



Tipo de artículo: Artículos originales
Temática: Inteligencia artificial
Recibido: 02/05/2023 | Aceptado: 02/08/2023 | Publicado: 30/09/2023

Identificadores persistentes:
DOI: [10.48168/innosoft.s12.a97](https://doi.org/10.48168/innosoft.s12.a97)
ARK: [ark:/42411/s12/a97](https://nbn-resolving.org/urn:ark:/42411/s12/a97)
PURL: [42411/s12/a97](https://purl.org/42411/s12/a97)

Detección de Ideas Suicidas en Twitter

Suicidal Idea Detection on Twitter

Juliana Berrios Butron ^{1*}, Daria López Franco ², Dante Rodolfo Tárraga ³

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

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

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

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

Resumen

El análisis de sentimientos es una nueva tendencia' hoy en día para entender cómo se sienten' las personas en diferentes situaciones de su vida cotidiana. Los datos de las redes sociales se utilizan durante todo el proceso de análisis y' clasificación, que consiste en datos de texto. Utilizando las redes sociales, se puede monitorizar o analizar el nivel emocional. En este trabajo de investigación clasificaremos los datos de redes sociales como twitter relativos al suicidio y los clasificaremos como: pensamiento suicida activo, pensamiento suicida pasivo, sarcasmo relacionado con el pensamiento suicida, tweets relacionados con el suicidio (concienciación sobre el suicidio, noticias, charla sobre el suicidio) y otros.

Palabras clave: Suicidio, pensamientos suicidas, procesamiento de lenguaje natural.

Abstract

Sentiment analysis is a new trend' nowadays to understand how people' feel in different situations of their daily life. Social network data is used during the whole process of analysis and' classification, which consists of text data. Using social networks, the emotional level can be monitored or analyzed. In this research work we will classify data from social networks such as twitter regarding suicide and classify it as: active suicidal thinking, passive suicidal thinking, sarcasm related to suicidal thinking, tweets related to suicide (suicide awareness, news, suicide talk) and others.

Keywords: *Suicide, suicidal thoughts, natural language processing.*

Introducción

Las tasas de suicidio han sido un tema preocupante en muchos países, entre ellos Perú. Según una revista de Neuro-Psiquiatría, la cantidad de personas que se suicidaron en un lapso del 2017-2021 incluye un total de 2579 suicidios [8]. Esto resalta la urgencia de implementar estrategias efectivas de prevención del suicidio. Estudios han demostrado que el suicidio es prevenible a través de intervenciones oportunas y específicas. En este contexto, las aplicaciones de NLP pueden desempeñar un papel importante a la hora de señalar e identificar a las personas en situación de riesgo mediante el análisis de patrones lingüísticos y sentimientos en las publicaciones de las redes sociales.

El objetivo de este proyecto de aplicación NLP es detectar pensamientos suicidas expresados en Twitter. Sabemos que en la actualidad las redes sociales se han convertido en una parte integral de la vida diaria de muchas personas y pueden tener un impacto significativo en la salud mental. Muchas personas utilizan las redes sociales para conectarse con otros, pero también pueden exponerse a una gran cantidad de información negativa, comparación social y ciberacoso. Twitter es una de las plataformas de redes sociales más populares y la gente la utiliza a menudo para expresar sus emociones y pensamientos. El proyecto desarrolla un modelo de aprendizaje automático que pueda analizar tweets para identificar patrones lingüísticos asociados a los pensamientos suicidas y marcar tweets que puedan indicar un riesgo de suicidio entre las categorías de activo o pasivo, usando modelos pre entrenados como Roberta y Bert. Identificar los pensamientos suicidas en los tweets podría ayudar a intervenir precozmente, lo que a su vez puede mejorar la calidad de vida de las personas afectadas y salvar vidas.

Preguntas de Investigación

El objetivo del proyecto es desarrollar un algoritmo basado en procesamiento de lenguaje natural (NLP) para clasificar el contenido de texto subido a las redes sociales, para identificar tweets respecto al suicidio.

- RQ1: ¿Que palabras indican posibles pensamientos suicidas? (activa, pasiva, sarcástica, relacionado al tema u otros)
- RQ2: ¿Qué modelo será más efectivo para este proyecto? Ya sea de Bert o Roberta

DATOS

Los datos clasificados se obtuvieron de Kaggle [6], hay tres columnas principales, los ids de los tweets, la clasificación y especificar si el dato se usara para entrenar al modelo, validarlo y otros. Estos datos cuentan con cerca de 81519

tweets. Este conjunto de datos se caracteriza por ser el único que toma en cuenta una clasificación para los tweets relacionados al suicidio considerando las categorías de activo, cuando el texto denota alguna intención suicida, pasiva las cuales son potencialmente activas en un futuro, sarcasmo dado que es propio de la plataforma, relacionado al suicidio como noticias o charlas y finalmente otros. La importancia de una clasificación es que los tipos de pensamientos suicidas pasivos o activos requieren un apoyo y ayuda diferente.

Para obtener los datos se tuvo de base el código proporcionado por raza8899 (usuario de Github). El código toma el id del tweet, que lo obtenemos del dataset, y se devolvemos la sentencia que es lo que se va a analizar. Finalmente se crea un nuevo dataset con las sentencias, su respectiva etiqueta y el dataset al que pertenece. Se mostrará el pseudocódigo.

Funciones

- `Get_specific_tweet`: Recibe el id del tweet para retornar solo el contenido de este.
- `Snscraper_get_tweet`: Esta función usa `TwitterTweetScrapper` de la librería `snscraper` tomando como parametros el id del tweet y el modo `single`.
- `Get_items`: Devuelve los objetos (tweets) obtenidos de un conjunto de datos.

Pseudocódigo

```
importar libreria snsrape
importar libreria json
importar libreria pandas
importar libreria numpy

función get_specific_tweet(id_tweet):
for tweet in enumerate(snscraper_get_tweet()
.get_items())
dato = tweet transformado a json

# Verificamos que se obtenga el contenido
if "rawContent" in dato:
retornamos el rawContent
else:
return '0'

df = leemos el dataset usando pandas
```

```
ids = lista de la columna de ids del dataset
lab = lista de la columna de labels del
dataset
dat = lista de la columna de dataset del
dataset
test_list = convertimos los ids a enteros
tweet = array vacio para agregar los tweets
con el contenido

for i in range (longitud del dataset o
longitud deseada):
try:
tweet = get_specific_tweet(test_list[i])
cargamos el tweet al arreglo de tweets
junto con su respectivo label y dataset
except:
pass #pasamos los tweets a los que no se
tuvo acceso

arr = convertir el arreglo de tweets a un
array de numpy
con el nuevo arreglo de datos lo pasamos a
formato csv usando de cabecera sentence,
label y dataset
```

REVISION LITERARIA

Para identificar otras soluciones propuestas se realizó una revisión literaria respecto al tema utilizando palabras claves como: suicide, suicide ideation, nlp, clasification.

A. Natural language processing applied to mental illness detection: a narrative review

Este artículo [4] propone la revisión de investigaciones relacionadas al procesamiento de lenguaje natural aplicado en la detección de enfermedades mentales. Principalmente identifica varias fuentes de datos como son posts en redes sociales, entrevistas, repositorio clínico y escritura narrativa. Detalla los métodos de machine learning y deep learning que son en su mayoría utilizados. Se concluye que ambos métodos son eficientes pero que dependen mucho de los datos utilizados, en especial el volumen y la fuente, debido a que no existen muchos conjuntos de datos. Los temas más

recurrentes en este ámbito de investigación son la depresión y el suicidio, las bases de datos más usadas son Twitter y Reddit debido a su facilidad de obtención de los datos, el lenguaje más usado para la entrenar a los modelos es el inglés.

B. Detection of suicide-related posts in Twitter data streams. IBM Journal of Research and Development

La detección temprana de señales de advertencia de suicidio en las redes sociales, específicamente en Twitter, es necesaria para prevenir el suicidio y mejorar la salud mental de las personas. Sin embargo, la identificación manual de publicaciones relacionadas con el suicidio es un proceso muy lento y costoso. Por ello se necesita un enfoque innovador que pueda detectar automáticamente estas publicaciones y proporcionar una alerta temprana a los profesionales de la salud mental. Este artículo [2] presenta un nuevo enfoque que utiliza técnicas avanzadas de NLP para detectar señales de advertencia de suicidio en Twitter.

Para resolver el problema de la detección de pensamientos suicidas en las redes sociales en línea, los autores proponen un nuevo enfoque que utiliza Twitter para cuantificar las señales de advertencia de suicidio y detectar los mensajes con contenido relacionado con el suicidio. Para ello se utilizan dos métodos diferentes: el procesamiento del lenguaje natural (NLP) y el aprendizaje automático. El método de NLP combina características generadas a partir del texto basadas en un conjunto de léxicos, que se componen de temas lingüísticos comúnmente exhibidos por las personas en situación de riesgo. El segundo método, denominado clasificador de angustia, se basa en el aprendizaje automático.

En el estudio, se concluye que el enfoque propuesto resulta efectivo en la detección de contenido relacionado con el suicidio en Twitter. Los autores se enfocaron en detectar contenido relacionado con la angustia y el suicidio, utilizando los dos enfoques mencionados anteriormente. Aunque el estudio tiene algunas limitaciones, como la prueba con solo dos usuarios de Twitter, se sugiere que futuras investigaciones podrían ampliar las pruebas con más usuarios y líneas temporales, especialmente con aquellos que se identifiquen como personas con enfermedades mentales. En general, este trabajo presenta un enfoque innovador con el potencial de salvar vidas y prevenir el suicidio al detectar las ideas suicidas en las redes sociales en línea.

C. Exploring and Learning Suicidal Ideation Connotations on Social Media with Deep Learning

Debido a esta creciente disponibilidad de contenidos en las redes sociales (como Twitter, Facebook y Reddit, etc.), urge identificar a las personas afectadas y ofrecerles ayuda [3].

El artículo [3] busca realizar la comparación entre 3 arquitecturas de deep learning. Primero hizo la recolección de los datos, los clasifico manualmente y luego se utilizaron en tres modelos para determinar cuál era el mejor.

Los datos fueron recolectados de 3 fuentes principales, sitios de micro-blogueo, Tumnlr y Reddit. Donde 300 datos fueron de forums de suicidio y 2000 datos fueron de posts seleccionados aleatoriamente en Tumblr, entre otras plataformas el conjunto de datos total consistía en 5217 sentencias. Para la clasificación utilizaron a tres observadores quienes manualmente identificaron si el dato se consideraba relacionado al suicidio o no. En caso de ambigüedad se optó por colocarlo como suicida. Para la categoría de suicidio se considera que este explicito el intento de suicidio del usuario. Para la categoría de no suicidio se considera sarcasmo, reportes, o textos relacionados a la sensibilización del tema.

Para el preprocesamiento se quitaron las urls, la mención de usuarios, hashtags y los signos como el guion y el apóstrofe (-, '), se consideraron las palabras que repitieran las letras. Luego esta información paso por tres arquitecturas. Recurrent Neural Network (RNN) modelo que se destaca por ser eficiente al clasificar información secuencial de longitud arbitraria. Long Short Term Memory (LSTM), método adecuado para clasificar, procesar y predecir series temporales y captar dependencias a largo plazo en las frases. Finalmente, C-LSTM, este modelo combina la eficiencia de codificación espacial y la extracción automática de características de la red neuronal convolucional (Convolutional neural network, CNN), combinada con la eficaz clasificación de textos mediante LSTM.

El procesamiento de clasificación se basó en tres etapas. Integración de palabras, integración de frases y clasificación. Para la primera etapa se usaron las N palabras más frecuentes en un tweet para formar una capa, se utilizó word2vec. La segunda etapa se usó en el modelo C-LSTM, donde se añade un CNN unidimensional y luego una capa de integración. La integración de sentencias alimenta la capa de LSTM. Finalmente, la clasificación donde el modelo alimenta las sentencias aprendidas. En comparación a los tres modelos el que dio mejores resultados fue el modelo híbrido C-LSTM.

D. Detection of Suicidal Ideation on Twitter using Machine Learning and Ensemble Approaches

El suicidio es uno de los principales problemas de salud pública que consume muchas vidas. Según las estadísticas de la Organización Mundial de la Salud (OMS), alrededor de un millón de personas mueren por suicidio cada año. La OMS menciona que el suicidio es la principal causa de muerte entre los adolescentes y la sexta entre los adultos. Los factores incluyen problemas personales como desesperanza, abuso de sustancias, ansiedad, esquizofrenia; factores sociales como el aislamiento de la sociedad, la pérdida de seres queridos, el desempleo, etc. A pesar del creciente número de casos de suicidio, se puede prevenir hasta cierto punto mediante la comprensión de los factores de riesgo relacionados con el comportamiento suicida en las primeras etapas del proceso suicida. El proceso de suicidio comienza

con pensamientos o ideas suicidas. Luego madura hasta el intento de suicidio y finalmente hasta el suicidio consumado [1].

Hay una creciente inclinación de las personas a formar comunidades en línea e interactuar entre sí. Las investigaciones han revelado que las personas se sienten cómodas para hablar sobre tendencias suicidas en línea en lugar de hablar de ello cara a cara; la razón es una sensación de autocontrol y la característica anónima de las redes sociales. Existen varios casos en los que las víctimas de suicidio también revelaron sus sentimientos finales antes de su muerte en las redes sociales.

El trabajo de investigación hace contribuciones novedosas a la literatura ya existente de las siguientes maneras.

- El trabajo extrae una serie de características relevantes para diferenciar entre tweets suicidas y no suicidas con la ayuda de una nueva técnica de ingeniería de características.
- El documento construye un nuevo conjunto de datos extrayendo los tweets relacionados con el suicidio y el no suicidio de Twitter. El conjunto de datos se anota y luego se usa para entrenar su modelo para la diferenciación binaria (suicida y no suicida) de los tweets. El trabajo revela el conocimiento oculto desde la perspectiva de la minería de datos.
- La evaluación comparativa se realiza en varios métodos de aprendizaje automático y conjuntos para distinguir su eficacia.

En el documento, se implementan varios algoritmos de aprendizaje automático y enfoques de conjuntos como árboles de decisión, Naive Bayes, Multinomial Naive Bayes, Support vector machine (SMO), Regresión, Bagging, Random Forest, AdaBoost, votación y Stacking para clasificar los tweets relacionados con el suicidio. En dos grupos usando los tweets reales extraídos a través de la API de Twitter. Random Forest resultó ser el más efectivo y funciona mejor con una precisión del 98,5%, una precisión del 98,7% y un valor de recuperación del 98,2%.

DISEÑO

Dada la comparativa entre los dos modelos se hicieron dos pipelines, uno para Roberta sin un preprocesamiento y otro para Bert con los datos limpios, de esta forma medir la eficiencia de cada modelo.

A. Pipeline Bert

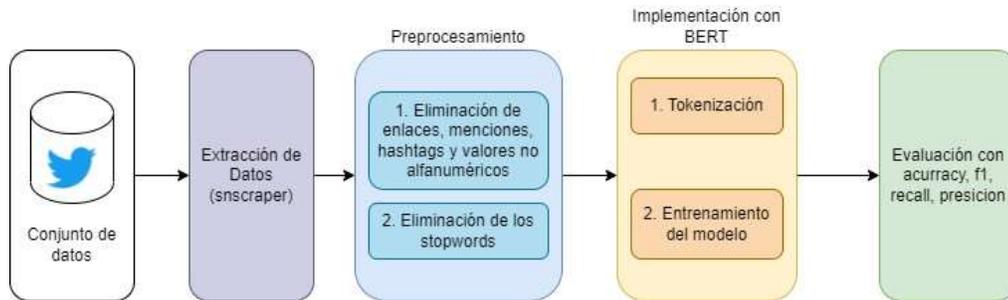


Figura 1. Marco de Trabajo para la Detección de suicidio en tweets con Bert

B. Pipeline Roberta

IMPLEMENTACION Y RESULTADOS

Se utilizo Google Colab para entrenar los modelos, ah 1 se utilizaron los códigos de (referencias a los códigos). En ambos casos se utilizaron los recursos de la GPU para el entrenamiento del modelo.

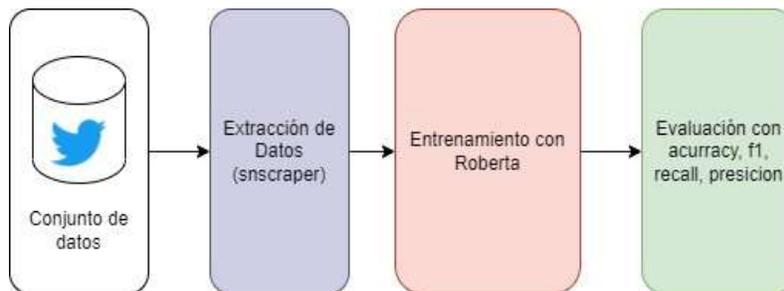


Figura 2. Marco de Trabajo para la Detección de suicidio en tweets con Roberta

A. Bert

BERT, que significa Bidirectional Encoder Representations from Transformers, es un modelo de lenguaje basado en Transformers que fue presentado por Google en 2018. Representa un avance significativo en el procesamiento del lenguaje natural y ha demostrado un rendimiento excepcional en una amplia gama de tareas relacionadas con el lenguaje. Por eso es ampliamente utilizado y reconocido por su capacidad para comprender y generar texto en un contexto más profundo.

Ventajas:

- **Contexto bidireccional:** Una de las principales ventajas de BERT es su capacidad para comprender el contexto en ambos sentidos, es decir, tanto las palabras que lo preceden como las que lo siguen. Esto le permite capturar las sutilezas de las relaciones entre palabras y generar representaciones de palabras más ricas y precisas.
- **Representaciones profundas:** BERT utiliza una arquitectura basada en Transformers con múltiples capas, lo que le permite capturar relaciones semánticas y sintácticas complejas en el texto. A través del entrenamiento en grandes cantidades de datos, BERT desarrolla una comprensión profunda del lenguaje y es capaz de inferir significados incluso en oraciones con estructuras gramaticales complejas.
- **Preentrenamiento y adaptabilidad:** BERT se preentrena en grandes corpus de texto sin etiquetas, lo que le permite aprender representaciones generales del lenguaje. Una vez preentrenado, se puede adaptar a tareas específicas de NLP mediante un proceso de afinación, lo que permite su uso en una variedad de aplicaciones y dominios.
- **Rendimiento sobresaliente:** BERT ha demostrado un rendimiento excepcional en una amplia gama de tareas de procesamiento del lenguaje natural, como la clasificación de texto, el etiquetado de entidades, la respuesta a preguntas y la traducción automática. Su capacidad para comprender el contexto y las relaciones complejas entre palabras lo convierte en una opción destacada en comparación con modelos previos.

Desventajas:

- **Alta demanda computacional:** Debido a su arquitectura basada en Transformers y su gran tamaño, BERT requiere una cantidad significativa de recursos computacionales para entrenar y utilizar. Esto puede ser una limitación en entornos con recursos limitados o en aplicaciones que requieren un procesamiento rápido en tiempo real.
- **Dependencia de grandes conjuntos de datos:** Para alcanzar su máximo rendimiento, BERT necesita ser entrenado en grandes cantidades de datos. Esto puede ser un desafío en dominios de nicho o en idiomas con recursos limitados, donde puede ser difícil reunir suficiente información para entrenar el modelo de manera efectiva.
- **Dificultad para manejar ambigüedades:** Aunque BERT tiene un rendimiento sobresaliente, aun puede tener dificultades para manejar casos de ambigüedad en el lenguaje. En situaciones donde el contexto no es suficiente para determinar el significado exacto de una palabra o frase, BERT puede producir resultados imprecisos o inconsistentes.

- **Falta de aplicabilidad:** Dado que BERT es un modelo de aprendizaje automático de caja negra, puede resultar difícil comprender como toma decisiones o llega a ciertas conclusiones. Esto puede plantear desafíos en escenarios donde se requiere una explicación clara y transparente del proceso de toma de decisiones.

B. RoBERTa

RoBERTa es un modelo de lenguaje basado en Transformers que se basa en el éxito de BERT y mejora a un mas su rendimiento. Fue presentado por Facebook AI en 2019 y se enfoca en el preentrenamiento sin supervisión para aprender representaciones de alta calidad del lenguaje.

Ventajas:

- **Mejor preentrenamiento:** RoBERTa utiliza una estrategia de preentrenamiento más extensa en comparación con BERT. Se entrena en grandes conjuntos de datos sin etiquetas, utilizando técnicas como el entrenamiento de máscara de palabras (masking) y la predicción de oraciones siguientes. Esto resulta en representaciones de palabras más sólidas y una mejor comprensión del lenguaje.
- **Mayor capacidad de generalización:** Al tener un enfoque de preentrenamiento más exhaustivo, RoBERTa logra una mayor capacidad de generalización. Puede capturar mejor las relaciones semánticas y sintácticas complejas en el texto, lo que le permite adaptarse a diferentes tareas de procesamiento del lenguaje natural con un rendimiento mejorado.
- **Afinación más efectiva:** RoBERTa se beneficia de técnicas de afinación mejoradas. Después del preentrenamiento, el modelo se ajusta a tareas específicas de NLP mediante un proceso de afinación en conjuntos de datos etiquetados más grandes y diversificados. Esto permite una adaptación más efectiva y un mejor rendimiento en comparación con BERT.
- **Resultados líderes en el estado del arte:** RoBERTa ha logrado resultados sobresalientes en una amplia gama de tareas de procesamiento del lenguaje natural, incluyendo clasificación de texto, extracción de información, traducción automática y más. Su capacidad para capturar el contexto y las relaciones complejas en el lenguaje lo coloca en la vanguardia del rendimiento en NLP.

Desventajas:

- **Requiere recursos computacionales significativos:** Al igual que BERT, RoBERTa es computacionalmente intensivo y requiere una gran cantidad de recursos para su entrenamiento y uso eficiente. Esto puede limitar su aplicabilidad en entornos con recursos limitados o en aplicaciones que necesitan un procesamiento rápido en tiempo real.

- Dificultad para la interpretación y explicación: Debido a su complejidad y enfoque de aprendizaje automático de caja negra, RoBERTa puede ser difícil de interpretar y explicar. Comprender como el modelo toma decisiones o llega a conclusiones específicas puede resultar desafiante, lo que puede ser problemático en aplicaciones que requieren transparencia y aplicabilidad.
- Dependencia de grandes conjuntos de datos: Al igual que BERT, RoBERTa también se beneficia del entrenamiento en grandes conjuntos de datos. Esto puede ser un desafío en dominios o idiomas con recursos limitados, donde puede ser difícil reunir suficiente información para el entrenamiento efectivo del modelo.
- Sensibilidad al ruido en los datos de entrenamiento: RoBERTa puede ser sensible al ruido presente en los datos de entrenamiento. La presencia de datos incorrectos o mal etiquetados puede afectar negativamente su rendimiento y conducir a resultados imprecisos o inconsistentes en ciertos casos.

Resultados

Se utilizó sklearn para evaluar ambos modelos. Los resultados de ambos modelos fue el siguiente, Bert obtuvo 0.96684 su mejor valor en precisión lo que significa que el modelo tiene una baja tasa de falsos positivos; mientras que Roberta obtuvo 0.95800 en recall, indica que el modelo es efectivo al capturar la mayoría de las instancias positivas y tiene una baja tasa de falsos negativos.

Tabla I Resultados de los modelos.

Modelo	Accuracy	F1	Recall	Precision
Bert	0.95802	0.95756	0.95802	0.96684
Roberta	0.93798	0.94447	0.95800	0.93187
Roberta DataSet Limpio	0.954039	0.95213	0.95486	0.951526

CONCLUSIONES

El modelo Bert muestra una mayor precisión (0.96684) y exactitud (0.95802), lo que lo hace adecuado para aplicaciones donde minimizar los falsos positivos y lograr una precisión global en la clasificación son cruciales. Sin embargo, muestra un nivel ligeramente inferior de recall (0.95802) en comparación con Roberta, lo que indica que podría perder algunos casos positivos. Además, su puntuación F1 (0.95756) sugiere un ligero equilibrio entre precisión y recall.

Por otro lado, el modelo Roberta demuestra un mayor recall (0.95800) y puntuación F1 (0.94447), lo que indica su fortaleza al capturar casos positivos y mantener un buen equilibrio entre precisión y recall. Esto lo convierte en una

opción confiable cuando reducir los falsos negativos es importante. Sin embargo, tiene una precisión más baja (0.93187) y una exactitud menor (0.93798) en comparación con Bert, lo que sugiere una posible mayor tasa de falsos positivos y un mayor índice general de clasificación incorrecta.

Es importante tener en cuenta que la elección entre Bert y Roberta debe basarse en los requisitos específicos de la tarea en cuestión, así como en la importancia relativa de la precisión, el recall y la exactitud en el contexto de la aplicación.

En términos de precisión y exactitud, tanto Bert como Roberta muestran resultados inferiores en comparación con el modelo Random Forest mencionado en el artículo, que logra una precisión del 98,7%. Esto sugiere que el modelo Random Forest es más preciso y capaz de clasificar los casos correctamente con mayor certeza que los modelos de lenguaje natural como Bert y Roberta.

Sin embargo, se debe tener en cuenta que los modelos de lenguaje natural como Bert y Roberta pueden tener otras ventajas, como la capacidad de comprender y generar texto en lenguaje natural de manera más general, lo que los hace más versátiles en diferentes aplicaciones.

Respecto a las palabras más recurrentes se recopilaron en una imagen.

Pensamientos Suicidas



Figura 3. Palabras Recurrentes

TRABAJOS FUTUROS

Para los trabajos futuros hay varias consideraciones a tener en cuenta al trabajar con datos relacionados con la ideación suicida en Twitter, especialmente teniendo en cuenta los desafíos encontrados durante la recopilación de datos.

En primer lugar, sería beneficioso expandir el conjunto de datos para mejorar el rendimiento de los modelos. Dado la dificultad para obtener datos relacionados con los pensamientos suicidas, se deben realizar esfuerzos para recopilar un conjunto de datos más amplio para las categorías de activo, pasivo, sarcasmo y relaciones dado que la suma de estos datos no se compara con la categoría de otros que son la mayor parte. Esto podría implicar colaborar con organizaciones de salud mental, investigadores o utilizar técnicas de búsqueda avanzadas para capturar una gama más amplia de tweets relacionados con la ideación suicida.

En segundo lugar, considerando la naturaleza sensible del tema, sería importante incorporar medidas adicionales para garantizar el manejo ético de los datos. Esto podría incluir la implementación de técnicas robustas de anonimización para proteger la privacidad y confidencialidad de las personas involucradas en los tweets. Colaborar con profesionales de salud mental y expertos en ética podría proporcionar orientación valiosa para desarrollar protocolos adecuados para la recopilación y el uso de datos.

Además, explorar la información contextual y los matices del lenguaje específicos de Twitter podría ser beneficioso para mejorar la comprensión de los modelos sobre los diferentes tipos de ideación suicida. Con ello obtener más información sobre que características tienen las personas relacionado a estos tipos de pensamientos. Esto podría mejorar la capacidad de los modelos para clasificar y diferenciar con precisión entre la ideación suicida activa y pasiva, el sarcasmo, las discusiones relacionadas con el suicidio y otras categorías relevantes.

Contribución de Autoría

Juliana Berrios Butron: [Conceptualización](#), [Análisis formal](#), [Investigación](#), [Visualización](#), [Metodología](#), [Software](#), [Validación](#), [Redacción - borrador original](#), [Curación de datos](#), [Escritura, revisión y edición](#). **Daria López Franco:** [Conceptualización](#), [Análisis formal](#), [Investigación](#), [Visualización](#), [Metodología](#), [Software](#), [Validación](#), [Redacción - borrador original](#), [Curación de datos](#), [Escritura, revisión y edición](#). **Dante Rodolfo Tárraga:** [Conceptualización](#), [Análisis formal](#), [Investigación](#), [Visualización](#), [Metodología](#), [Software](#), [Validación](#), [Redacción - borrador original](#), [Curación de datos](#), [Escritura, revisión y edición](#).

Referencias

- [1] Rabani, S. T., Khan, Q. R., & Khanday, A. M. U. D. (2020). Detection of suicidal ideation on Twitter using machine learning & ensemble approaches. *Baghdad Science Journal*, 17(4), 1328-1339.
- [2] M. J. Vioules, B. Moulahi, J. Az` e and S. Bringay, (2018). Detection of suicide-related posts in Twitter data streams. *IBM Journal of Research and Development*, vol. 62, no. 1, pp. 7:1-7:12, 1 Jan.-Feb. 2018, doi: 10.1147/JRD.2017.2768678.
- [3] Sawhney, R., Manchanda, P., Mathur, P., Shah, R., & Singh, R. (2018, October). Exploring and learning suicidal ideation connotations on social media with deep learning. In *Proceedings of the 9th workshop on computational approaches to subjectivity, sentiment and social media analysis* (pp. 167-175).
- [4] Zhang, T., Schoene, A. M., Ji, S., & Ananiadou, S. (2022). Natural language processing applied to mental illness detection: a narrative review. *NPJ digital medicine*, 5(1), 46.
- [5] Otter, D. W., Medina, J. R., & Kalita, J. K. (2020). A survey of the usages of deep learning for natural language processing. *IEEE transactions on neural networks and learning systems*, 32(2), 604-624.
- [6] Lech, N. (2008). Suicidal ideation on Twitter.[Database].
<https://www.kaggle.com/datasets/natalialech/suicidal-ideation-on-twitter>
<https://www.kaggle.com/datasets/natalialech/suicidal-ideationon-twitter>
- [7] Rabani, S. T., Khan, Q. R., & Khanday, A. M. U. D. (2020). Detection of suicidal ideation on Twitter using machine learning & ensemble approaches. *Baghdad Science Journal*, 17(4), 1328-1339.
- [8] Coco C. C., Joel A. P., Carlos S. Francklin C. H., Waldir P. H. (2022). Suicidios en el Perú: Descripción epidemiológica a través del Sistema Informático Nacional de Defunciones (SINADEF) en el periodo 2017- 2021.