mundo. Las plataformas de redes sociales nos permiten observar las interacciones, pensamientos y sentimientos de la vida diaria de las personas, incluidas las de los pacientes que padecen trastornos mentales. Existen estudios que han analizado la influencia de los trastornos mentales, incluida la depresión, en el comportamiento de los usuarios de las redes sociales, básicamente se han centrado en mensajes en inglés.

El estudio se centró en detectar características especiales acerca de los tweets depresivos en español. Esto se logró gracias al etiquetado de parte del discurso, Part-of-Speech (POS) realizado mediante la herramienta Freeing Natural Language Processing para analizar los patrones de uso de categorías gramaticales (p. Ej., Adjetivos, sustantivos o pronombres) en el texto de los tweets. Así mismo se contó de la confirmación y evaluación de especialistas de la salud para un análisis más detallado de los resultados de las frecuencias de parte del habla (POS) en tweets de control, usuarios depresivos y conjuntos de datos de tweets depresivos.

El estudio llegó a la conclusión que los usuarios de Twitter que sufren depresión modifican las características generales de su idioma y la forma en que interactúan en las redes sociales. En base a estos cambios, estos usuarios pueden ser monitoreados y apoyados, lo que presenta nuevas oportunidades para estudiar la depresión y brindar servicios de atención médica a las personas con este trastorno mental.

Adicionalmente a la intención de analizar el texto de las publicaciones, existe la necesidad de limpiar la información receptada. Puesto que todo lo que está en internet no es precisamente concreto llegando a tornarse muchas veces ambiguo y puede afectar la data recolectada y estudiada con falsas alarmas Kauffmann plantea seis clases de problemas al momento de validar la información que se encuentra en las redes sociales [11]:

- En dominios particulares, una palabra u oración puede tener un significado opuesto.
- Una oración interrogativa o una oración condicional puede no tener un sentimiento positivo o negativo, pero una palabra en particular puede tenerlo.

- Las frases sarcásticas pueden tener el sentimiento opuesto.
- Algunas oraciones pueden tener información de sentimiento, pero no usan palabras de sentimiento.
- Una sola palabra puede cambiar la polarización del sentimiento en dos oraciones similares, así como el hecho de que, para una persona diferente, una oración puede tener un sentimiento diferente.
- Los problemas del lenguaje natural cambian de lugar a lugar.

Existe un modelo llamado BERT mayormente utilizado para tareas como la inferencia del lenguaje y la respuesta de preguntas y ha demostrado tener muy buenos resultados gracias a la forma como procesa el texto. Por ello es muy popular y reconocido como técnica revolucionaria que da inicio a una nueva era en el procesamiento de lenguaje natural (NLP). Este es propuesto por Google como un modelo pre entrenado con un conjunto de información gigantesca para la comprensión del lenguaje natural.

El acrónimo BERT significa “Bidirectional Encoder Representations from Transformers” por ende la forma en la que analiza las oraciones es bidireccional para una mejor comprensión de texto, hace uso de una arquitectura basada en “Transformers” que consta de una codificación y una decodificación [19].

En Pandey, S. hacen el análisis de sentimientos sobre la red social de Twitter para poder hacer un sondeo de cuál fue el impacto en la vida social por el efecto de coronavirus, para el proyecto se utilizó BERT para el análisis que se realiza en la sección 4 de dicho documento, este análisis también toma en consideración los likes que son recibidos junto con los retweets, se puede presumir que entre una mayor de cantidad de likes los tweets tienen una mayor acogida [14].

Potena, D. tiene un trabajo parecido al de Pandey, S. con la diferencia de que a pesar de trabajar también con tweets estos aún conservan emoticones, los cuales son utilizados para poder tener una mayor polarización en el texto y ser mejor clasificado [14][15].

El trabajo de Pota es muy parecido al creado por Potena, D. pero Pota logró llegar un poco más allá logrando poder trabajar también no sólo el lenguaje del inglés sino que logró también el trabajar con italiano [15][16].

Así podemos darnos cuenta que BERT es una poderosa herramienta que no solo nos permite el entendimiento de texto en inglés, sino que también puede ser utilizado para varios lenguajes.

Materiales y métodos o Metodología computacional

Para la propuesta de este trabajo, se busca desarrollar una herramienta que permita clasificar una serie de tweets como depresivos o no depresivos. Todos los resultados obtenidos podrán ser visualizados a través de una interfaz web por medio de un gráfico Pie Chart.

Se hace uso del modelo BERT, tomándolo como base para el procesamiento de texto. Se añade una red neuronal lineal, la cual le permite que se especialice en una tarea en particular que en nuestro caso, es el análisis de tweets para reconocer cuando es depresivo o no lo es. Luego este modelo junto con la red neuronal deben pasar a un entrenamiento para alcanzar el objetivo de clasificación específica.

Para lograr el objetivo, primero es necesario conseguir los recursos para este entrenamiento. Se utiliza tablas procedentes del sitio web de Kaggle [21], [22]. Este conjunto de datos está estructurado de la siguiente manera: Tweet y su respectiva clasificación (depresivo o no depresivo).

Los tweets incluyen emoticones, lo cuales dificultan la interpretación del texto, elevando la ambigüedad, por ello se suprimieron antes de pasar al tratamiento del texto de los tweets procedentes de las tablas. Este tratamiento consiste en dividir el texto en tokens (tokenización) que luego son codificados por BERT Tokenizer. Estas etapas previas son necesarias para que el modelo BERT pueda entender el texto.

Una vez terminado el entrenamiento, la herramienta estará lista para clasificar tweets. El uso de esta herramienta web será por medio de una interfaz gráfica en el que el usuario podrá buscar tweets por un tema específico para luego ser clasificados. Los resultados serán entregados finalmente en el gráfico Pie Chart el cual indica los porcentajes de cada clase de tweets (depresivo y no depresivo).

A. Pipeline

El diseño de la propuesta puede verse reflejado en el siguiente pipeline:

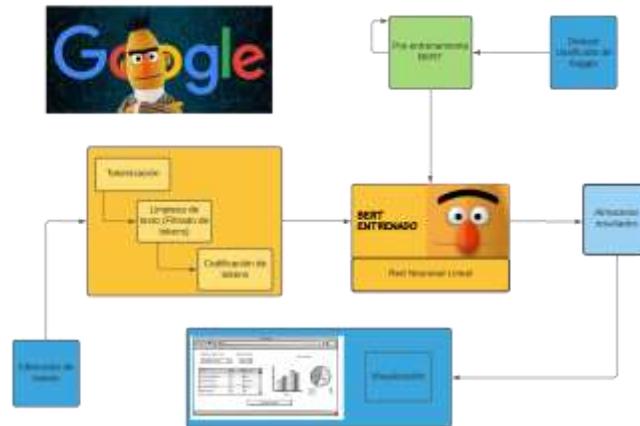


Figura 1. Pipeline descriptivo de la propuesta.

La entrada principal es la data de Tweets que se extrajeron de la API Twitter, la cual será analizada. Luego pasa por un proceso de tokenización tarea que es realizada el mismo modelo de BERT, segundo una limpieza de texto la cual consiste en eliminar los tokens que no aportan en el análisis como símbolos, menciones (no propias al lenguaje común), por último una codificación de tokens.

En el modelo se hará uso de una red neuronal lineal para el proceso de pre entrenamiento y entrenamiento. Como entrada para el modelo se tiene un dataset ya clasificado por dos clases: depresivo no depresivo, la cual servirá como input para el entrenamiento.

Como salida del modelo de BERT se obtendrán los resultados de clasificación de tweets los cuales pasarán a ser procesados con el fin que tenga una correcta estructura que sirva para una futura visualización de los datos. Luego pasarán a ser almacenados y por último los resultados serán presentados mediante una interfaz gráfica para representar de manera visual así como estadísticas para tener un mejor entendimiento de la información.

B. Datos

La data a analizar será extraída del API Tweepy, y tendrá la siguiente estructura:

Tabla 1. Data a analizar

Atributos	Descripción
tweet_id	Identificador del tweet
tweet	Cuerpo del mensaje del tweet
classification	Valoración: depresivo, no depresivo
twitter_username	Nombre de usuario
twitter_user_location	Ubicación del usuario
hashtags	Son los hashtags que se encuentran en el Tweet

En cuanto al dataset para el entrenamiento del modelo base se tendrá un archivo CSV con 116933 tweets y se tendrá los siguientes parámetros: Texto, Sentimiento.

C. Principios de diseño

Con la intención de desarrollar un software de calidad, se planteó el diseño basado en dos clases de principios:

- Kiss (“Keep it simple, Stupid!”)

Se propone hacer el diseño lo más simple posible, de forma tal que la complejidad no sea tanta como para evitar que el proyecto no logre reflejar sus objetivos por perderse en detalles innecesarios; ni sea tan simple como para desmerecer la calidad del servicio planteado [17].

- Solid

Representa cinco principios básicos de la programación orientada a objetos:

S – Single Responsibility Principle (SRP)

- O – Open/Closed Principle (OCP)
- L – Liskov Substitution Principle (LSP)
- I – Interface Segregation Principle (ISP)
- D – Dependency Inversion Principle (DIP)

En cuanto a la etapa de desarrollo, se aplican los principios SOLID. Se busca mantener responsabilidades únicas sobre el código planteado, de modo tal que no sean necesarias de modificar, a menos que estrictamente deban de hacerlo debido a que la función real (no la abstraída) cambie su concepto.

Las clases programadas, estarán dispuestas a extenderse para lograr mejorar, y no ser modificadas con otro fin que no sea la mejora de las mismas.

Aunque el proyecto es corto, en caso sea necesario extenderlo a través de subclases, estas no deberán evitar la funcionalidad haciéndose exclusivas, sino que se deberá poder seguir operando bajo las clases base, primeramente, planteadas para dicho fin.

El diseño se está planteando con secciones que realizan una tarea específica, cada parte del programa se encarga de un contexto único, y las funcionalidades diferentes son parte de otra clase. Por ejemplo: tokenización, limpieza y análisis [18].

D. Propuesta UI

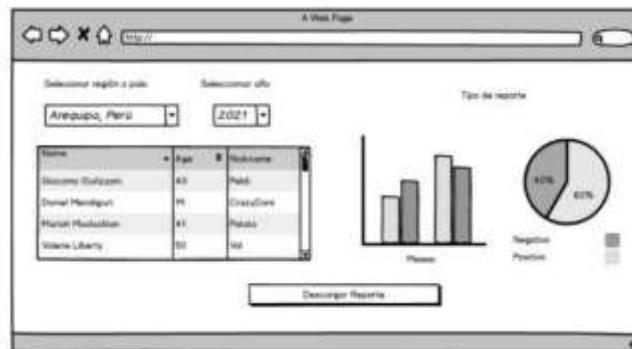


Figura 2. Prototipo de la interfaz de usuario.

Para la propuesta UI se desarrollará una aplicación web, en la cual se podrán visualizar los resultados del análisis de tweets por medio de gráficos estadísticos (Pie Chart). La extracción de tweets se hará de forma automática cuando el usuario busque un tema en concreto, además se mostrará una leyenda para reconocer las etiquetas de los tweets clasificados en depresivo o no depresivo.

E. Implementación

Para la propuesta se siguieron los pasos descritos en el algoritmo de la siguiente subsección, que estará implementado en Python. Se dió inicio a obtener el conjunto de datos de Tweets relacionados con un contexto especificado por un especialista interesado en obtener información y alertas sobre casos de depresión. Esto se logrará gracias al API Tweepy la que se encargará de proporcionar todos los tweets con su respectiva información.

La implementación se realizó en un entorno virtual de Google Colab con las siguientes características: GPU del backend de Google Compute Engine en python 3, RAM 12.69 GB, Disco 68.35 GB. Para el entrenamiento de BERT para el reconocimiento de sentimiento se tardó 2 horas 21 minutos y 23 segundos utilizando la GPU del tipo CUDA proporcionado por Google Colab.

Con respecto a la evaluación del tweet, en primer lugar, es necesario un preprocesamiento de los tweets para convertirlo en tokens entendibles para el modelo entrenado BERT, así mismo una limpieza y estandarización del texto. Luego el algoritmo procede a clasificar cada tweet analizado si es depresivo o no depresivo.

Para el entrenamiento del modelo base de BERT junto a la red neuronal lineal, se hará uso de un dataset con ejemplos de tweets, frases y citas ya clasificados. BERT se encargará de la comprensión de texto mientras que la red neuronal lineal se encargará de determinar si un tweet es o no depresivo.

Algoritmo 1: Obtención de Tweets

Input: TOPIC => Contexto, N_TWEETS_REQUESTED
=> número de tweets solicitados
Output: TWEETS => Tweets de la API tweepy
TWEETS = []
for i = 1 ... N_TWEETS_REQUESTED
=> número de tweets solicitados
return TWEETS

Algoritmo 2: Entrenamiento BERT

Input: DATA_SET => Grupo de datos
clasificados, MODEL => Modelo
pre entrenado BERT, DATA_TEST => Grupo de datos clasificados.
Output: PORCENTAJE_DE_EFICIENCIA => Porcentaje que indica que tan efectivo es el modelo
clasificando tweets en el entrenamiento.

DATA_LOADER = []

for i = 1 DATA_SET.tamaño
DATA_LOADER.añadir (tokenización (DATA_SET[1].texto), DATASET[1].clasificación)
MODEL.train (DATA_LOADER)
DATA_LOADER_TEST = []
for i = 1 DATA_TEST.tamaño
DATA_LOADER_TEST.añadir (tokenización (DATA_SET[1].texto),
DATA_TEST[1].clasificación)

PORCENTAJE_DE_EFICACIA = MODEL.evaluación

return PORCENTAJE_DE_EFICACIA

Algoritmo 3: Clasificación de tweets

Input: TWEETS => Tweets, Tweets de la API tweepy
Output: CLASIFICACIONES => Si es o no depresivo.

TWEETS_TOKENIZADOS = []

TOKENIZADO = []

LIMPIO = []

```

CODIFICADO = []

for i = 1 TWEETS.tamaño
    TOKENIZADO = tokenización (TWEETS [i][Text])
    LIMPIO = limpieza (tokenizado)
    CODIFICADO = codificación (LIMPIO)

    TWEETS_TOKENIZADOS.añadir (CODIFICADO)

CLASIFICACIONES = []
for i = 1 TWEETS.tamaño
    CLASIFICACIONES.añadir (
        MODEL.clasificar tweet(
            (TWEETS_TOKENIZADOS [ i ]))
        return CLASIFICACIONES
    
```

Resultados y discusión

A. Resultados de la implementación del modelo de BERT

Se obtuvieron los siguientes resultados de efectividad durante el entrenamiento de BERT. El entrenamiento duró 2 horas con 21 minutos y 23 segundos con una cantidad de 57732 tweets clasificados. Para ambas clases en el dataset (depresivo y no depresivo) se ha considerado la misma cantidad de tweets.

Tabla 2. Resultados del entrenamiento de BERT

Época	Entrenamiento		Validación	
	Loss	Accuracy	Loss	Accuracy
1	0.02922	0.99214	0.01582	0.99740
2	0.00869	0.99840	0.01578	0.99801
3	0.00544	0.99911	0.01257	0.99827
4	0.00268	0.99959	0.01311	0.99835
5	0.00104	0.99983	0.01594	0.99809

Como se puede observar en la tabla 2, se obtuvo luego de 5 épocas del entrenamiento de BERT un Accuracy de 0.99983 y un 0.99809 en la validación.

Estos son algunos ejemplos que se utilizaron para la prueba de clasificación de tweets.

Tabla 3. Ejemplos de tweets y sus resultados de clasificación.

Tweet	Clasificación
Up early to watch the GP today. Dodgy hamstring is still dodgy, so I wouldn't have been able to run this morning anyway	No depresivo.
is very sad because she can't be with her Mom during this sad time.	Depresivo
Suicide is the best way to get out of any situation...	Depresivo
@shaundiviney I want pizzaaaa.	No depresivo

B. Métricas de exactitud

Se evaluó el rendimiento de los algoritmos Machine Learning de clases múltiples mediante el uso de the 10-Fold Cross-Validation strategy [20].

Para ello se utilizó la evaluación F1 por medio de la librería de python llamada sklearn.

La medida de evaluación F1 corresponde a la media armónica de Precisión (Ecuación 1) y Recuperación (Ecuación 2), donde T P (Verdadero Positivo) se refiere al número de documentos de una clase en la que el algoritmo clasificó correctamente; FP (Falso Positivo) se refiere al número de documentos que no pertenecen a una clase en la que el algoritmo clasificó erróneamente como pertenecientes; y F N (Falso Negativo) se refiere al número de documentos de una clase en los que el algoritmo clasificó erróneamente como otra clase. La ecuación 3 define la medida F1.

$$P = \frac{TP}{TP+FP}, \quad R = \frac{TP}{TP+FN}, \quad F_1 = \frac{2TP}{2TP + FP + FN} \quad (1)$$

Aplicando la evaluación F1 con el dataset de tweets se obtuvo como resultado un valor de 0.8806.

C. Interfaz gráfica

Para una mejor visualización de datos y usabilidad para el usuario, se plantea el uso de herramientas gráficas fáciles de comprender.

Se desarrolló una herramienta web gracias a los frameworks Vue.js y Flask para poder obtener de manera automática los tweets de un tema específico usando la librería Tweepy.



Figura 3. Interfaz de búsqueda de tweets por tema.

Así mismo mediante la herramienta web construida es permisible obtener un gráfico del tipo Pie chart para la visualización estadística por clases.

Esta herramienta permite filtrar los resultados de la búsqueda antes realizada. Así mismo muestra una leyenda por colores para identificar las clases de tweets en el gráfico, siendo estas: depresivo en color celeste y no depresivo en color anaranjado.

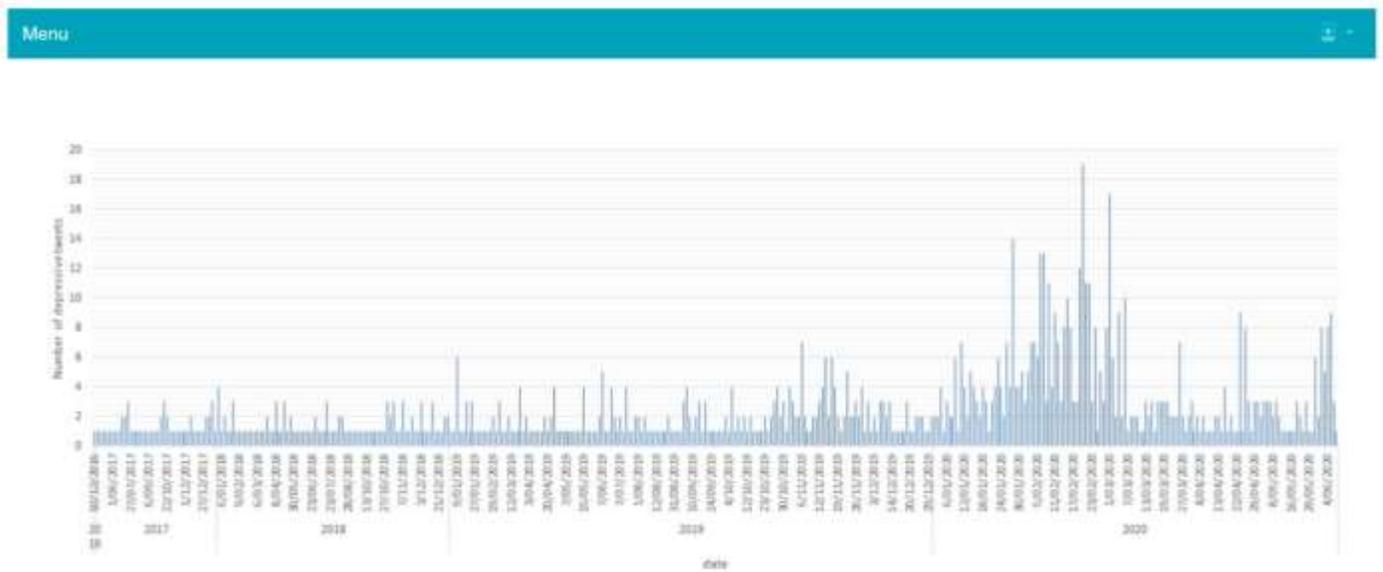


Figura 4. Histograma generado por la herramienta.

Conclusiones

A pesar de existir varios modelos para el análisis, es importante considerar la eficiencia de los resultados y la calidad que se busca ofrecer, así que para el proyecto se utilizó una técnica de clasificación de textos bastante eficiente como es en la actualidad BERT, logrando un valor en la métrica F1 de 0.8806 porcentaje superior a otras técnicas que existen en el estado del arte.

La aplicación del Procesamiento del Lenguaje Natural por medio del uso de distintas técnicas a lo largo de los años se ha ido refinando, siendo la inteligencia artificial la herramienta más valorada para NLP en la actualidad, la cual se refleja en el modelo BERT, que demuestra ser mucho más eficiente que técnicas usadas previamente con simple reconocimiento de texto sin considerar el contexto.

Se debe recordar que la evaluación de sentimientos de esta propuesta es solo para tweets en inglés.

En el presente trabajo, fue desarrollada una herramienta web llamada “UBDevs-Depression-Classifier” la cual permite a especialistas, organizaciones gubernamentales o privadas con implicancias o relaciones sobre la salud de la población,

identificar problemas de depresión en la red social Twitter. Así mismo, obtener de forma temprana alertas sobre problemas de depresión, muchas veces desencadenantes y/o agravantes de otros problemas de salud física.

Se realizó mayor énfasis en los tweets de COVID-19 debido a que los casos de depresión se incrementaron en la pandemia que atraviesa el mundo en los años 2020-2021.

Como trabajo futuro se espera mejorar aún más el porcentaje de la métrica F1 ampliando el dataset de entrenamiento y de prueba. Además, se espera mejorar esta herramienta web considerando una línea de tiempo que permita a los especialistas observar la variación de índices de depresión respecto a una fecha o periodo determinado.

Así mismo, se podría incluir una mayor cantidad de filtros para la vista de gráficos, como por ejemplo un filtro por localización. Gracias a ello será más útil en el estudio de los tweets y podrán orientar su uso a labores de prevención de la depresión sobre determinadas regiones, ciudades o países. Como trabajo futuro, se espera desarrollar o ampliar la funcionalidad de la herramienta para incluir la identificación de palabras más importantes y frecuentes en los tweets depresivos, si es posible también aumentar el idioma español para la herramienta.

Contribución de Autoría

Guillermo José Aleman-Zambrano: [Conceptualización](#), [Análisis formal](#), [Investigación](#), [Visualización](#), [Metodología](#), [Software](#), [Validación](#), [Redacción - borrador original](#), [Curación de datos](#), [Escritura, revisión y edición](#). **Marvik Irzovic Del Carpio-Lazo:** [Conceptualización](#), [Investigación](#), [Visualización](#), [Metodología](#), [Software](#), [Validación](#), [Redacción - borrador original](#), [Curación de datos](#). **Daniel Gustavo Mendiguri-Chávez:** [Conceptualización](#), [Análisis formal](#), [Investigación](#), [Visualización](#), [Metodología](#), [Software](#), [Validación](#), [Redacción - borrador original](#), [Curación de datos](#), [Escritura, revisión y edición](#). **Daniela Carolina Vilchez-Silva:** [Visualización](#), [Software](#), [Validación](#), [Redacción - borrador original](#), [Curación de datos](#). **Franco Tejada-Toledo:** [Conceptualización](#), [Investigación](#), [Metodología](#), [Análisis formal](#), [Supervisión](#).

Referencias

- [1] Chen, F., Zheng, D., Liu, J., Gong, Y., Guan, Z., & Lou, D. (2020). Depression and anxiety among adolescents during COVID-19: A cross-sectional study. *Brain, behavior, and immunity*, 88, 36.

- [2] Islam, M. A., Barna, S. D., Raihan, H., Khan, M. N. A., & Hossain, M. T. (2020). Depression and anxiety among university students during the COVID-19 pandemic in Bangladesh: A web-based cross-sectional survey. *PloS one*, 15(8), e0238162.
- [3] Lee, S. A., Jobe, M. C., Mathis, A. A., & Gibbons, J. A. (2020). Incremental validity of coronaphobia: Coronavirus anxiety explains depression, generalized anxiety, and death anxiety. *Journal of anxiety disorders*, 74, 102268.
- [4] Santini, Z. I., Jose, P. E., Cornwell, E. Y., Koyanagi, A., Nielsen, L., Hinrichsen, C., ... & Koushede, V. (2020). Social disconnectedness, perceived isolation, and symptoms of depression and anxiety among older Americans (NSHAP): a longitudinal mediation analysis. *The Lancet Public Health*, 5(1), e62-e70.
- [5] Bhuiyan, A. I., Sakib, N., Pakpour, A. H., Griffiths, M. D., & Mamun, M. A. (2020). COVID-19-related suicides in Bangladesh due to lockdown and economic factors: case study evidence from media reports. *International Journal of Mental Health and Addiction*, 1-6.
- [6] Reddy MS. Depression - the global crisis. *Indian J Psychol Med* 2012;34:201-3.
- [7] World Health Organization. (2014). Preventing suicide: A global imperative. World Health Organization.
- [8] Charoensukmongkol, P. (2018). The impact of social media on social comparison and envy in teenagers: The moderating role of the parent comparing children and in-group competition among friends. *Journal of Child and Family Studies*, 27(1), 69-79.
- [9] Anger, I., & Kittl, C. (2011, September). Measuring influence on Twitter. In *Proceedings of the 11th international conference on knowledge management and knowledge technologies* (pp. 1-4).
- [10] Sobrino Sande, J. C. (2018) Análisis de sentimientos en Twitter.
- [11] Kauffmann, E., Peral, J., Gil, D., Ferrández, A., Sellers, R., & Mora, H. (2020). A framework for big data analytics in commercial social networks: A case study on sentiment analysis and fake review detection for marketing decision-making. *Industrial Marketing Management*, 90, 523-537.
- [12] Back, B. H., & Ha, I. K. (2019). Comparison of sentiment analysis from large Twitter datasets by Naïve Bayes and natural language processing methods. *Journal of information and communication convergence engineering*, 17(4), 239-245.
- [13] Leis, A., Ronzano, F., Mayer, M. A., Furlong, L. I., & Sanz, F. (2019). Detecting signs of depression in tweets in Spanish: behavioral and linguistic analysis. *Journal of medical Internet research*, 21(6), e14199.
- [14] Singh, M., Jakhar, A. K., & Pandey, S. (2021). Sentiment analysis on the impact of coronavirus in social life using the BERT model. *Social Network Analysis and Mining*, 11(1), 1-11.

- [15] Chiorrini, A., Diamantini, C., Mircoli, A., & Potena, D. (2021). Emotion and sentiment analysis of tweets using BERT. In EDBT/ICDT Workshops.
- [16] Pota, M., Ventura, M., Fujita, H., & Esposito, M. (2021). Multilingual evaluation of pre-processing for BERT-based sentiment analysis of tweets. *Expert Systems with Applications*, 181, 115119.
- [17] Qadeer, S., & Wu, D. (2004). KISS: keep it simple and sequential. *ACM sigplan notices*, 39(6), 14-24.
- [18] Rotge, J. F. (2000, September). SGDL-Scheme: a high level algorithmic language for projective solid modeling programming. In *Proceedings of the Scheme and Functional Programming 2000 Workshop (Montreal, Canada (pp. 31-34))*.
- [19] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- [20] Tan, P., Steinbach, M., and Kumar, V. (2013). *Introduction to Data Mining: Pearson NewInternational Edition*. Pearson Education Limited.
- [21] Virahonda, S. (2021). Depressive and Anxious Tweets. Disponible en: <https://www.kaggle.com/datasets/sergiovirahonda/depression-anxiety-tweets>
- [22] Kazanova, M. (2018). Sentiment140 dataset with 1.6 million tweets. Disponible en: <https://www.kaggle.com/datasets/kazanova/sentiment140>
- [23] Hatzivassiloglou, V., & McKeown, K. R. (1997, July). Predicting the semantic orientation of adjectives. In *Proceedings of the 35th annual meeting of the association for computational linguistics and eighth conference of the european chapter of the association for computational linguistics (pp. 174-181)*. Association for Computational Linguistics