



ARTÍCULO RESULTADO DE INVESTIGACIÓN

Corporación Unificada
de Educación Superior

Análisis de la ideación suicida expresada en redes sociales mediante un modelo de aprendizaje profundo

*Analysis of suicidal ideation expressed
on social media using a Deep
learning model*

*Uso de Análise da ideação suicida expressa em
redes sociais por meio de um modelo de
aprendizagem profunda*

Recibido: Primer semestre 2024

Aprobado: Segundo semestre 2024

DOI: <https://doi.org/10.22490/27452115.8599>

Víctor Alfonso Guzmán Brand. Psicólogo-Especialista en analítica de datos. Corporación Unificada de Educación Superior. E-mail: victora.guzman@cun.edu.co. ORCID: <https://orcid.org/0000-0002-6051-3153>

Laura Esperanza Gélvez-García. Licenciada en lengua castellana, magíster en Lingüística española. Corporación Unificada de Educación Superior. E-mail: laura_gelvez@cun.edu.co. ORCID: <https://orcid.org/0000-0003-0164-2972>

RESUMEN

El artículo tiene como objetivo analizar la ideación suicida expresada en los comentarios realizados en un grupo de discusión en una red social, estudio llevado a cabo mediante un modelo pre-entrenado de aprendizaje profundo BERT (Bidirectional Encoder Representations from Transformers). Para abordar la investigación se emplea el método de descubrimiento de conocimiento en bases de datos (KDD), se caracteriza por la extracción de información de calidad, lo que permite generar conclusiones fundamentadas en relaciones o modelos identificados dentro de los datos. Como resultado cabe señalar que el análisis de datos de redes sociales indica una alta probabilidad de intención negativa con ideación suicida en la mayoría de los mensajes. La evaluación comparativa destaca la superioridad del modelo BERT en aplicaciones de deep learning, especialmente en el análisis de texto, gracias a su capacidad para entender patrones complejos en datos no estructurados. Se destaca además que la búsqueda de información sobre el suicidio en línea no solo revela la indagación sobre métodos suicidas, sino también la búsqueda de recursos de ayuda y apoyo para manejar el sufrimiento emocional. Algunas palabras clave en comentarios de redes sociales aumentan significativamente la probabilidad de publicar ideas suicidas en el futuro. El estudio concluye que la propuesta de una aplicación web resalta la necesidad de mejorar las herramientas para abordar los desafíos del suicidio en entornos digitales. El uso de inteligencia artificial y aprendizaje automático facilita la detección temprana y la intervención oportuna para promover la salud mental.

ABSTRACT

The objective of this article is to analyze the suicidal ideation expressed in the comments posted on a discussion group on social media, by means of a pre-trained deep learning model BERT (Bidirectional Encoder Representations from Transformers). To approach the research, the method of knowledge discovery in databases (KDD) is used, characterized by the extraction of quality information, which allows to generate conclusions based on relationships or models identified within the data. As a result, it should be noted that the analysis of social media data indicates a high probability of negative intent with suicidal ideation in most of the messages. Benchmarking highlights the superiority of the BERT model in deep learning applications, especially in text analysis, thanks to its ability to understand complex patterns in unstructured data. Additionally, it is noted that searching for information about suicide on-line not only reveals inquiry about suicidal methods, but also the search for resources for help and support to manage emotional distress. Some keywords in social media comments significantly increase the likelihood of posting suicidal ideation in the future. The study concludes that the proposal for a web application highlights the need for improved tools to address suicide challenges in digital environments. The use of artificial intelligence and machine learning facilitates early detection and timely intervention to promote mental health.

RESUMO

O artigo tem como objetivo analisar a ideação suicida expressa nos comentários feitos em um grupo de discussão em uma rede social, estudo realizado por meio de um modelo pré-treinado de aprendizagem profunda BERT (Bidirectional Encoder Representations from Transformers). Para abordar a pesquisa, emprega-se o método de descoberta de conhecimento em bases de dados (KDD), caracterizado pela extração de informações de qualidade, o que permite gerar conclusões fundamentadas em relações ou modelos identificados nos dados. Como resultado, destaca-se que a análise de dados de redes sociais indica uma alta probabilidade de intenção negativa com ideação suicida na maioria das mensagens. A avaliação comparativa ressalta a superioridade do modelo BERT em aplicações de aprendizagem profunda, especialmente na análise de texto, graças à sua capacidade de compreender padrões complexos em dados não estruturados. Além disso, destaca-se que a busca por informações sobre o suicídio online revela não apenas a investigação de métodos suicidas, mas também a procura por recursos de ajuda e apoio para lidar com o sofrimento emocional. Algumas palavras-chave em comentários de redes sociais aumentam significativamente a probabilidade de publicação futura de ideias suicidas. O estudo conclui que a proposta de um aplicativo web ressalta a necessidade de aprimorar as ferramentas para enfrentar os desafios do suicídio em ambientes digitais. O uso de inteligência artificial e de aprendizagem automática facilita a detecção precoce e a intervenção oportuna para promover a saúde mental.

PALABRAS CLAVES:

análisis, expresión, suicidio, redes sociales, aprendizaje profundo (Unesco Biblioteca Virtual de Salud).

KEYWORDS:

Analysis; Expression; Suicide; Social media; Deep learning (UNESCO Virtual Health Library).

PALAVRAS CHAVE:

análise, expressão, suicídio, redes sociais, aprendizagem profunda (Unesco Biblioteca Virtual da Saúde).

INTRODUCCIÓN

Cada año, alrededor de 703.000 personas fallecen por suicidio, con un impacto devastador en familias y comunidades. Este problema afecta a todas las edades y regiones del mundo, siendo la cuarta causa de muerte entre los jóvenes de 15 a 29 años en 2019. Además, más del 77% de los suicidios ocurren en países de ingresos bajos y medianos, destacando la necesidad de abordar este fenómeno a nivel global (Organización Mundial de la Salud, 2021). Desde el año 2016, se ha monitoreado los comportamientos suicidas en Colombia, reflejando una tendencia al alza tanto en intentos como en muertes consumadas. En el año 2018, se reportaron 28.615 casos de intentos de suicidio, lo que representa un promedio de 78.4 casos por día en todo el país (Ministerio de Salud y Protección Social, 2021).

El aumento de síntomas de ansiedad, depresión y comportamientos suicidas entre la población info-juvenil vulnerable se ha destacado como una preocupación creciente. Este fenómeno se atribuye en parte a la falta de supervisión en el uso indiscriminado de la tecnología digital, donde la accesibilidad sin filtros ni controles efectivos expone a los jóvenes a contenido perjudicial. La relevancia de la salud digital se vuelve crucial en este contexto, ya que puede tener un impacto significativo en la salud mental de los adolescentes (Paricio-Castillo et al., 2023).

Las conversaciones en línea destinadas a brindar ayuda se entremezclan con elementos de descalificación, agresión y motivación, convirtiendo las plataformas sociales en un espacio donde el tema del suicidio se expresa de manera diversa y a menudo perjudicial (Mendoza-Palma & Mera-Holguín, 2019). Esta cacofonía virtual puede ser especialmente dañina para aquellos que experimentan aislamiento social y carecen de una red de apoyo sólida que pueda contrarrestar la información negativa que puedan encontrar en línea (Molina & Restrepo, 2018).

De tal manera, que se resalta la importancia de analizar las redes sociales como una herramienta para la detección y prevención del suicidio. Estas plataformas ofrecen una oportunidad para identificar precozmente a individuos en riesgo, difundir información sobre recursos disponibles para personas en crisis y combatir el estigma asociado al tema del suicidio (Arilla-Andres et al., 2022). Sin embargo, también plantean desafíos, como la gestión de datos en tiempo real y la intervención efectiva. Dado que los adultos jóvenes, son un grupo particularmente vulnerable al suicidio, son los usuarios más activos en redes sociales, estas plataformas representan un medio clave para abordar este grave problema de salud pública.

Se ha observado la utilidad de los algoritmos de machine Learning para la detección de ideas o expresiones suicidas en los comentarios de las redes sociales. Este aspecto subraya la importancia de las técnicas avanzadas de procesamiento de datos, en la identificación temprana de contenido suicida. Es así como, al explorar secuencias y codificar representaciones vectoriales de términos, estos modelos pueden captar patrones lingüísticos y contextuales que pueden pasar desapercibidos para otros métodos de detección (Biscaia-Fernández, 2023; Gordon & Turnbull, 2023; Mulas-Camara et al., 2022; Zafra-Cremades, 2019).

De tal manera que, el objetivo del presente artículo es analizar la ideación suicida en los comentarios realizados en un grupo creado en una red social, estudio llevado a cabo mediante un modelo pre-entrenado de aprendizaje profundo desarrollado por Google y de acceso libre denominado BERT (Bidirectional Encoder Representations from Transformers).

METODOLOGÍA

Se emplea el método de descubrimiento de conocimiento en bases de datos (KDD), este hace referencia al proceso no trivial de descubrir conocimiento e información de utilidad potencial, que reside en los datos almacenados en un repositorio de información. Este proceso no es automático, sino más bien iterativo, implicando una exploración exhaustiva de volúmenes significativos de datos con el propósito de identificar relaciones relevantes (Joyanes, 2019). Se caracteriza por la extracción de información de calidad, lo que permite generar conclusiones fundamentadas en relaciones o modelos identificados dentro de los datos (García et al., 2018). Este enfoque para la minería de datos se organiza en varias etapas que facilitan un análisis riguroso y detallado. A continuación, se presenta la descripción de dichas etapas:

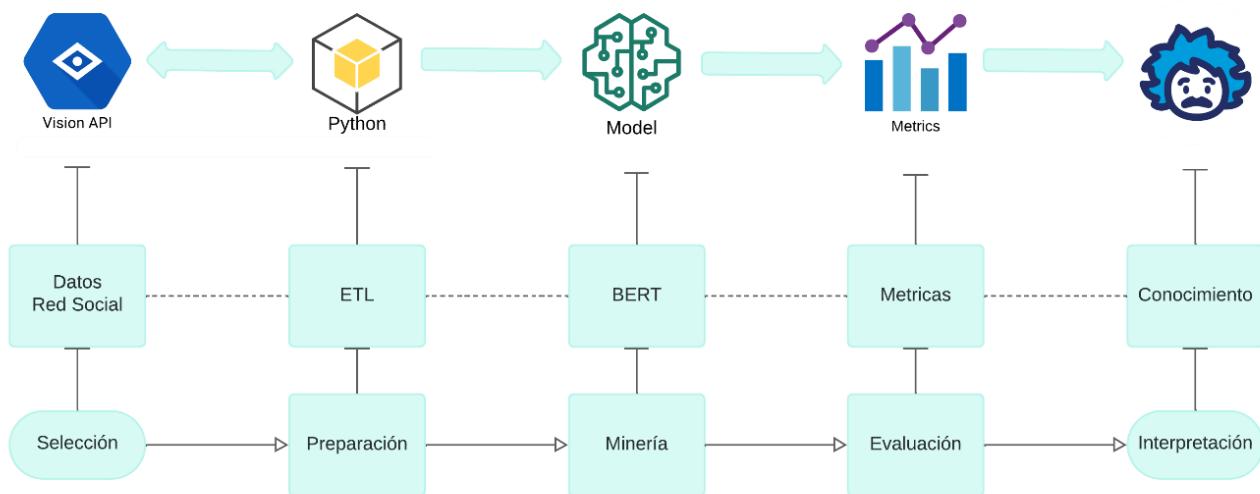


Figura 1. Ejecución del sistema de descubrimiento de conocimiento en bases de datos (KDD)

Nota: estructura adaptada al proceso KDD para minería de datos, teniendo en cuenta a (Calvache-Fernández et al., 2018); (Hernández et al., 2005). Creación propia

I. Selección de los datos

En cuanto a la selección de datos, se opta por registrar la actividad en una red social o foro denominado Reddit, donde los usuarios comparten opiniones y discuten sobre diversos temas. Plataforma que permite, tras una solicitud formal, extraer información de los comentarios realizados por los usuarios. En este caso, se conecta a la plataforma de desarrollo Google Colaboratory utilizando la interfaz de programación de aplicaciones (API) correspondiente. A través del lenguaje de programación Python, se realizan solicitudes para obtener los comentarios del grupo de discusión relacionado con el suicidio. Posteriormente, se lleva a cabo la transformación de los datos obtenidos para su análisis.

II. Preparación o preprocessamiento

Con respecto al preprocessamiento de datos en el análisis de sentimientos en el texto de comentarios de una red social, se requiere una serie de pasos para limpiar y normalizar el texto antes de que pueda ser utilizado en el análisis. Por un lado, es crucial eliminar caracteres especiales, signos de puntuación y emoticones del texto, ya que estos elementos no aportan información relevante para el análisis de sentimientos. Esto se puede lograr mediante el uso de expresiones regulares para identificar y eliminar estos caracteres no deseados. Además, se aplica la corrección ortográfica, la codificación de etiquetas y la normalización de palabras, para garantizar la calidad y consistencia de los datos utilizados en el análisis de sentimientos.

Consideraciones éticas

Este proceso investigativo se rige en razón a lo dispuesto en el decreto 1743 de 2016 sobre el reconocer el proceso de anonimización de microdatos el cual es entendido como artículo 2.2.3.1.1. “proceso técnico que consiste en transformar los datos individuales de las unidades de observación, de tal modo que no sea posible identificar sujetos o características individuales de la fuente de información, preservando así las propiedades estadísticas en los resultados” (Función Pública, 2023a).

RESULTADOS

III. Minería de datos

Después de depurar los datos, se procede a la etapa de experimentación, donde se implementa el modelo de inteligencia artificial BERT (Bidirectional Encoder Representations from Transformers), el cual está diseñado específicamente para preentrenar representaciones bidireccionales profundas utilizando texto no etiquetado, capturando simultáneamente el contexto tanto izquierdo como derecho en todas sus capas (Márquez-Díaz, 2023). Como consecuencia de este enfoque, este modelo puede ser afinado con una sola capa de salida adicional para generar modelos de vanguardia para diversas tareas, como la respuesta a preguntas y la inferencia del lenguaje, sin necesidad de modificaciones sustanciales en la arquitectura. La simplicidad conceptual y el poder empírico de BERT son destacables en el ámbito de la inteligencia artificial y el procesamiento del lenguaje natural (Devlin et al., 2019).

En términos generales, BERT Base utiliza los siguientes parámetros preentrenados: $L = 12$, $H = 768$, $A = 12$, con un total de 110 millones de parámetros. Aquí, L representa el número de capas (o bloques de Transformers), H indica el tamaño del espacio oculto, y A denota la cantidad de capas de atención. El tamaño del filtro de alimentación hacia adelante se establece en $4H$, equivalente a 3072 para $H = 768$. Modelo que procesa una secuencia de tokens con una longitud máxima de 512, generando una representación de la secuencia en un vector de 768 dimensiones, según la documentación base de BERT. Se prefiere utilizar un tamaño de lote más grande para obtener una estimación más estable del gradiente en el conjunto de datos completo (Carrión & Serrano, 2021).

Procesamiento del lenguaje natural (PLN)

El procesamiento del lenguaje natural (PLN) emplea técnicas de aprendizaje automático para descifrar la estructura y el significado presentes en los textos. Mediante las aplicaciones de PLN, las organizaciones tienen la capacidad de analizar el texto y extraer información pertinente sobre individuos, ubicaciones y eventos, lo que les permite obtener una comprensión más profunda de las opiniones expresadas en las redes sociales y las interacciones con los clientes (Google Cloud, 2023). Esta disciplina integra la lingüística, la informática y la inteligencia artificial con el objetivo de permitir a las computadoras comprender, procesar y producir lenguaje humano (Giraldo-Forero & Orozco-Duque, 2023).

Estas aplicaciones son herramientas fundamentales para la extracción de información valiosa de conjuntos de datos no estructurados basados en texto, así como para acceder y aprovechar la información extraída con el propósito de generar nuevos conocimientos a partir de dichos datos (Sancho-Escrivá et al., 2020). Además, es una combinación de la lingüística computacional, que implica el modelado basado en reglas del lenguaje humano, con modelos estadísticos, de aprendizaje automático y de aprendizaje profundo, posibilita que los sistemas informáticos procesen el lenguaje humano, ya sea en forma de datos de texto o de voz, y logren entender su significado integral, incluyendo la intención y el sentimiento del hablante o escritor (IBM, 2023).

Para poder alcanzar este proceso el modelo en su estructura contiene redes neuronales artificiales Transformers, estas se destacan por su capacidad para comprender el contexto y el significado de los datos secuenciales, como las palabras en una oración, mediante el análisis de las relaciones entre ellos. Estos modelos utilizan técnicas matemáticas en constante evolución, conocidas como atención o atención propia, para identificar conexiones sutiles entre los elementos de datos en una secuencia (Merritt, 2022).

Los transformers emplean codificadores posicionales para etiquetar los elementos de datos que ingresan y salen de la red, permitiendo que las unidades de atención creen un mapa algebraico que muestra cómo cada elemento se relaciona con los demás. Las consultas de atención se ejecutan típicamente en paralelo, calculando una matriz de ecuaciones en lo que se conoce como atención de múltiples encabezados (Beltrán & Rodríguez, 2021). Este enfoque permite a los modelos transformer capturar y comprender la complejidad de las relaciones secuenciales de manera más efectiva, lo que los hace especialmente adecuados para tareas que requieren una comprensión profunda del contexto y las interdependencias entre los datos (Martínez-Huertas, 2022).

Estructura del modelo

Después de la preparación de los datos, se procede a la fase de construcción del modelo. En este proceso, inicialmente se carga el modelo BERT (Bidirectional Encoder Representations from Transformers) se ajusta e incorpora una capa adicional diseñada específicamente para el análisis de sentimientos. Esta capa de dropout aplica la función Softmax para obtener las probabilidades predichas por el modelo entrenado, basadas en las salidas generadas. Este enfoque también se implementa con el fin de mitigar el riesgo de sobreajuste del modelo. Posteriormente, se realiza la conversión de las etiquetas a una codificación única, lo que permite que el modelo entrenado pueda clasificar comentarios etiquetados como negativos o positivos. De esta manera, se puede utilizar el modelo para predecir la etiqueta correspondiente de nuevos comentarios, clasificándolos en una de estas dos categorías. El esquema del modelo se presenta en la siguiente figura:

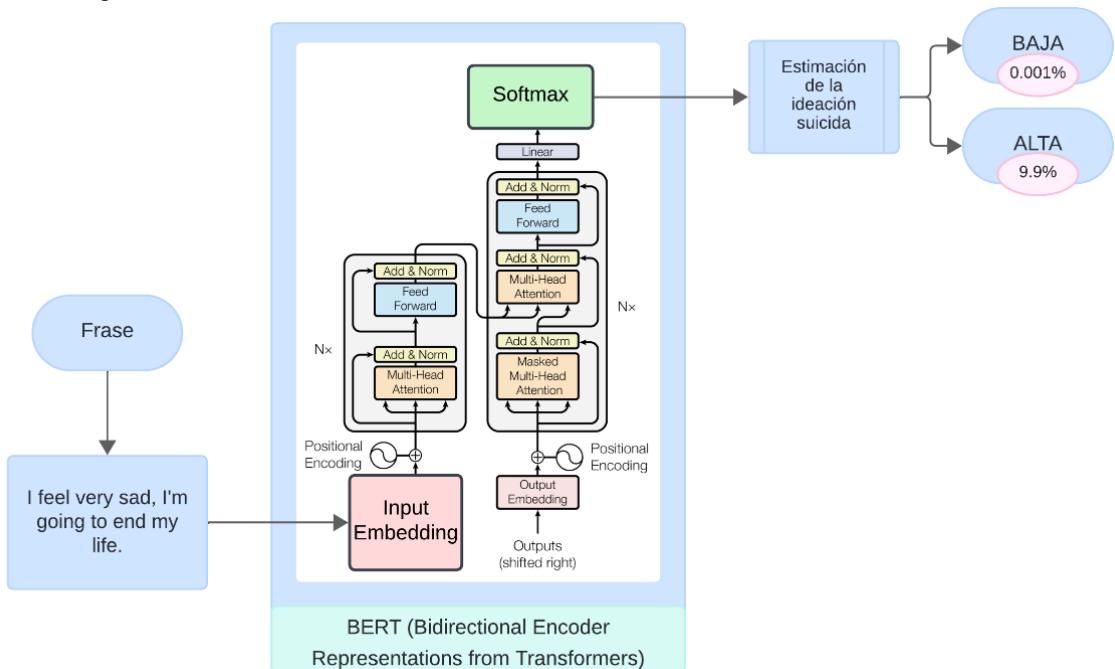


Figura 2. Estructura del modelo

Nota: organización del análisis de la información mediante la implementación del modelo BERT. Creación propia

Por otro lado, es importante destacar los parámetros utilizados durante el proceso de entrenamiento del modelo, los cuales fueron ajustados para lograr un rendimiento óptimo. Estos parámetros son fundamentales para influir en la capacidad del modelo para aprender y generalizar a partir de los datos proporcionados. Los parámetros empleados son los siguientes:

- Tamaño del lote: 16
- Tasa de aprendizaje: 2e-5
- Longitud máxima de secuencia: 200
- Épocas: 5
- Número de clases: 2
- Semilla aleatoria: 42

En el análisis de los datos, se recolectaron un total de 2075 mensajes o comentarios de la red social, los cuales presentaban una variedad de sentimientos. Estos mensajes se distribuyeron en un 20% para la validación y el 80% restante para el entrenamiento del modelo. De la totalidad de los mensajes, se observa que el 87.35% sugiere una alta probabilidad de que la persona esté expresando una intención negativa con ideación suicida, mientras que el 12.65% restante sugiere una alta probabilidad de que la persona esté expresando una intención positiva con baja ideación suicida. Estos resultados se presentan en la gráfica

siguiente:

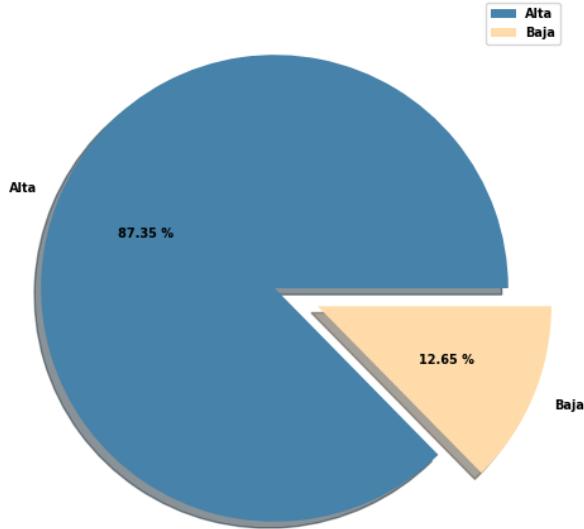


Figura 3 Análisis de los datos según su ideación suicida

Nota: desarrollo en Python Google Colab, información de acuerdo con los datos procesados por el modelo. Creación propia

Al mismo tiempo se genera un análisis con una nube de palabras sobre el texto el cual revela varias tendencias y patrones significativos en la conversación del grupo social. Las palabras más frecuentes, como (Suicidio) y (Morir), destacan como las más grandes y prominentes en la nube, sugiriendo que son temas principales de discusión entre los miembros del grupo. Además, la presencia destacada de (Ayuda) indica que los participantes están buscando apoyo y recursos para lidiar con estos temas difíciles, mientras que la palabra (Sentir) refleja la expresión abierta de emociones por parte de los miembros del grupo.

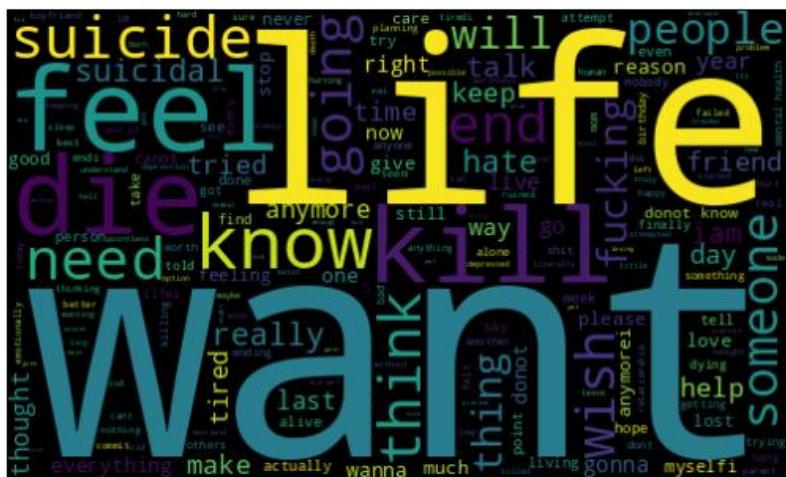


Figura 4 Nube de palabras sobre el texto

Nota: estructura jerárquica de las palabras con mayor frecuencia en el texto. Creación propia

Entre otras palabras notables, (Solo) y (Dolor) indican sentimientos de aislamiento y sufrimiento entre los miembros del grupo. Sin embargo, también se observan signos de esperanza, como lo evidencia la presencia de (Esperanza), sugiriendo que, a pesar de las dificultades, algunos miembros del grupo mantienen una visión optimista hacia el futuro. De esta forma, la nube de palabras refleja la complejidad de las emociones y pensamientos relacionados con el suicidio dentro del grupo social analizado. Mientras algunos expresan desesperanza y dolor, otros buscan ayuda y mantienen esperanzas para el futuro.

IV. Evaluación de rendimiento del modelo

El rendimiento del modelo BERT se analiza en comparación con otros algoritmos de aprendizaje automático que se entrenan con los mismos datos, estos tienen un buen rendimiento en la tarea de clasificación. Este análisis proporciona una evaluación integral de la eficacia y la capacidad predictiva de BERT en relación con otras técnicas de machine Learning.

Naive Bayes Kernel (NB)

Este algoritmo es conocido por su eficacia en la clasificación de datos, opera mediante la generación de una hipótesis para las clases de salida, las cuales se determinan en función de los atributos de entrada. Su enfoque probabilístico se fundamenta en el cálculo de la probabilidad de que una instancia pertenezca a una clase determinada, considerando los valores de los atributos proporcionados (Albán et al., 2022). Este método asume que los atributos son independientes entre sí, una simplificación que, si bien puede no ser válida en todos los casos, suele funcionar bien en la práctica (Ramírez-Gómez, 2020).

Es así como para la aplicación de este algoritmo se utiliza la librería *from sklearn.naive_bayes import MultinomialNB*, siendo especializada para la clasificación de características discretas (scikit-learn.org, 2024). Para llegar a un buen ajuste del modelo, se aplica un proceso de derivación y lematización en Python usando NLTK, siendo este último el que arroja mayor precisión. Esta técnica consiste en la agrupación de las múltiples formas flexionadas de una palabra para considerarlas como una única entidad durante el análisis. A diferencia de la derivación, este proceso proporciona contexto a las palabras, identificando y relacionando términos con significados similares para representarlos mediante una única forma (Pipis, 2022).

Regresión Logística

La Regresión Logística constituye un método estadístico fundamental para la predicción de clases binarias, donde la variable objetivo presenta una naturaleza dicotómica, es decir, se divide en dos categorías exclusivas (Alderete, 2006). Considerado uno de los algoritmos más simples y ampliamente empleados en el ámbito del Aprendizaje Automático, se destaca por su facilidad de implementación y su utilidad como punto de partida para problemas de clasificación binaria. Su función principal radica en describir y estimar la relación entre una variable binaria dependiente y las variables independientes pertinentes en el modelo (González, 2019).

Al emplear la Regresión Logística en conjunto con técnicas de NLP, es posible realizar tareas como la clasificación de documentos, el análisis de sentimientos, la detección de spam, entre otros. Este enfoque aprovecha la capacidad de la Regresión Logística para modelar relaciones entre variables binarias, mientras se procesan y analizan datos textuales de manera efectiva mediante técnicas de NLP para obtener insights significativos.

Tabla 1

Resultados de evaluación

Algoritmo	Accuracy	F1 Score	ROC AUC	Recall
Naive Bayes Kernel - MultinomialNB	0.96	0.96	0.96	0.96
Regresión Logística	0.95	0.95	0.96	0.95
BERT (Bidirectional Encoder Representations from Transformers)	0.99	0.99	0.99	0.99

Nota: creación propia

V. Implementación / Despliegue

....El despliegue del servicio basado en el modelo de aprendizaje profundo implica una serie de etapas cruciales

para garantizar su funcionamiento óptimo y su disponibilidad para los usuarios finales. Para ello, se utiliza el framework de desarrollo para ciencia de datos Streamlit, el cual permite construir una aplicación web robusta e interactiva.

Desarrollo rápido de aplicaciones (RAD)

El modelo RAD (Desarrollo Rápido de Aplicaciones) se destaca por su enfoque ágil y dinámico en la creación de prototipos, especialmente dirigido al ámbito del desarrollo web. Su principal característica radica en la retroalimentación constante que se establece con el cliente, lo que posibilita un desarrollo simultáneo y colaborativo (Delgado-Olivera & Díaz-Alonso, 2021). Esta interacción continua garantiza que el producto evolucione de manera adaptativa y se ajuste a las necesidades y expectativas del cliente en tiempo real. Además, gracias a esta flexibilidad y adaptabilidad, el modelo se mantiene en constante actualización, lo que lo hace idóneo para proyectos en los que los requisitos y las especificaciones pueden cambiar con frecuencia (Garcés-Gómez et al., 2021).

Planeación de sprint

Estos períodos de tiempo definidos, generalmente de una a cuatro semanas, representan unidades de trabajo enfocadas y manejables dentro de un proyecto más amplio. Durante cada sprint, el equipo se compromete a completar un conjunto específico de tareas o funcionalidades, lo que permite un avance incremental y constante hacia el objetivo final del proyecto. Cada sprint se enfoca en cumplir con un conjunto específico de objetivos, lo que permite un desarrollo iterativo y progresivo de la aplicación, asegurando así que se satisfagan los requerimientos de los usuarios de manera efectiva.

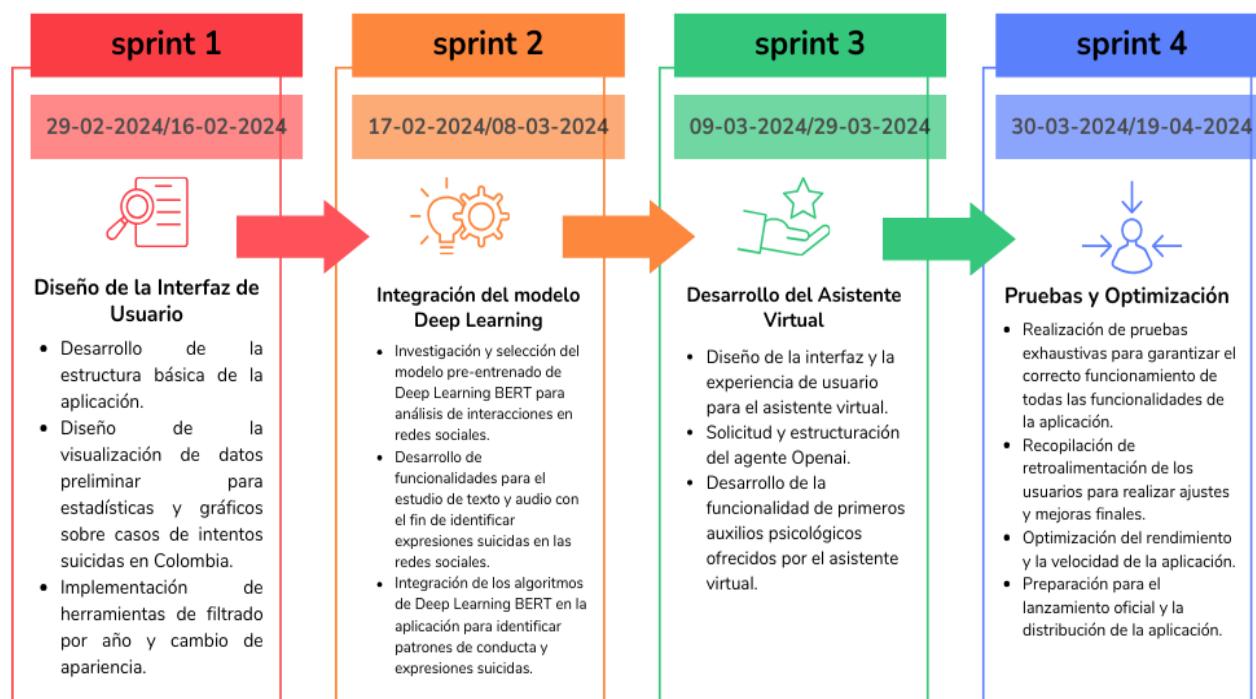


Figura 5 Sprint de desarrollo

Nota: Elaboración propia

Fase de identificación de necesidades o requerimientos

En esta etapa se recopilan o elaboran los requisitos del usuario, teniendo en cuenta que estos surgen en función de las necesidades identificadas en la literatura relevante sobre el tema. Además, se presenta esta propuesta como una herramienta para abordar los intentos de suicidio, en particular en la población con mayor incidencia, como se observó en el caso de la joven. Los requisitos se detallan en la tabla siguiente:

Tabla 2

Necesidades o requerimientos

Requerimientos	Descripción
Visualización de Datos	La aplicación debe presentar estadísticas y gráficos claros y comprensibles sobre los casos de intentos suicidas en Colombia.
	Debe permitir a los usuarios filtrar los datos por año y cambiar la apariencia.
	Se requiere una interfaz intuitiva que facilite la navegación y la comprensión de los datos para usuarios con diferentes niveles de conocimiento.
Interfaz de Usuario	La aplicación debe ser compatible con dispositivos móviles y otros dispositivos con diferentes tamaños de pantalla
	Se deben proporcionar herramientas de retroalimentación para que los usuarios puedan informar sobre posibles mejoras o problemas en la aplicación
Implementación de Algoritmos Deep Learning BERT	Se requiere la implementación del modelo pre-entrenado de Deep Learning BERT para analizar las interacciones en las redes sociales y encontrar patrones en la conducta, así como identificar expresiones suicidas
	Se requiere la funcionalidad de estudio de texto y audio para identificar expresiones suicidas en las interacciones en las redes sociales
Experiencia con un Asistente Virtual	La aplicación debe incluir un asistente virtual que brinde primeros auxilios psicológicos, ofreciendo apoyo y recursos a las personas en riesgo
	Es esencial que la aplicación garantice la privacidad y confidencialidad de los usuarios y los datos procesados

Nota: elaboración propia

Fase Diseño

Diagrama de casos de uso

Luego se definieron los casos de uso siendo una técnica utilizada en el análisis y diseño de sistemas de software para capturar los requisitos funcionales de un sistema desde la perspectiva del usuario final. En esencia, representan interacciones específicas entre el sistema y los actores (usuarios o sistemas externos) que lo utilizan para lograr un objetivo particular. Estas interacciones se describen en términos de acciones que el usuario realiza y las respuestas del sistema a esas acciones (Departamento Nacional de Planeación, 2020). De esta forma se muestran de la siguiente manera:

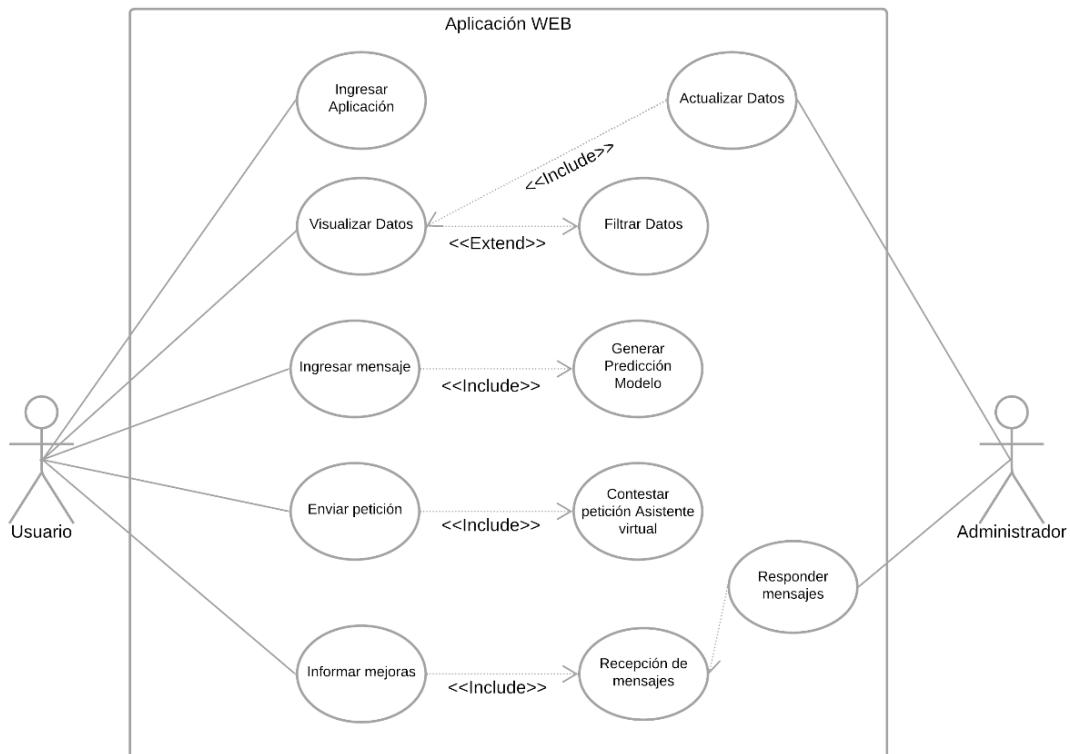


Figura 6 Diagramas de caso de uso

Nota: Elaboración propia

Diagrama de actividades

Por otro lado, se estructura un diagrama de actividades este muestra el flujo de trabajo de la aplicación web para exponer las características de los casos de intentos suicidas en Colombia durante el periodo de 2016 a 2022. Además, identificar las expresiones suicidas en las interacciones en las redes sociales mediante aprendizaje profundo. De esta manera, la interfaz permite el acceso al usuario de manera directa, encontrando una página de inicio, una segunda donde de información general del tema. Además, la aplicación permite a los usuarios filtrar datos por año en el panel de control, solicitar una predicción de las expresiones suicidas en redes sociales, e iniciar un asistente virtual que brinda primeros auxilios psicológicos.

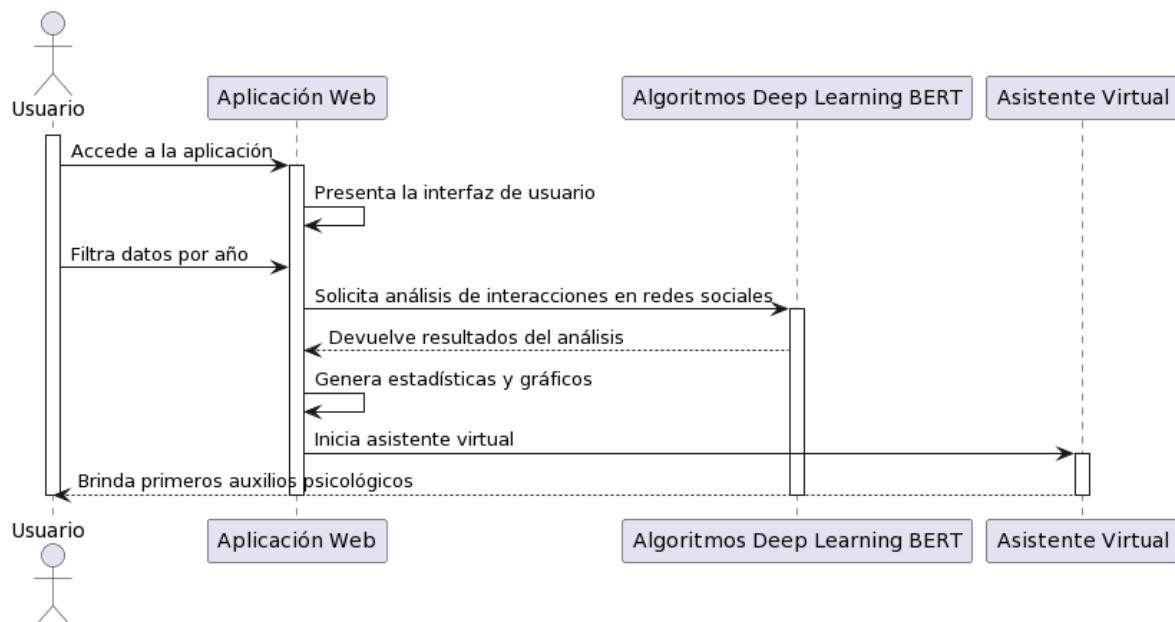


Figura 7 Diagramas de actividades
Nota: Elaboración propia

Diagrama de componentes

El diagrama de componentes el cual permite tener una visión de manera general la implementación la aplicación web simulando su forma física, a su vez muestra la relación entre cada componente. Esta estructura se expone de la siguiente forma:

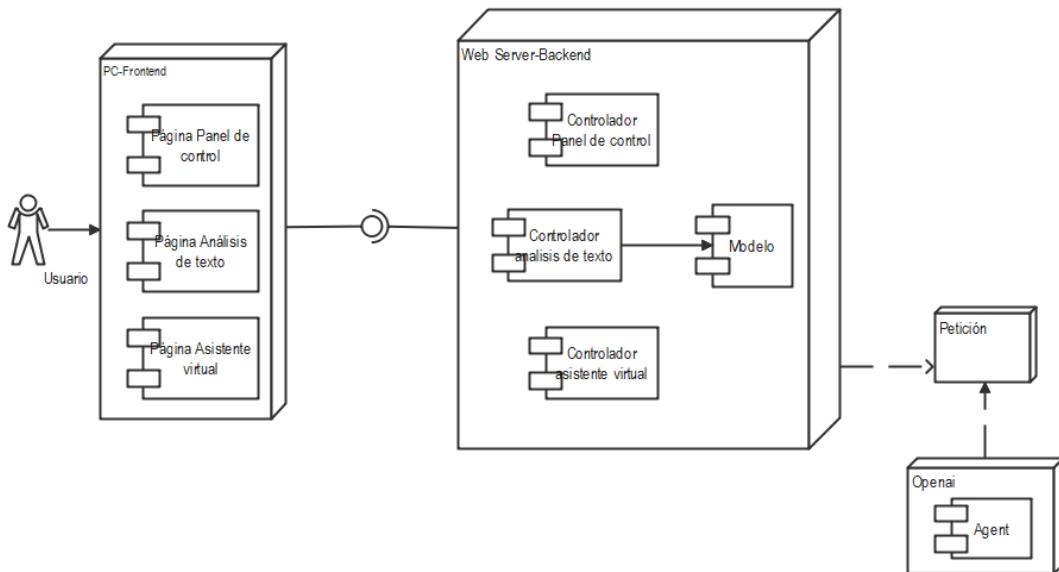


Figura 8 Diagrama de componentes

Nota: Elaboración propia

Desarrollo de la aplicación

Es crucial tener una comprensión clara de la idea y realizar una investigación exhaustiva sobre la temática que se pretende abordar. Al construir una aplicación web, es fundamental considerar ciertas características básicas, como la accesibilidad y la usabilidad. En este sentido, la accesibilidad se refiere a la capacidad del sitio para que sus contenidos sean percibidos por los usuarios, sin importar sus habilidades digitales o las herramientas que utilicen. Esto, a su vez, contribuye a mejorar la experiencia del usuario (Aragón-Payares et al., 2023). De esta forma, se desarrolla cada una de las partes y se muestran más adelante:



Figura 9 Estructura de la aplicación web

Nota: Elaboración propia

La aplicación web se organiza en varias páginas, cada una con un propósito específico. En la primera página, se presenta información general sobre la aplicación, seguida de alguna información de la situación problema, profundizando en estadísticas y conceptos teóricos relevantes. En la tercera página muestra un panel de control interactivo con estadísticas sobre los intentos de suicidio en Colombia durante el período 2016-2022.

En la cuarta página, se exhibe el modelo de Deep Learning en funcionamiento, donde el usuario puede ingresar texto o usar la función de reconocimiento de voz. Este texto es procesado y se muestra la probabilidad de ideación suicida. La quinta página contiene un asistente virtual para la atención de primeros auxilios psicológicos, mientras que en la sexta página se encuentra un formulario de contacto donde los usuarios pueden expresar sus opiniones y realizar recomendaciones sobre la aplicación.

Para disfrutar de la experiencia completa y acceder a la aplicación web que está alojada en la nube, te invitamos a utilizar el enlace facilitado. Además, para mayor comodidad, puedes escanear el código QR adjunto utilizando tu dispositivo móvil o cualquier otro dispositivo compatible. De esta manera, se podrá acceder de manera rápida y sencilla a la aplicación y explorar todas sus funcionalidades desde cualquier lugar. Accede aquí:



Figura 10 Código QR de la aplicación

Nota: Elaboración propia

Enlace de la aplicación web:

<https://deeppsychology.streamlit.app/>

Evaluación de la aplicación

Para evaluar la aplicación web, se utiliza un formulario creado en Google Forms. Este formulario incluye una encuesta con una variedad de preguntas diseñadas para explorar las experiencias de los usuarios al navegar por la plataforma, interactuar con su contenido y utilizar sus servicios. Además de brindar la oportunidad de identificar posibles fallos, este proceso permite realizar correcciones y mejoras de manera inmediata, además de promocionar la aplicación web.

Se tuvieron en cuenta aspectos como navegación, contenido, experiencia de uso, identificación de fallas y sugerencias de mejora percibidas por (30) los usuarios. Instrumento que arrojó que la navegación es intuitiva y fácil de entender en un 81%, el contenido es adecuado y bien organizado en un 90%, la experiencia de satisfacción al utilizar la aplicación es de un 91%, los fallos percibidos son de un 5%, sobre las mejoras que se proponen son crear un informe de resultados para ser descargado y avisos.

DISCUSIONES

Las personas que buscan información sobre el suicidio en Internet no solo encuentran sitios que ofrecen detalles sobre métodos y formas para llevar a cabo este acto, sino que también buscan recursos de ayuda, apoyo y orientación para lidiar con el sufrimiento asociado a sus pensamientos suicidas. Esto incluye enfrentar sentimientos de tristeza, soledad y ansiedad derivados de estas ideas (Molina & Restrepo, 2018). Es importante reconocer que la búsqueda de información sobre el suicidio en línea puede ser tanto una señal de angustia como un intento de encontrar soluciones o consuelo.

En cuanto a las palabras con mayor frecuencia y peso en los comentarios se destacan (Suicidio), (Morir), (Ayuda), (Sentir), (Solo), (Dolor) y (Esperanza), siendo las más grandes y prominentes en la nube, sugiriendo que son temas principales de discusión entre los miembros del grupo. En razón a esto, se han detectado palabras que aumentan notablemente la probabilidad de que un usuario de redes sociales publique ideas suicidas en el futuro, estas son "depresión", este incremento es del 30%, para "suicidio" es del 32%, para "sin amigos" es del 51%, para "llorar" es del 51%, y para "mata" es del 53% (De Choudhury et al., 2016).

Durante el análisis de los datos, se recopilaron un total de 2075 mensajes de una red social con diversas emociones. Del total, el 87.35% sugiere una alta probabilidad de intención negativa con ideación suicida, mientras que el 12.65% restante sugiere una probabilidad de intención baja de ideación suicida. Los resultados se alinean con los hallazgos de (Arilla-Andrés et al., 2022), quien indica que la publicación en redes sociales de ideas asociadas al suicidio antes de un intento autolítico es significativa, abarcando tanto publicaciones directas como indirectas, pensamientos negativos e imágenes desagradables.

La evaluación comparativa del rendimiento del modelo BERT en relación con otros algoritmos revela consistentemente su superioridad en diversas aplicaciones. Esta superioridad se evidencia en estudios previos que han confrontado algoritmos de deep learning con métodos tradicionales, destacando el rendimiento superior de los primeros, particularmente en tareas relacionadas con el análisis de texto. Esta tendencia hacia un mejor desempeño de los algoritmos de deep learning, como BERT, en el procesamiento de texto se atribuye a su capacidad inherente para capturar y comprender patrones complejos en datos no estructurados, como el lenguaje natural (Arango-Pastrana & Osorio-Andrade, 2021; López-Condori & Gonzales-Saji, 2021; Lovera & Cardinale, 2023).

Entre las limitaciones del estudio, cabe destacar la falta de acceso a datos adicionales, como el contexto completo de las publicaciones en redes sociales o la información demográfica de los usuarios, lo que podría haber Enriquecido el análisis y la interpretación de los resultados. Además, la investigación se centró en una única plataforma de redes sociales, lo que limita la generalización de los hallazgos a otros entornos digitales. Asimismo, la naturaleza automatizada del análisis de los datos podría haber pasado por alto matices importantes en el contenido de las publicaciones que solo podrían ser captados mediante un análisis cualitativo más detallado.

CONCLUSIONES

El presente estudio arroja aspectos cruciales para la detección y prevención de las expresiones suicidas en entornos digitales. Se identifican elementos clave que permiten desarrollar estrategias para abordar esta problemática. Destacando la importancia de reconocer la búsqueda de información sobre suicidio en línea como un indicador de angustia y búsqueda de ayuda.

El estudio identifica palabras clave en comentarios de redes sociales que se asocian con una mayor probabilidad de expresiones suicidas. Esta información es valiosa para desarrollar estrategias de intervención temprana y prevención. Además, subraya la eficacia de los modelos avanzados de aprendizaje automático, como el Deep Learning, en la detección de contenido suicida. Estos modelos son capaces de captar matices lingüísticos y contextuales que pueden pasar desapercibidos para otros métodos de detección, lo que los convierte en herramientas para identificar y abordar riesgos en entornos digitales.

La comparación del rendimiento del modelo BERT con otros algoritmos muestra consistentemente su superioridad en diversas aplicaciones, especialmente en tareas relacionadas con el análisis de texto. Esta tendencia hacia un mejor desempeño de los algoritmos de Deep Learning destaca la importancia de adoptar

enfoques avanzados en la detección.

Al proponer una aplicación web como recurso, se destaca la importancia de seguir desarrollando y perfeccionando herramientas y estrategias para abordar los desafíos asociados con la conducta suicida en entornos digitales. La utilización del aprendizaje automático y la inteligencia artificial permite avanzar hacia una detección temprana y una intervención oportuna para fomentar la salud mental.

Para futuras investigaciones, se sugiere explorar la efectividad de intervenciones específicas basadas en las palabras clave identificadas en este estudio para brindar apoyo y recursos a aquellos que expresan ideaciones suicidas en entornos digitales. Además, sería favorable investigar cómo mejorar y optimizar aún más los modelos de aprendizaje profundo para la detección temprana y precisa de contenido suicida, así como evaluar su implementación en tiempo real en plataformas de redes sociales para una intervención más rápida y efectiva.

Los autores refieren las siguientes anotaciones

Fuentes de financiamiento.

La investigación no recibió ningún financiamiento específico de organismos públicos o de sectores comerciales o sin fines de lucro.

Conflictos de interés

Todos los autores declaramos que no tenemos ningún conflicto de interés financiero, profesional o personal, que pueda influir de forma inapropiada en los resultados obtenidos o las interpretaciones propuestas.

Contribuciones de los autores

Autor 1: conceptualización, curación de datos, análisis formal, investigación, metodología, recursos, software, validación, visualización, escritura -preparación del borrador original, escritura, revisión y edición.

Autor 2: análisis formal, investigación, metodología, administración del proyecto, recursos, supervisión, escritura -preparación del borrador original, escritura, revisión y edición

REFERENCIAS



- Albán, D., Urvina, M., & Andrade, R. (2022). Análisis y Diseño de un Modelo Predictivo para Detección de Phishing Basado en Url y Corpus del Correo Electrónico. *Revista Politécnica*, 50(3), 1-15. <https://doi.org/10.33333/rp.vol50n3.03>
- Alderete, A. (2006). Fundamentos del Análisis de Regresión Logística en la Investigación Psicológica. *Revista Evaluar*, 6(1), 20-35. <https://doi.org/10.35670/1667-4545.v6.n1.534>
- Aragón-Payares, C., Marín-Betancourt, K., & Guzmán-Brand, V. (2023). Las Tecnologías de la Información y de la Comunicación (TIC) para la construcción de la interculturalidad en el aula | Revista RETOS XXI. *Revista RETOS XXI*, 6(6), 1-20. <https://revistaseug.ugr.es/index.php/RETOSXXI/article/view/25355>
- Arango-Pastrana, C., & Osorio-Andrade, C. (2021). Aislamiento social obligatorio: Un análisis de sentimientos mediante machine learning. *Suma de Negocios*, 12(26), 1-13. <https://doi.org/10.14349/sumneg/2021.v12.n26.a1>
- Arilla-Andrés, S., García-Martínez, C., & López-Del Hoyo, Y. (2022). Detección del riesgo de suicidio a través de las redes sociales. *Revista Internacional de Tecnología Ciencia y Sociedad*, 2(2), 1-25. <https://doi.org/10.37467/revtechno.v11.4384>
- Arilla-Andrés, S., García-Martínez, C., & López-Del Hoyo, Y. (2022). Detección del riesgo de suicidio a través de las redes sociales. *Revista Internacional de Tecnología Ciencia y Sociedad*, 5(6), 2-14. <https://doi.org/10.37467/revtechno.v11.4384>
- Beltrán, N., & Rodríguez, E. (2021). Procesamiento del lenguaje natural (PLN) - GPT-3.: Aplicación en la Ingeniería de Software. *Tecnología Investigación y Academia*, 8(1), 18-37. <https://revistas.udistrital.edu.co/index.php/tia/article/view/17323>
- Biscaia-Fernández, J. (2023). La inteligencia artificial en la prevención de conductas suicidas: Aspectos técnicos y consideraciones ético-legales | Revista de Bioética y Derecho. *Revista Bio y Derecho*, 5(5), 181-203. <https://revistes.ub.edu/index.php/RBD/article/view/42759>
- Carrión, J., & Serrano, V. (2021). Revisión sistemática de literatura: Características y funcionamiento respecto a los modelos BERT y SQuAD. *CEDAMAZ*, 11(1), Article 1. <https://revistas.unl.edu.ec/index.php/cedamaz/article/view/1041>
- De Choudhury, M., Kiciman, E., Dredze, M., Coppersmith, G., & Kumar, M. (2016). Discovering Shifts to Suicidal Ideation from Mental Health Content in Social Media.. *Proceedings of the SIGCHI conference on Human factors in computing systems*. New York, NY, USA: ACM.

human factors in computing systems . CHI Conference, 16(8), 2098-2110. <https://doi.org/10.1145/2858036.2858207>

- Departamento Nacional de Planeación. (2020). *Guía para la elaboración y presentación de casos de uso* (pp. 1-21). DNP. <https://colaboracion.dnp.gov.co/CDTI/Oficina%20Informatica/Sistemas%20de%20informaci%C3%B3n/Gu%C3%A1das%20Formatos%20Plantillas/Gu%C3%A1da%20para%20la%20Elaboraci%C3%B3n%20y%20Presentaci%C3%B3n%20de%20Casos%20de%20Uso.pdf?>
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Arxiv*, 1(1), 4171-4186. <http://arxiv.org/abs/1810.04805>
- García, J., Molina, J., Berlanga, A., Patricio, M., Bustamante, A., & Padilla, W. (2018). *Ciencia de datos. Técnicas analíticas y aprendizaje estadístico*. Alfaomega Grupo Editor.
- Giraldo-Forero, A., & Orozco-Duque, A. (2023). Evolución del procesamiento natural del lenguaje. *TecnoLógicas*, 26(56), 1-20. <https://doi.org/10.22430/22565337.2687>
- González, L. (2019). *Machine Learning con Python Aprendizaje Supervisado*. Editorial Independiente.
- Google Cloud. (2023). ¿Qué es el procesamiento del lenguaje natural? Google Cloud. <https://cloud.google.com/learn/what-is-natural-language-processing?hl=es>
- Gordon, S., & Turnbull, B. (2023). Vista de Adopción de la inteligencia artificial en el campo de la psicología | Psicología Iberoamericana. *Psicología Iberoamericana*, 1(1), 1-5. <https://psicologiaiberoamericana.ibero.mx/index.php/psicologia/article/view/547/1075>
- Hernández, J., Ramírez, M., & Ferri, C. (2005). *Introducción a la minería de datos*. Pearson Educación.
- IBM. (2023). ¿Qué es el procesamiento del lenguaje natural? <https://www.ibm.com/es-es/topics/natural-language-processing>
- Joyanes, L. (2019). *Inteligencia de negocios y analítica de datos. Una visión global de Business Intelligence & Analytics*. Alfaomega Grupo Editor.
- López-Condori, J., & Gonzales-Saji, F. (2021). Análisis de sentimiento de comentarios en español en Google Play Store usando BERT. *Ingeniare. Revista chilena de ingeniería*, 29(3), 557-563. <https://doi.org/10.4067/S0718-33052021000300557>
- Lovera, F., & Cardinale, Y. (2023). Análisis de sentimientos en Twitter: Un estudio comparativo. *Revista Científica de Sistemas e Informática*, 3(1), 1-20. <https://doi.org/10.51252/rksi.v3i1.418>

- Márquez-Díaz, J. (2023). *Modelos de lenguaje natural en la investigación científica: Una descripción técnica.* https://innovacionyciencia.com/articulos_cientificos/modelos-de-lenguaje-natural-en-la-investigacion-cientifica-una-descripcion-tecnica
- Martínez-Huertas, J. (2022). Introducción a las Redes Neuronales Transformers (Material docente). *Universidad Nacional de Educación a Distancia*, 1(1), 1-22. <http://e-socio.uned.es/fez/view/bibliuned:DptoMCC-FPSI-MatDocente-Jamartinez-0003>
- Mendoza-Palma, K., & Mera-Holguín, G. (2019). Adicción a las redes sociales y conducta suicida en los adolescentes de Montecristi. *Revista Científica y Arbitrada de Psicología NUNA YACHAY - ISSN: 2697-3588.*, 2(3), Article 3.
- Merritt, R. (2022, abril 19). *¿Qué Es un Modelo Transformer?* Blog oficial de NVIDIA Latino América. <https://la.blogs.nvidia.com/2022/04/19/que-es-un-modelo-transformer/>
- Ministerio de Salud y Protección Social. (2021). *Prevención de la Conducta Suicida en Colombia* (pp. 1-57). Ministerio de Salud y Protección Social. <https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/VS/P/ENT/estrategia-nacional-conducta-suicida-2021.pdf>
- Molina, M., & Restrepo, D. (2018). Internet y comportamiento suicida en adolescentes: ¿cuál es la conexión? *Pediatria*, 51(2), 120-135. <https://doi.org/10.14295/pediatr.v51i2.109>
- Mulas-Cámara, P., Fernández-Calvillo, R., Martínez-Cabezali, C., & Molina-Cañizares, M. (2022). Análisis de los datos obtenidos de la red social Twitter para la identificación precoz de la tendencia al suicidio de los usuarios. *Comunitania. Revista Internacional de Trabajo Social y Ciencias Sociales*, 23(24), Article 24. <https://doi.org/10.5944/comunitania.24.2>
- Organizacion Mundial de la Salud. (2021). *Suicidio.* <https://www.who.int/es/news-room/fact-sheets/detail/suicide>
- Paricio-Castillo, R., García-Murillo, L., Mallol-Castaño, L., Del Sol-Calderón, P., Pascual-Aranda, A., & Palanca-Maresca, I. (2023). Redes sociales y conductas suicidas en la infancia y la adolescencia durante la pandemia de COVID-19: Una relación difícil de estimar. *Revista de Psiquiatría Infanto-Juvenil*, 40(3), Article 3. <https://doi.org/10.31766/revpsij.v40n3a2>
- Pipis, G. (2022, noviembre 29). *Stemming and Lemmatization in Python using NLTK* [Medium]. Geek Culture.

<https://medium.com/geekculture/stemming-and-lemmatization-in-python-using-nltk-df8fd800a13f>

Ramírez-Gómez, C. (2020). Aplicación del Machine Learning en agricultura de precisión. *Revista CINTEX*, 25(2), 1-26.
<https://doi.org/10.33131/24222208.356>

Sancho-Escrivá, J., Fanjul-Peyró, C., Iglesia-Vayá, M., Montell, J., & Escartí-Fabra, M. (2020). Aplicación de la Inteligencia Artificial con Procesamiento del Lenguaje Natural para textos de investigación cualitativa en la relación médico-paciente con enfermedad mental mediante el uso de tecnologías móviles. *Revista de Comunicación y Salud*, 10(1), 1-15. [https://doi.org/10.35669/rcys.2020.10\(1\).19-41](https://doi.org/10.35669/rcys.2020.10(1).19-41)

scikit-learn.org. (2024). *Sklearn.naive_bayes.MultinomialNB*. scikit-learn.org.
https://scikit-learn/stable/modules/generated/sklearn.naive_bayes.MultinomialNB.html

Zafra-Cremades, S. (2019). Redes sociales para la prevención del suicidio juvenil. *Cuadernos de desarrollo aplicados a las TIC*, 8(2), 54-69.
<http://dx.doi.org/10.17993/3ctic.2019.82.54-69>