



EDUCACIÓN

SECRETARÍA DE EDUCACIÓN PÚBLICA



TECNOLÓGICO
NACIONAL DE MÉXICO

Tecnológico Nacional de México

Centro Nacional de Investigación
y Desarrollo Tecnológico

Tesis de maestría

Detección automática de tendencias suicidas
en redes sociales utilizando embeddings de
texto

presentada por

Lic. Juan Pablo Álvarez López

como requisito para la obtención del grado de
Maestro en Ciencias de la Computación

Director de tesis

Dr. Noé Alejandro Castro Sánchez

Codirector de tesis

Dra. Leonor Rivera Rivera

Cuernavaca, Morelos, México. Enero de 2025.



Cuernavaca, Mor.,
No. de Oficio:
Asunto:

13/diciembre/2024
SAC/387/2024
Autorización de
impresión de tesis

JUAN PABLO ÁLVAREZ LÓPEZ
CANDIDATO AL GRADO DE MAESTRO EN
CIENCIAS DE LA COMPUTACIÓN
PRESENTE

Por este conducto, tengo el agrado de comunicarle que el Comité Tutorial asignado a su trabajo de tesis titulado "**Detección automática de tendencias suicidas en redes sociales utilizando embeddings de texto**", ha informado a esta Subdirección Académica, que están de acuerdo con el trabajo presentado. Por lo anterior, se le autoriza a que proceda con la impresión definitiva de su trabajo de tesis.

Esperando que el logro del mismo sea acorde con sus aspiraciones profesionales, reciba un cordial saludo.

ATENTAMENTE

Excelencia en Educación Tecnológica®
"Conocimiento y Tecnología al Servicio de México"

CARLOS MANUEL ASTORGA ZARAGOZA
SUBDIRECTOR ACADÉMICO

c.c.p. Departamento de Ciencias Computacionales
Departamento de Servicios Escolares

CMAZ/lmz





TECNOLÓGICO
NACIONAL DE MÉXICO



Centro Nacional de Investigación
y Desarrollo Tecnológico
Departamento de Ciencias Computacionales

Cuernavaca, Mor. 10/diciembre/2024
Asunto: Liberación de producto académico
Comité tutorial

JUAN PABLO ÁLVAREZ LÓPEZ
CANDIDATO AL GRADO DE MAESTRO EN CIENCIAS
DE LA COMPUTACIÓN
PRESENTE

Por este conducto, tengo el agrado de comunicarle que el Comité Tutorial asignado a su trabajo de tesis titulado "Detección automática de tendencias suicidas en redes sociales utilizando embeddings de texto", avala que usted tiene el siguiente producto académico, derivado de su investigación.

Artículo: "Word Embeddings and Machine Learning Classifiers Applications for Automatic Detection of Suicide-Tendencies in Social Media". ISSN 0361-7688, Programming and Computer Software, 2024, Vol. 50, No. 8, pp. 611-619.

Sin más por el momento, reciba un cordial saludo.

ATENTAMENTE

Excelencia en Educación Tecnológica-
"Conocimiento y Tecnología al Servicio de México"

Dr. Noé Alejandro Castro Sánchez
Director de Tesis

Dra. Leonor Rivera Rivera
Codirectora de Tesis

Dr. Juan Gabriel González Serna
Revisor 1

Dr. Dante Mújica Vargas
Revisor 2

Dr. Nimrod González Franco
Revisor 3

C.c.p. Jefe (a) del Depto. de Servicios Escolares



Av. Interior Inter... N, Col. Palmira, C. P. 62490, Cuernavaca, Morelos
Tel. 01 (777) 3627770, ext. 3201, E-mail: dcc@cenidet.tecnm.mx | cenidet.tecnm.mx





TECNOLÓGICO
NACIONAL DE MÉXICO



Centro Nacional de Investigación
y Desarrollo Tecnológico
Departamento de Ciencias Computacionales

Cuernavaca, Mor., 11/diciembre/2024
OFICIO/DCC/318/2024

Asunto: Aceptación de documento de tesis
CENIDET-AC-M14-OFICIO

CARLOS MANUEL ASTORGA ZAGOZA
SUBDIRECTOR ACADÉMICO
PRESENTE

Por este conducto, los integrantes del Comité Tutorial de JUAN PABLO ÁLVAREZ LÓPEZ, con número de control M22CE042, de la Maestría en Ciencias de la Computación, le informamos que hemos revisado el trabajo de tesis de grado titulado "Detección automática de tendencias suicidas en redes sociales utilizando embeddings de texto", y hemos encontrado que se han atendido todas las observaciones que se le indicaron, por lo que hemos acordado aceptar el documento de tesis y le solicitamos la autorización de impresión definitiva.

ATENTAMENTE

*Excelencia en Educación Tecnológica-
"Conocimiento y Tecnología al Servicio de México"*



Dr. Noé Alejandro Castro Sánchez
Director de Tesis



Dra. Leonor Rivera Rivera
Codirectora de Tesis



Dr. Juan Gabriel González Serna
Revisor 1



Dr. Dante Mújica Vargas
Revisor 2



Dr. Nimrod González Franco
Revisor 3



Av. Interior Inter... N, Col. Palmira, C. P. 62490, Cuernavaca, Morelos
Tel. 01 (777) 3627770, ext. 3201, E-mail: dcc@cenidet.tecnm.mx | cenidet.tecnm.mx



2024
Felipe Carrillo
PUERTO

Resumen

El suicidio es un problema de salud mental que afecta a las personas debido a varios factores como depresión, problemas económicos, familiares o personales, entre otros y afecta a cualquier estatus social en el mundo, donde cerca de 700,000 personas mueren al año según la Organización Mundial de la Salud. Debido a tabúes o prejuicios la gente tiende a no consultar con profesionales por ayuda, por lo cual recurren a medios digitales como foros o chats en redes sociales para compartir pensamientos y sentimientos. Con esto en cuenta se concluye que estas redes sociales tienen un gran contenido sobre pensamientos de gente con tendencias suicidas que puede ser analizado de forma automática mediante modelos de inteligencia artificial. Este trabajo consistió en analizar textos con tendencias suicidas e implementar herramientas de procesamiento de lenguaje natural junto con modelos de inteligencia artificial para detectar de forma automática comentarios de personas con posible tendencia suicida. Los resultados indican que utilizando modelos largos de embeddings como GPT-3 Large y junto a modelos de redes neuronales se pueden obtener resultados de +90% en métricas como precisión, exactitud, sensibilidad y puntuación F1.

Abstract

Suicide is a mental health problem that affects people due to various factors such as depression, economic, family or personal problems, among others, and affects any social status in the world, where about 700,000 people die each year according to the World Health Organization. Due to taboos or prejudices people tend not to consult professionals for help, so they resort to digital media such as forums or chats on social networks to share thoughts and feelings. It is concluded that these social networks have a large amount of content about thoughts of people with suicidal tendencies that can be analyzed automatically using artificial intelligence models. This work consisted of analyzing texts with suicidal tendencies and implementing natural language processing tools together with artificial intelligence models to automatically detect people with possible suicidal tendencies. The results indicate that using long embedding models such as GPT-3 Large and together with neural network models, results of +90% can be obtained in metrics such as precision, accuracy, recall and F1 score.

Dedicatoria

Este trabajo se lo dedico a mi familia: madre, padre y hermano. Sin ustedes no sería la persona que soy hoy. Todos y cada uno de sus sentimientos se han cultivado en mí para ser una mejor persona. Su apoyo incondicional, sus palabras de aliento, su motivación, sus regaños y su cariño vivirán siempre conmigo y con mis ideales personales y profesionales.

A mis abuelos, que han cumplido su deber como abuelos, de consentirme y guiarme, de contarme sus experiencias y anécdotas para que siga por el mejor camino.

A mis amigos, que son las personas que más quiero en el mundo y que han estado para mí en las buenas y en las malas, sin esperar nada a cambio. Me han animado, apoyado, inspirado, aconsejado y regañado, pensando siempre en mí y mi bienestar. Los quiero a todos por igual y no los cambiaría por nada del mundo.

Finalmente, a mi pajera, que ha sido mi inspiración y mi fuente de ideas, que me ayuda a reflexionar y pensar en todas las posibilidades antes de cerrarme las puertas solo y que decidió creer en mí cuando yo mismo dejaba de creer. Es la persona más inteligente, amable, linda y respetuosa que he conocido en mi vida y espero que esa energía pueda siempre estar conmigo.

“Al final todo estará bien, y si no está bien,
entonces aún no es el final”

Pelusa Caligari

Agradecimientos

Al Consejo Nacional de Humanidades Ciencias y Tecnologías (CONAHCyT) por su Programa Nacional de Posgrados de Calidad (PNCP) por medio del cual me fue otorgada una beca para ser estudiante de tiempo completo y desarrollar este trabajo de investigación, así como al Tecnológico Nacional de México, Campus Centro Nacional de Investigación y Desarrollo Tecnológico (CENIDET).

Doy las gracias a los profesores que aportaron con mi formación como Maestro en Ciencias de la Computación, especialmente al Dr. Noe Alejandro Castro Sanchez y a mi codirectora Dra. Leonor Rivera Rivera. De igual manera agradezco a mis revisores Dr. Dante Mújica Vargas, Dr. Juan Gabriel González Serna y Dr. Nimrod González Franco.

Índice

1	Introducción.....	8
1.1	<i>Planteamiento del problema.....</i>	<i>10</i>
1.2	<i>Objetivo.....</i>	<i>11</i>
1.2.1	Objetivo General	11
1.2.2	Objetivos Específicos	11
1.3	<i>Justificación.....</i>	<i>11</i>
1.4	<i>Alcances y limitaciones.....</i>	<i>12</i>
1.4.1	Alcances	12
1.4.2	Limitaciones.....	12
1.5	<i>Organización del documento</i>	<i>13</i>
2	Fundamento Teórico	14
2.1	<i>Marco Conceptual.....</i>	<i>14</i>
2.1.1	Suicidio	14
2.1.2	Factores de riesgo suicida.....	15
2.1.3	Proceso suicida	16
2.1.4	Programa nacional de prevención del suicidio	17
2.1.5	Redes sociales.....	19
2.1.6	Procesamiento de lenguaje natural (PLN)	19
2.1.7	Enfoques para el análisis de texto	20
2.1.8	Traducción de textos.....	21
2.1.9	Algoritmos de clasificación.....	21
2.2	<i>Estado del arte.....</i>	<i>25</i>
2.2.1	Discusión del estado del arte.....	34
3	Metodología de solución	37
3.1	<i>Búsqueda y recopilación de datos, generación y limpieza de CSV.....</i>	<i>38</i>
3.2	<i>Preprocesamiento de datos.....</i>	<i>40</i>
3.3	<i>Representaciones numéricas.....</i>	<i>42</i>
3.4	<i>Algoritmos de clasificación empleados.....</i>	<i>45</i>
4	Experimentos y resultados.....	46
4.1	<i>Clasificación con textos traducidos.....</i>	<i>53</i>
4.2	<i>Clasificación con textos traducidos y redes neuronales.....</i>	<i>54</i>
5	Conclusiones	58
6	Producción científica.....	60
7	Referencias.....	62

Índice imágenes

Imagen 1 Metodología de solución.....	37
Imagen 2 Tamaño por conjunto de datos del repositorio	38
Imagen 3 Resultados con Word2Vec.....	48
Imagen 4 Resultados con GloVe.....	49
Imagen 5 Resultados con MPNet.....	50
Imagen 6 Resultados con GPT-3 Small	51
Imagen 7 Resultados con GPT-3 Large.....	51
Imagen 8 Mejores combinaciones basándose en la puntuación de la validación cruzada, umbral 85%+	52
Imagen 9 Mejores combinaciones basándose en la puntuación de la validación cruzada, umbral 85%+ en idioma español.....	55
Imagen 10 Diseño de red neuronal.....	56
Imagen 11 Resultados inglés - español de la red neuronal	57

Índice de tablas

Tabla 1	Tabla comparativa del estado del arte.....	35
Tabla 2	Comparativa de técnicas de representaciones numéricas.	42
Tabla 3	Puntajes de la validación cruzada entre embedding y clasificador ...	46
Tabla 4	Hiperparámetros implementados por embedding – clasificador.....	47
Tabla 5	Sets de datos para traducción	53
Tabla 6	Efectividad en textos traducidos con algoritmos de clasificación	54

1 Introducción

El suicidio se define como el acto de quitarse la vida tras infligirse daño a sí mismo (2022). El suicidio constituye un grave problema de salud pública que impacta a familias, comunidades y países, y genera consecuencias duraderas para los allegados de la víctima (WHO, 2021). De acuerdo con la Organización Mundial de la Salud, alrededor de 700,000 personas mueren por suicidio cada año. Este fenómeno no discrimina entre sectores de la población, afectando a todos los estratos sociales; sin embargo, se observa una incidencia particular en los países con ingresos bajos, como lo demuestra el dato de que en 2019 más del 77% de los suicidios globales ocurrieron en países de ingresos bajos y medianos.

Existen factores que pueden contribuir al suicidio, tales como la depresión o trastornos de salud mental, el consumo de alcohol y otras sustancias psicoactivas, antecedentes familiares, violencia intrafamiliar, posesión de armas de fuego, o la vivencia de eventos estresantes o traumáticos. (2022)

Este comportamiento puede ser provocado por una diversidad de factores tales como problemas de salud, económicos, sociales, personales o familiares. El suicidio es prevenible, siempre y cuando se actúe de manera oportuna y exista un apoyo constante por parte del entorno cercano de la persona con tendencias suicidas. Lamentablemente, persisten tabúes en torno a los temas de salud mental y suicidio, lo que dificulta que las personas busquen ayuda. Este fenómeno es particularmente notable en los hombres, quienes, según las estadísticas, son menos propensos a solicitar apoyo y presentan una mayor prevalencia de trastornos relacionados con el consumo de alcohol y otras sustancias, lo que puede contribuir a conductas suicidas. (Moutier, 2021).

El impacto de este problema no se reduce únicamente a las tasas de suicidio consumado, sino también a la ideación suicida o a las tentativas de suicidio previas. La ideación suicida y los intentos previos son precedentes frecuentes del suicidio consumado.

Cabe señalar, que en los últimos años la ideación e intento suicida ha incrementado. Los estudios realizados en México principalmente en población adolescentes han demostrado un incremento importante principalmente en mujeres (Rivera–Rivera, Fonseca–Pedrero, Sérís–Martínez, Vázquez–Salas, & Reynales–Shigematsu, 2020). De acuerdo con los resultados de la Encuesta Nacional de Salud y Nutrición (ENSANUT) en el año 2018 se encontró que el

6.6% de las mujeres y el 3.6% de los hombres presentaron ideación suicida y para el 2022 este incremento fue preocupante ya que el 10.22% de las mujeres y el 5.1% de los hombres presentaron este problema (Valdez–Santiago, Villalobos Hernández, Arenas–Monreal, Benjet, & Vázquez García, 2023).

Una de las formas más comunes de expresar pensamientos y emociones en la actualidad son las redes sociales. Dependiendo de su estructura, estas plataformas permiten a los usuarios mantener su identidad de manera anónima, lo que facilita una expresión más libre y natural de sus ideas (Ma, Hancock, & Naaman, 2016). En este sentido, el análisis de la información publicada en redes sociales puede ser útil para identificar a usuarios o usuarias con comportamientos relacionados a daños en la salud mental. Esto es especialmente relevante en el caso de las publicaciones de texto plano, que pueden ser analizadas mediante herramientas científicas, como la computación y la lingüística.

La lingüística, también conocida como la ciencia del lenguaje, se ocupa del estudio del lenguaje humano desde múltiples perspectivas (Coseriu, 1951). Gracias a los avances tecnológicos, hoy en día es posible realizar análisis automáticos del lenguaje, lo que ha dado lugar al campo de la lingüística computacional.

Comprender el suicidio, el papel de las redes sociales y el procesamiento del lenguaje natural es esencial para llevar a cabo investigaciones en esta área, ya que dichos conceptos son clave para analizar textos o publicaciones en redes sociales y detectar patrones que reflejen posibles conductas suicidas. Las redes sociales que permiten la descarga de publicaciones, como Twitter y Reddit, así como foros públicos y repositorios de datos accesibles, proporcionan fuentes valiosas para obtener datos sobre personas con tendencias suicidas.

1.1 Planteamiento del problema

Diversos estudios han explorado el uso de las redes sociales como un medio para detectar conductas suicidas. No obstante, la mayoría de los datos empleados en estos trabajos se basan en el inglés como idioma principal, siendo menos frecuente el uso del español. Estos trabajos tienen como objetivo principal el uso de algoritmos de clasificación automática, así como de modelos de transformación de texto a números para la detección de tendencias suicidas en redes sociales.

En los estudios previos se destaca la carencia de técnicas, algoritmos y herramientas que permitan un mejor aprovechamiento de textos breves y su conversión en datos numéricos, enfocándose en un par de técnicas como lo puede ser el utilizar una sola representación de textos.

Muchos trabajos se limitan a utilizar modelos clásicos de representación numérica, como TF-IDF, que tienen la desventaja de no contextualizar la información, ya que se restringen al vocabulario proporcionado por el usuario, sin considerar transformaciones más avanzadas como los *word embeddings*, que ofrecen una solución más efectiva, ya que permiten representar el contexto de una oración letra por letra, capturando el significado de una manera más completa y extrayendo más información, especialmente en textos cortos.

En este trabajo se evaluará la calidad de la información que los *word embeddings* pueden extraer de un texto y explorar la mejor combinación entre un *embedding* y un clasificador automático para detectar ideación suicida en redes sociales. Los resultados obtenidos serán comparados mediante métricas de clasificación, con el propósito de desarrollar una alternativa automatizada para la detección de posibles tendencias suicidas.

1.2 Objetivo

A continuación, se presenta el objetivo general del proyecto, así como cada uno de los objetivos específicos para completarlo.

1.2.1 Objetivo General

Detectar posibles tendencias suicidas en textos publicados en redes sociales mediante modelos de *embeddings* y clasificadores automáticos.

1.2.2 Objetivos Específicos

- Analizar trabajos relacionados con la detección automática de textos con tendencias o comportamientos suicidas.
- Desarrollar y entrenar una serie de modelos de inteligencia artificial que identifiquen publicaciones con tendencias suicidas utilizando el corpus de textos creado.
- Comparar los resultados obtenidos de las diferentes combinaciones de *embeddings* y clasificadores automáticos para la detección de tendencias suicidas en redes sociales.
- Evaluar los resultados de clasificación de la herramienta desarrollada.

1.3 Justificación

El suicidio es un problema global que afecta a millones de personas anualmente, y su prevención constituye una prioridad en el ámbito de la salud pública. Las redes sociales han emergido como un espacio donde las personas expresan sus pensamientos y emociones, incluidas aquellas relacionadas con la ideación suicida. La detección temprana de estas conductas en publicaciones de redes sociales puede permitir intervenciones oportunas y brindar apoyo a quienes lo necesitan, lo que podría salvar vidas.

La utilización de técnicas de procesamiento del lenguaje natural (PLN) y aprendizaje automático para la detección automática de tendencias suicidas ofrece una herramienta poderosa para identificar señales de advertencia tempranas. No obstante, la precisión y eficacia de estos modelos puede variar considerablemente, lo que subraya la necesidad de realizar investigaciones exhaustivas en este ámbito.

El objetivo de este estudio es comparar diferentes enfoques, como los *word embeddings* y los clasificadores, para determinar cuáles resultan más efectivos en la detección de ideación suicida en redes sociales. La relevancia de esta

investigación no solo radica en la mejora potencial de las herramientas para la prevención del suicidio, sino también en su contribución al campo del PLN, al abordar un problema de gran impacto social. Además, un enfoque exitoso podría ser implementado en sistemas de monitoreo y alerta, proporcionando a los profesionales de la salud mental una herramienta adicional para la prevención del suicidio.

1.4 Alcances y limitaciones

1.4.1 Alcances

- Se seleccionarán al menos 3 modelos de *embeddings* y clasificación para la detección automática de tendencias suicidas según su relevancia en la literatura.
- Se creará un *corpus* de texto suicida basándose en la recopilación de repositorios en internet cuyos datos sean utilizados en el estado del arte y sean relevantes para la detección automática.

1.4.2 Limitaciones

- El modelo no contemplará la identificación de modismos o regionalismos variante del idioma español.
- No se considerarán imágenes o videos de la información recopilada de internet para el análisis y entrenamiento de modelos de clasificación.
- El tamaño del corpus dependerá de las revisiones del estado del arte y de los datos públicos disponibles relacionados con el tema.

1.5 Organización del documento

La estructura de este documento consta de seis capítulos, organizados de la siguiente manera:

- **Fundamento teórico:** Se ofrece un análisis exhaustivo de la literatura relacionada con el tema, abordando el estado del arte sobre la detección de tendencias suicidas en redes sociales. Se revisan estudios previos y enfoques actuales que integran salud mental, análisis de redes sociales y procesamiento del lenguaje natural. Asimismo, se explican conceptos clave como el suicidio, los factores de riesgo asociados, y la relación entre las redes sociales y la ideación suicida, en el contexto del PLN.
- **Metodología de solución:** Se describe en detalle el enfoque metodológico adoptado para resolver el problema de investigación. Esto incluye el proceso de recolección de datos, especificando las plataformas de redes sociales seleccionadas y los criterios de inclusión de los textos. También se explican las técnicas de preprocesamiento del lenguaje (tokenización, lematización, filtrado de *stopwords*, entre otros) y los modelos de aprendizaje automático o profundo utilizados para clasificar las publicaciones en tendencias suicidas o no suicidas.
- **Experimentos y resultados:** Esta sección detalla los experimentos realizados para entrenar y evaluar los modelos propuestos. Se especifican los conjuntos de datos empleados, los parámetros del modelo y las técnicas de validación cruzada. Los resultados se presentan mediante métricas cuantitativas, como la precisión, sensibilidad, exactitud y F1-score, proporcionando un análisis comparativo de los diferentes modelos y técnicas evaluadas.
- **Conclusiones:** Se sintetizan los principales hallazgos de la investigación, evaluando si los objetivos planteados fueron alcanzados. Además, se discuten las implicaciones de los resultados obtenidos y se reflexiona sobre las limitaciones del estudio, como los posibles sesgos en los datos o en el modelo. Finalmente, se proponen futuras líneas de investigación orientadas a mejorar la precisión y eficacia en la detección automática de señales suicidas.
- **Producción científica:** Se documentan los productos científicos derivados del desarrollo de la tesis, tales como publicaciones en revistas académicas, presentaciones en congresos, participación en conferencias o talleres especializados, y cualquier otra forma de difusión científica del trabajo realizado.

2 Fundamento Teórico

2.1 Marco Conceptual

2.1.1 Suicidio

La etimología de la palabra "suicidio" se remonta al latín, donde "sui" significa "sí mismo" y "cidium" se traduce como "matar". Esto puede interpretarse como un acto de atentar contra la propia vida o de quitarse la vida. No obstante, su definición varía considerablemente según la cultura o el autor. En la antigua Grecia, el término se expresaba como "autokeiria", derivado de "autos" (sí mismo) y "keiros" (mano), lo que se puede interpretar como una muerte elegida por el individuo o una muerte llevada a cabo por su propia mano (Barrionuevo, 2009). Desde la perspectiva de la sociología, Emilio Durkheim define el suicidio de la siguiente manera: "Se llama suicidio a todo caso de muerte que resulta directa o indirectamente de un acto positivo o negativo, cumplido por la víctima misma, que sabía que debía producir ese resultado" (Durkheim, 1989)

El suicidio, así como la ideación e intento suicida, no se considera una enfermedad mental; sin embargo, puede ser un síntoma asociado a diversos trastornos mentales, según lo indica la Organización Mundial de la Salud (OMS). La OMS también señala que, en países de altos ingresos, existe una relación bien documentada entre el suicidio y las enfermedades mentales. De acuerdo con esta información, la mayoría de los casos de suicidio ocurren en momentos de crisis, cuando las personas afectadas carecen de la capacidad para afrontar sus problemas cotidianos. Entre estos problemas se incluyen dificultades económicas, rupturas de relaciones y padecimientos crónicos o terminales (WHO, 2021).

Asimismo, se ha registrado que la experiencia de eventos traumáticos puede dar lugar a conductas suicidas. Algunos de estos eventos comprenden la inestabilidad familiar, la violencia, los abusos y la pérdida de seres queridos. Un dato adicional es que la tasa de suicidio es particularmente alta entre grupos que presentan marcadas diferencias en la sociedad, los cuales tienden a ser vulnerables y a experimentar discriminación. No obstante, de todos los factores de riesgo identificados, el mayor riesgo de desarrollar tendencias suicidas se observa en aquellas personas que han tenido un intento previo de suicidio (WHO, 2021).

Un intento de suicidio, como su denominación sugiere, se refiere a la acción de una persona que ha intentado quitarse la vida sin lograr completar el

proceso. Este fenómeno puede atribuirse a diversos factores, entre los cuales se incluyen el fracaso del método empleado para causar la muerte, el temor a morir que puede impedir que la persona continúe con su intento, o incluso un cambio en la mentalidad del individuo en momentos críticos.

Adicionalmente, un suicidio puede transformarse en un intento suicida cuando el proceso es interrumpido por un tercero. Generalmente, esta interrupción ocurre cuando la persona que busca acabar con su vida solicita la ayuda o el consejo de otra persona, lo que puede modificar el curso de su acción.

Según la OMS es difícil conseguir datos precisos sobre el suicidio ya que es insuficiente la información disponible de estos, fundamentado en que únicamente 80 miembros de la OMS disponen de datos que se pueden utilizar para estimar tasas de suicidio. Otro de las razones es la notificación insuficiente o la clasificación errónea de casos de suicidio, ya que estos en ciertos países son considerados temas tabúes o ilegales hasta cierto punto. Cabe destacar que los casos de corrupción también pueden afectar la medición de casos suicidas en un país.

2.1.2 Factores de riesgo suicida

Según Moutier (2021) las autopsias psicológicas llevadas a cabo en personas que fallecieron por suicidio evidencian que los individuos afectados presentaban múltiples factores de riesgo asociados a este comportamiento. Las investigaciones indican que entre el 85% y el 95% de las personas que se quitan la vida padecen algún trastorno de salud mental. Entre los factores de riesgo más prevalentes se encuentra la depresión, la cual tiende a intensificarse cuando se combinan con otros factores de riesgo.

Otros factores de riesgo para el suicidio pueden ser:

- Trastornos de salud mental graves
- Intentos previos de suicidio
- Trastornos de personalidad
- Impulsividad y agresión
- Experiencias infantiles traumáticas
- Antecedentes familiares de suicidio
- Enfermedades psiquiátricas
- Uso de alcohol, drogas ilegales y analgésicos recetados
- Enfermedades físicas graves o crónicas
- Periodos de duelo
- Conflictos en relaciones

- Interrupción de trabajo
- Periodos de transición profesional
- Estrés financiero
- Acoso escolar

Entre las enfermedades asociadas al riesgo de suicidio, la esquizofrenia se destaca como un factor que incrementa la frecuencia de estos actos. Asimismo, los trastornos de personalidad también predisponen a los individuos a un mayor riesgo suicida, especialmente aquellos que padecen trastorno de personalidad limítrofe o antisocial.

Según el estudio de Sierra (2022), desde una perspectiva léxica, ciertos factores de riesgo presentes en la escritura pueden ser utilizados para identificar textos que reflejan tendencias suicidas. Entre estos factores se incluyen el uso recurrente de la primera persona, la utilización de verbos en primera persona, la manifestación de emociones negativas, así como expresiones de ansiedad, enojo, tristeza, muerte y elementos relacionados con el hogar.

2.1.3 Proceso suicida

En su trabajo "Adolescencia y suicidio," Baron (2000) hace referencia a lo que plantea G. Bouchard en su artículo "Le Suicide a l'adolescence," publicado en el año 2000. Según Bouchard, el proceso suicida se comprende como el periodo que transcurre desde el momento en que se produce la crisis hasta la ejecución del acto suicida. Se destaca que, en el caso de los adolescentes, este proceso puede ser considerablemente más rápido, pudiendo completarse en cuestión de horas, especialmente si ya ha habido intentos previos de llevar a cabo un acto suicida.

Este proceso se plantea en 5 momentos:

- **Búsqueda de soluciones ante una crisis.** Algunas personas pueden encontrar diversas soluciones en algún problema, mientras que otras no son capaces de encontrar soluciones que respondan a sus necesidades. En este punto el suicidio no es una respuesta o posible solución.
- **Ideación suicida.** Durante esta fase solo es necesario una pequeña introducción al tema de la muerte: una imagen, video, alusión, o una referencia. En este punto la idea de la muerte comienza a parecer una posible solución, apareciendo con mayor frecuencia e intensificándose con la construcción de escenarios imaginarios.

- **Rumiación.** En esta etapa se experimenta una gran inconformidad debida la situación anterior. La incapacidad de resolver la crisis y la sensación de no tener opciones genera más angustia, lo que genera un aumento de los pensamientos suicidas. Estos pensamientos se vuelven constantes y recurrentes, incrementando la angustia, el sufrimiento y el dolor, creando un ciclo repetitivo.
- **Cristalización.** Llegando a este punto la persona se encuentra sin esperanzas de encontrar una solución. El suicidio se percibe como la solución definitiva a todos sus problemas, lo que lleva a la planificación del acto suicida. Se experimenta un falso sentido de alivio y mejoría temporal, ya que se cree que existe una solución definitiva. Frecuentemente la persona comienza a romper vínculos con su entorno y a aislarse, creando un sentimiento de soledad. En este punto, solo hace falta un evento detonante para llevar a cabo el acto suicida.
- **Elemento desencadenante.** En este punto, el paso al acto se vuelve inminente. Solo es necesario un evento, relevante o no, que la persona considere una derrota más en su vida.

Es importante destacar que, en cualquier etapa de este proceso, una intervención adecuada puede ayudar a interrumpir estos pensamientos.

2.1.4 Programa nacional de prevención del suicidio

El gobierno mexicano ha implementado un programa de prevención del suicidio que abarca diversos objetivos y estrategias específicas. El objetivo general de este programa es establecer mecanismos de intervención para atender a las personas que buscan ayuda debido a comportamientos suicidas, así como reducir la mortalidad asociada al suicidio (Secretaría de Salud, 2022).

Los objetivos específicos del programa son:

- Mejorar la vigilancia epidemiológica sobre el comportamiento suicida y muertes por suicidio a nivel nacional.
- Promover acciones homologadas de evaluación e intervención sobre el comportamiento suicida y muertes por suicidio, que permitan dirigir líneas de investigación que den respuesta a las necesidades de la población mexicana.
- Difundir entre la población general información veraz sobre los determinantes bio-psico-sociales (factores de riesgo y factores de protección) asociados al suicidio y las enfermedades mentales.

- Generar convenios interinstitucionales que garanticen la referencia oportuna para una adecuada atención y tratamiento a las familias de personas con comportamiento suicida.
- Capacitar y formar a los equipos de salud para el manejo y tratamiento del comportamiento suicida con base en el Modelo Dinámico de Ordenes de Riesgo de Suicidio (MODORIS).
- Establecer los mecanismos de vinculación interinstitucional con los diferentes actores clave, tanto del gobierno como de la sociedad civil para activar las redes de atención en salud mental en los casos de comportamiento suicida y muertes por suicidio.

De forma específica se mencionan cinco estrategias alineadas con los objetivos específicos las cuales permitirían construir una política clara sobre cómo las 32 entidades federativas deben atender el problema del suicidio (Secretaría de Salud, 2022)

- Coordinación intersectorial e interinstitucional: Consiste en establecer los mecanismos de vinculación con los diferentes actores clave, tanto del gobierno como de la sociedad civil para activar la colaboración en las redes de atención en salud mental.
- Vigilancia epidemiológica: Contar con un sistema homologado de recopilación de datos sobre comportamiento suicida y muertes por suicidio, para tener un monitoreo permanente de los casos de autolesiones, ideación suicida, intentos de suicidio y suicidios, así como los factores de riesgo asociados en cada entidad del país a través de la inclusión de la vigilancia del comportamiento suicida en los sistemas de información de DGIS y DGE.
- Intervención: Consiste en implementar a nivel intersectorial acciones de prevención, atención y seguimiento del comportamiento suicida para reducir la mortalidad por esta causa, mediante estrategias clínicas y comunitarias basadas en evidencia. Dirigida
- Capacitación: consiste en diseñar e implementar programas de capacitación y actualización dirigidos al personal de salud con el propósito de potenciar sus habilidades clínicas, para responder a las necesidades psicoemocionales de la persona con comportamiento suicida, según los cuatro ordenes de riesgo del MODORIS.
- Investigación: Con la finalidad de generar evidencia en torno a las políticas e intervenciones más efectivas y eficaces en materia de prevención del suicidio, se promoverán mayores esfuerzos de reflexión,

recopilación, procesamiento y análisis de información epidemiológica, factores de riesgo y protección individuales y psicosociales, asociados al estudio de los determinantes sociales de la salud (DSS). Ello permitirá fortalecer estrategias universales, así también focalizar intervenciones para grupos específicos con mayor riesgo por conducta suicida y/o localidades de creciente incidencia.

2.1.5 Redes sociales

Las redes sociales se pueden definir como servicios en línea basados en la web que permiten a los usuarios relacionarse, compartir información, coordinar acciones y en general, mantenerse en contacto (Orihuela, 2008). Estos espacios virtuales, en los que nos relacionamos, nos permiten construir nuestra identidad la cual puede ser observada por el mundo entero.

Desde la creación de internet en 1972, ha habido numerosos avances en la interacción con las computadoras. Este proceso culminó con la creación de lo que conocemos como la *World Wide Web* (WWW) en 1990. A partir de ese momento, la forma en que los humanos nos conectaríamos cambiaría para siempre. Hoy en día las redes sociales más influyentes son Facebook, Twitter e Instagram las cuales ofrecen una mayor libertad de expresión en cuanto al contenido que se desea publicar. Además, permiten establecer filtros que muestran solo lo que más nos interesa (Orihuela, 2008). Estos filtros son personalizados según la interacción del usuario con las plataformas, e incluso recomiendan contenido basado en las interacciones de amigos y contactos.

Lo anterior, permite reflexionar que vivimos en un mundo extremadamente conectado, donde una simple acción como tomar el teléfono puede brindar el acceso a una red de información tan basta como la biblioteca de Alejandría. Por otro lado, también permite compartir información de cualquier índole a familia, amigos o conocidos en menos de 5 minutos. Las redes sociales se han convertido así en una fuente invaluable de información personal y cotidiana que no se puede obtener de ningún otro lugar.

2.1.6 Procesamiento de lenguaje natural (PLN)

El procesamiento del lenguaje natural (PLN) se refiere a la utilización del lenguaje natural para la comunicación con computadoras, de manera que estas puedan comprender las oraciones que se les presentan. Este enfoque busca facilitar el desarrollo de programas que realicen tareas relacionadas con el

lenguaje o crear modelos que ayuden a comprender los mecanismos humanos asociados al lenguaje (Augusto Cortez Vásquez, Hugo Vega Huerta, Jaime, & Quispe, 2009). Existen diversas aplicaciones del PLN, tales como la traducción automática entre diferentes idiomas, la evaluación de la sensibilidad de la información, la extracción y generación de resúmenes, la corrección ortográfica, y el reconocimiento de voz, entre otras.

Las lenguas están regidas por normas de comportamiento, conocidas como lenguaje, y se organizan en distintos niveles: fonético/fonológico, morfológico, sintáctico, semántico y pragmático. (Castro Sánchez, 2021). En el contexto del procesamiento del lenguaje natural (PLN), estos cinco niveles son utilizados para abordar problemas lingüísticos específicos mediante sistemas informáticos. Es fundamental que cualquier texto destinado a su análisis sea procesado adecuadamente, dado que la computadora no puede comprender directamente el significado de las palabras.

2.1.7 Enfoques para el análisis de texto

Existen varios enfoques para abordar problemas de interés relacionados con la lingüística computacional. Algunos ejemplos de estos enfoques incluyen aquellos basados en reglas, los estadísticos tradicionales, los conexionistas y los híbridos. (Castro Sánchez, 2021)

El enfoque basado en reglas, también denominado heurístico o basado en conocimiento, modela un fenómeno a través de reglas que reflejan las condiciones de uso o la aparición de unidades lingüísticas. Estas reglas pueden implementarse utilizando expresiones regulares, representadas como gramáticas libres de contexto, o mediante estructuras de control, como las condiciones if-else comúnmente empleadas en lenguajes de programación. Para la aplicación efectiva de este método, es fundamental poseer un conocimiento profundo del problema en cuestión, lo que permite generar las reglas adecuadas.

El enfoque basado en aprendizaje automático abarca tanto los enfoques estadísticos tradicionales como los conexionistas. Cada uno de estos enfoques gestiona los datos de manera diferente: el enfoque estadístico tradicional se basa en la información presente en los datos y, mediante estadísticas y cálculos concretos, resuelve problemas. Por otro lado, los enfoques conexionistas, como las redes neuronales, llevan a cabo cálculos que les permiten aprender de la información proporcionada, adquiriendo así el conocimiento necesario para realizar tareas y encontrar soluciones. La principal

diferencia entre estos métodos radica en la manera en que manejan la información; mientras que el enfoque estadístico trabaja con los datos tal como se presentan, el enfoque conexionista aprende y se adapta según la información disponible.

El principal desafío en el procesamiento de textos radica en que las palabras no tienen una representación natural o directa en valores numéricos. Para poder trabajar con textos en el ámbito del aprendizaje automático, se han desarrollado diversas técnicas que ofrecen soluciones según las necesidades específicas. Comúnmente, los textos se representan como vectores numéricos en un modelo conocido como espacio vectorial. En este modelo, cada término del texto se convierte en un índice o valor numérico que se integra en el vector correspondiente.

2.1.8 Traducción de textos

El concepto de equivalencia es uno de los más utilizados en la traducción entre idiomas. En términos generales, se considera que una unidad traducida ha alcanzado un nivel de equivalencia con respecto al texto original cuando lo traducido conserva el mismo significado que en la lengua de origen (Márquez, 2008). Al llevar a cabo una traducción a otro idioma, se discuten dos enfoques principales: la equivalencia formal, que busca una aproximación al texto original, y la equivalencia dinámica, que intenta adaptarse más al lector y a la cultura del idioma objetivo.

La traducción por equivalencia formal tiene como objetivo principal preservar la forma lingüística del texto original en la lengua de destino. Este enfoque busca imitar la sucesión de palabras, la sintaxis y, en la medida de lo posible, los sonidos del texto original. Este tipo de traducción es comúnmente conocido como traducción literal.

Por otro lado, la traducción dinámica no busca replicar la lengua de origen, sino transmitir el mensaje utilizando los recursos propios del idioma de destino, conservando los aspectos pragmáticos. Este tipo de traducción debe adaptarse a todo el contexto proporcionado por el idioma original, haciéndolo de la manera más natural posible, evitando forzar la expresión y asegurando que sea adecuada para la comunidad lingüística de destino.

2.1.9 Algoritmos de clasificación

Los algoritmos de clasificación desempeñan un papel crucial en el procesamiento de lenguaje natural (PLN) al automatizar la asignación de etiquetas o categorías a textos. Estos algoritmos son esenciales en una amplia variedad de tareas, tales como el análisis de sentimientos, la clasificación de correos electrónicos y la categorización de noticias. En este contexto, los *word embeddings* han revolucionado el rendimiento de los algoritmos de clasificación. Representaciones como *Word2Vec*, *GloVe* o *BERT* modelan palabras como vectores densos en un espacio continuo, capturando tanto las características léxicas de las palabras como sus relaciones semánticas. Esto proporciona a los algoritmos de clasificación una representación numérica más rica, que refleja de manera más precisa el significado y el contexto de las palabras, lo que se traduce en una mayor precisión en tareas de PLN.

Los algoritmos de clasificación tradicionales, como la Regresión Logística (LR), las Máquinas de Soporte Vectorial (SVM), los Árboles de Decisión (DT) y los Bosques Aleatorios, tienden a utilizar técnicas de preprocesamiento como *TF-IDF* o *Bag of Words* para convertir los textos en vectores dispersos y de alta dimensionalidad, donde cada palabra o término se trata como una característica independiente. Sin embargo, estas técnicas no son eficaces para capturar las relaciones semánticas entre palabras, lo que limita su capacidad para modelar el contexto lingüístico. Aunque estos métodos tradicionales son rápidos y relativamente fáciles de interpretar, su desempeño en tareas complejas de PLN es inferior debido a su incapacidad para capturar el significado contextual de las palabras.

En contraste, los algoritmos basados en redes neuronales, como las Redes Neuronales Convolucionales (CNN), las Redes Neuronales Recurrentes (RNN) y los modelos basados en *Transformers* (como *BERT* o *GPT*), pueden aprovechar de manera mucho más efectiva los *word embeddings*. Estas redes procesan las representaciones vectoriales de las palabras en conjunto, lo que les permite captar tanto la secuencia como el contexto global de los textos, extrayendo patrones más complejos y sutiles que los modelos tradicionales no logran detectar. Además, los modelos basados en redes neuronales pueden ser preentrenados en grandes cantidades de datos no supervisados, lo que facilita la transferencia de conocimiento a tareas de clasificación específicas.

Aunque es posible utilizar *word embeddings* con modelos de clasificación tradicionales, la complejidad inherente de las representaciones vectoriales puede dificultar que estos clasificadores capturen correctamente las relaciones semánticas del texto, lo que resulta en una menor eficiencia en el rendimiento de la clasificación. Por ello, los algoritmos basados en redes neuronales, que

están diseñados para manejar tales complejidades, suelen ser más adecuados para aprovechar el potencial completo de los *embeddings* en tareas avanzadas de PLN.

Las ventajas y desventajas de los distintos algoritmos de clasificación se pueden resumir de la siguiente manera:

Algoritmos tradicionales

- Ventajas:
 - Son rápidos de entrenar y requieren menos recursos computacionales, lo que los hace ideales para tareas con limitaciones de tiempo o hardware.
 - Son fáciles de interpretar y depurar, lo que facilita entender cómo las características individuales influyen en la clasificación.
 - Funcionan bien con conjuntos de datos pequeños o cuando se dispone de representaciones simples y de alta calidad del texto, como cuando se emplean técnicas como *TF-IDF* o *Bag of Words*.
- Desventajas:
 - Tienen una capacidad limitada para modelar relaciones semánticas entre palabras, ya que no capturan el contexto en el que aparecen los términos.
 - Su rendimiento suele ser inferior al de los modelos basados en redes neuronales, especialmente en tareas más complejas o con grandes volúmenes de datos.
 - Las representaciones dispersas de características, como *Bag of Words* o *TF-IDF*, pueden llevar a la creación de un "espacio de características" muy grande, lo que resulta costoso en términos computacionales.

Algoritmos basados en redes neuronales

- Ventajas:
 - Capturan de manera mucho más eficaz el contexto y las relaciones semánticas entre palabras, gracias al uso de *word embeddings* y su capacidad para procesar secuencias de palabras.
 - Tienen una mejor capacidad de generalización, especialmente cuando se entrenan con grandes cantidades de datos, lo que los hace más adecuados para tareas complejas y variadas.
 - Los modelos preentrenados, como *BERT* o *GPT*, pueden ser reutilizados en tareas específicas mediante el ajuste fino (*fine-*

tuning), lo que ahorra tiempo de entrenamiento y mejora el rendimiento.

- Desventajas:
 - Son mucho más costosos desde el punto de vista computacional, tanto en el entrenamiento como en la inferencia, debido a la cantidad de parámetros que manejan y su mayor complejidad.
 - Requieren grandes cantidades de datos etiquetados para entrenarse desde cero, aunque los modelos preentrenados ayudan a reducir este problema.
 - Pueden ser difíciles de interpretar, lo que complica la comprensión de por qué el modelo toma ciertas decisiones o la identificación de errores en su funcionamiento.

2.2 Estado del arte

Se presentan 17 artículos relevantes en el estado del arte, los cuales están ligados al núcleo temático: “detección de tendencias suicidas en redes sociales por medio de procesamiento de lenguaje natural” Para la construcción del trabajo se consideraron documentos en los idiomas inglés y español buscados en MedlinePlus, PubMed y Google Scholar. Solamente se consideraron los artículos que tuvieran máximo 5 años de ser publicados.

Exploring Behavior of People with Suicidal Ideation in a Chinese Online Suicidal Community (Wang, Yu, & Tian, 2019)

El propósito de este artículo fue entender el comportamiento de las personas con ideación suicida en comunidades web particularmente en la red social Weibo en China. El artículo se basó en la interacción en la red de un usuario llamado “Zoufan”, el cual dejó un mensaje de despedida programado automáticamente para publicarse después de su suicidio. Durante los últimos 6 años se han publicado 560,000 comentarios dejados por 160,000 usuarios de los cuales se detectó que 3039 eran usuarios con ideación suicida. Como resultado se encontraron un total de 4489 comentarios suicidas: 3300 comentarios (73.5%) expresaron ideaciones suicidas y 513 (11.4%) expreso tanatofobia. Algunos temas menos frecuentes fueron: búsqueda de compañeros suicidas con 304 comentarios (6.8%), preguntas sobre métodos suicidas con 263 comentarios (5.9%) y descripciones de intentos suicidas previos con 117 comentarios (2.6%). Se concluyó que la interacción de personas suicidas motiva a otras personas con problemas similares a ir por el mismo camino, como si de una infección se tratara.

Exploring the Risk of Suicide in Real Time on Spanish Twitter: Observational Study (García-Martínez et al., 2022)

Para este trabajo se recolectaron y analizaron tweets por medio de un marco de trabajo desarrollado en un proyecto anterior por el autor. Los datos fueron extraídos de Twitter los cuales solo considera a los relacionados con el suicidio. Cada uno de los tweets fue asignado a uno de los 25 psicólogos con experiencia con el objetivo de evaluar los tweets para ser parte de las variables de estudio. Posteriormente los tweets fueron asignado a 3 evaluadores los cuales de manera independiente realizaron una revisión sobre los textos ya evaluados. Como resultado 2059 tweets fueron obtenidos. De los cuales 2018 no fueron relevantes por los 3 evaluadores, mientras que 257 fueron relevantes para 2 de 3 evaluadores. Finalmente 216 tweets fueron encontrados como relevantes y entre estos 68 fueron los tweets más relevantes por los 3

evaluadores. Los resultados de este estudio permiten identificar usuarios en riesgo suicida mediante el uso de tweets en español, además que no se necesita solamente un lexicón relacionado al suicidio para funcionar, sino que también es posible utilizar el contenido emocional de los textos.

Detection of Suicidality Among Opioid Users on Reddit: Machine Learning-Based Approach (Yao et al., 2020)

La investigación realizada tuvo como objetivo combinar datos de algunos subreddits y el usar machine learning para predecir y detectar ideaciones suicidas en el contexto de consumo de opioides. La mayor parte de la información fue obtenida de los subreddits *r/suicidewatch* y *r/depression* así como foros relacionados a los opioides como *r/opiates* o *r/heroin*. La información recolectada fue procesada usando técnicas de procesamiento de lenguaje natural para su posterior uso con métodos de clasificación. El método principal para la clasificación fue un modelo de redes neuronales convolucionales (CNN). Otros métodos utilizados para la comparación de eficacia fueron regresión logística (LR), bosques aleatorios (RF) o máquinas de soporte vectorial (SVM). Como resultado dependiendo de la técnica de preprocesamiento de datos utilizada puede ayudar a mejorar el rendimiento de los modelos mencionados. La configuración con mejor rendimiento fue el uso de *Term Frequency – Inverse Document Frequency* (TF-IDF) y *word2vec* obteniendo una puntuación F1 de 96% con la CNN, mientras que con los métodos clásicos SVM obtuvo un resultado de 94.1% seguido de LR con un 94% y finalmente RF con un 92%.

Detecting Potentially Harmful and Protective Suicide-Related Content on Twitter: Machine Learning Approach (Metzler, Baginski, Niederkrotenthaler, & Garcia, 2022)

El estudio implementa procesamiento de lenguaje natural y métodos de aprendizaje máquina para clasificar largas cantidades de datos de redes sociales de acuerdo con las características identificadas como potencialmente dañinas o benéficas en redes sociales con respecto al suicidio. Utilizando Twitter como su red social de estudio y el idioma inglés como predefinido en los datos a analizar se recolectaron y etiquetaron 3202 *tweets* que contenían al menos uno de los términos suicidas en los trabajos investigados previos a este documento (por ejemplo: *suicide*, *suicidal*, *kill himself*, *hang himself*, etc...). Para este trabajo se utilizó un modelo de SVM con TF-IDF como herramienta de transformación de texto, además se utilizaron paradigmas de aprendizaje profundo como lo son BERT Base y XLNet Base.

Se tuvo como resultado que de un conjunto de prueba de 641 tweets que la SVM con TF-IDF alcanzo un 66% de exactitud mientras que XLNet obtuvo un 74% seguido de BERT base con un 73%, concluyendo que los métodos de aprendizaje profundo tienen un mejor desempeño que los modelos clásicos.

Deep Hierarchical Ensemble Model for Suicide Detection on Imbalanced Social Media Data (Li, Zhou, An, Cheng, & Hu, 2022)

Desarrollado en el idioma chino y utilizando la red social Weibo se extrajeron 2019 usuarios con ideación suicida de los cuales se obtuvieron 164,854 publicaciones originales en los rangos de fechas 1 de julio de 2020 a 30 junio de 2021. También se obtuvieron 489,543 publicaciones de 2997 usuarios comunes, sin relación con temas de suicidio o depresión. El modelo propuesto “Deep Hierarchical Ensemble – Suicide Detection” (DHE-SD) pretende solucionar el problema de los datos no balanceados, en el cual se expresa que la saturación de datos en una sola categoría puede afectar de manera negativa a las predicciones que realice un clasificador. Con este modelo propuesto se utilizaron también redes neuronales profundas (TextCNN, FastText, DPCNN) para su evaluación y respectiva comparación. Finalmente, como resultados se utilizaron solo dos métricas de evaluación: exactitud y puntuación F1. Los métodos de aprendizaje profundo fueron los primeros en evaluarse, dando como resultado que TextCNN tuvo un mejor desempeño con un 93% de exactitud y un 91% de puntuación F1. Posteriormente y con los resultados anteriores se utilizaron como base de comparación para el método propuesto, el cual tuvo un mejor desempeño, teniendo este un 95% de exactitud y una puntuación F1 de 93%.

Exploring temporal suicidal behavior patterns on social media: Insight from Twitter analytics (Luo, Du, Tao, Xu, & Zhang, 2020)

El objetivo de este trabajo fue examinar los posibles patrones temporales de potenciales ideaciones y comportamientos suicidas en Twitter para la mejor comprensión los factores de riesgo de las personas suicidas. Se obtuvieron 716,899 tweets públicos de los rangos de fecha enero a noviembre de 2016 usando términos relacionados con el suicidio, todo esto por medio de la API de Twitter. Se utilizó el paradigma CNN para la clasificación de oraciones. Otros modelos utilizados para la evaluación del dataset fueron SVM, RF, LR, Extra Trees, Bi-dimensional Long Short-Term Memory. Como resultados se encontró que el paradigma CNN obtuvo una Precisión de 78%, una Sensibilidad de 88% y una Puntuación F1 de 83%. Posteriormente el modelo ya entrenado fue utilizado para clasificar un corpus de prueba de 191,473 tweets que fue usado posteriormente en la investigación.

Los resultados del documento indican que se encontraron 13 factores de riesgo, entre los cuales se encuentran: la vida, pérdida de energía, depresión, moda, emoción, cambio, entre otros.

Detection of traits in students with suicidal tendencies on internet applying web mining (Castillo-Zúñiga, Luna-Rosas, & López-Veyna, 2022)

Este trabajo presenta un modelo de análisis de datos en internet basado en minería web con el objetivo de encontrar conocimiento sobre grandes cantidades de datos en el ciberespacio enfocado en la identificación y detección de estudiantes con tendencias suicidas. El idioma trabajado es el español y por medio de un web *scraper* de código abierto, junto con algunas técnicas de *scraping* se obtuvo un corpus el cual contiene los nombres de páginas web obtenidas, dirección URL de estas y copias en texto plano del contenido HTML de la página web. Para los métodos de clasificación se utilizaron bosques aleatorios, redes neuronales, arboles de decisión y regresión logística. Finalmente se utilizó también un paradigma de aprendizaje profundo: redes neuronales recurrentes (RNN). Como resultado de los métodos tradicionales el que mejores resultados tuvo fue LR con un 97.41% de exactitud. Por otro lado, en los métodos ejecutados con computo paralelo el árbol de decisión fue el que mejor desempeño tuvo, con un 95.11% de exactitud y finalmente, con el aprendizaje profundo el paradigma RNN tuvo un favorable resultado de 98% de exactitud.

Assessment of supervised classifiers for the task of detecting messages with suicidal ideation (Acuña Caicedo, Gómez Soriano, & Melgar Sasieta, 2020)

Este estudio realizó un análisis sistemático de 28 métodos supervisados usando diferentes características de un corpus de la vida con el objetivo de detectar mensajes con ideación suicida y depresión. Este trabajo tiene una combinación de idiomas, utilizando información en inglés y en español, la cual no está recopilada de ninguna red social o foro público. El corpus de la vida es una serie de textos recopilados por el grupo de investigación de la universidad de Alicante. Este corpus consiste en 102 mensajes de texto, de los cuales 71 son en inglés y 31 son en español, todos siendo mensajes de índole suicida. Para los clasificadores se utilizó la paquetería Weka, una colección de métodos para tareas de minería de datos la cual contiene los 28 clasificadores a implementar.

Los resultados se obtuvieron después de realizar 3472 experimentos con diferentes combinaciones en las características utilizadas del corpus de la vida. Finalmente, el método Kstar fue el más sobresaliente de todos los demás

evaluados, obteniendo un Puntuación F1 de 76% y una ROC de 81%. El punto decisivo de este trabajo fue el preprocesamiento de datos, en los cuales siguiendo ciertas combinaciones se pueden obtener mejores resultados.

Detection of suicide ideation in social media forums using deep learning (Tadesse, Lin, Xu, & Yang, 2020)

Este estudio tuvo como objetivo presentar el trabajo de reconocimiento de publicaciones suicidas de manera automática. El corpus extraído de Reddit fue creado en el trabajo de (Ji, Yu, Fung, Pan, & Long, 2018) llamado “Supervised learning for suicidal detection in online user content”. Este corpus consiste en 3549 publicaciones suicidas y 3652 publicaciones no suicidas. Este trabajo tuvo como clasificador elegido el Long Short Term Memory – Convolutional Neural Network (LSTM–CNN) los cuales son paradigmas de aprendizaje profundo. Esta combinación fue comparada con otros métodos de clasificación tradicionales, así como con paradigmas de aprendizaje profundo con el fin de probar la eficacia del LSTM–CNN.

Para la evaluación de este proyecto se utilizaron las métricas de exactitud, precisión, Sensibilidad y Puntuación F1. Es importante mencionar que el procesamiento de datos previos jugó un papel importante pues la eficiencia del clasificador puede mejorar con una selección de características útiles. El resultado más óptimo lo tuvo LSTM–ID con un 93% de exactitud, Puntuación F1 y precisión y un 94% en Sensibilidad.

An ensemble deep learning technique for detecting suicidal ideation from posts in social media platforms (Renjith, Abraham, Jyothi, Chandran, & Thomson, 2022)

Este artículo propuso una técnica de aprendizaje profundo para detectar ideación suicida a partir de publicaciones en plataformas de redes sociales, concretamente en Reddit. El modelo propuesto combinó modelos LSTM y CNN para analizar publicaciones en redes sociales y detectar intenciones suicidas. El conjunto de datos utilizado para el entrenamiento y las pruebas es el Reddit Suicidality Dataset de la Universidad de Maryland, que contiene publicaciones del subreddit SuicideWatch etiquetadas con cuatro categorías: sin riesgo, riesgo bajo, riesgo moderado y riesgo grave. El modelo propuesto alcanzó una precisión del 90,3% y una puntuación F1 del 92,6%, superior a la de los modelos de referencia. Los autores también realizaron un análisis de datos para identificar los n–gramas comunes en los mensajes con ideaciones suicidas y los compararon con los de los mensajes de bajo riesgo. Sugieren que su modelo puede ayudar a reducir la tasa de suicidios detectando

mensajes suicidas en las redes sociales y proporcionando la asistencia médica adecuada.

Detection of Suicidal Ideation on social media: Multimodal, Relational, and Behavioral Analysis (Ramírez–Cifuentes et al., 2020)

Este trabajo de investigación abordó la detección de tendencias suicidas en redes sociales, recopilando datos públicos de los usuarios, se tomó en cuenta datos como textos, imágenes y características propias de los usuarios, como las relaciones con otros usuarios y su comportamiento en la red social. Se evaluó la información de 252 usuarios clínicamente detectados con tendencias suicidas. Se caracterizó información basada en su escritura, patrones de publicaciones, interacciones con otros usuarios e imágenes y textos publicados. Para la clasificación de los datos se consideraron cuatro algoritmos: bosques aleatorios, perceptrón multicapa, regresión logística, máquinas de soporte vectorial y redes neuronales convolucionales. Los resultados en la clasificación se obtuvieron mediante la combinación de características, siendo los textos con imágenes y relaciones entre comportamientos e interacciones de los usuarios los más representativos y efectivos en las tareas de clasificación, obteniendo resultados de precisión, sensibilidad y puntuación F1 arriba del 80%, mientras que la exactitud y el área bajo la curva obtuvieron valores arriba de 90%. Los algoritmos de clasificación más utilizados fueron máquinas de soporte vectorial, perceptrón multicapa y regresión logística.

Detection of Suicidal Intent in Spanish Language Social Networks using Machine Learning (Valeriano, Condori–Larico, & Sulla–Torres, 2020)

Este artículo expresa la preocupante problemática del suicidio en la población y como las intervenciones pueden ayudar a la prevención del suicidio. Unos de los objetivos principales del artículo fue el uso de redes sociales para realizar la detección automática de ideación suicida en el idioma español utilizando procesamiento de lenguaje natural. Por medio de palabras clave asociadas con el riesgo suicida se extrajeron datos de la red social Twitter obtenido un total de 10000 publicaciones. Posteriormente se creó un subconjunto aleatorio de datos de 2068 publicaciones de las cuales mediante un procedimiento de anotado manual se clasificaron en dos etiquetas (suicida y no suicida) dando como resultado 498 tweets suicidas y 1570 no suicidas. Se utilizaron técnicas de representación de vectores, como lo es TF–IDF, y de Word Embeddings, como lo es Word2Vec. Los algoritmos utilizados en las tareas de clasificación fueron máquinas de soporte vectorial y regresión logística. Los resultados del experimento indicaron que las representaciones con Word Embeddings y el

algoritmo de regresión logística tuvieron un mejor desempeño con una exactitud, precisión, sensibilidad y puntuación F1 de 79%.

Suicidal ideation and mental disorder detection with attentive relation networks (Ji, Li, Huang, & Cambria, 2022)

La salud mental es un problema en la sociedad moderna y los medios digitales se involucran más en la vida cotidiana de las personas, siendo estos utilizados como medios de expresión de sentimientos. La temprana detección de padecimientos mentales, en este caso ideación suicida, puede ayudar a realizar una intervención social efectiva. Clasificar ideación suicida es un problema complejo puesto que puede maquillarse con algún desorden mental. Este artículo utilizó textos publicados en redes sociales, así como puntuaciones de sentimientos que, junto con redes relacionales, sea posible detectar ideación suicida en publicaciones. Para realizar este experimento se recopiló 3 conjuntos de datos, extraídos de Reddit y de Twitter. En los algoritmos de clasificación se utilizaron *fastText*, CNN, LSTM, RCNN, SSA y RN. Como resultados finales los diversos conjuntos de datos tuvieron un desempeño diferente, siendo el conjunto de datos de Twitter el más efectivo con un 83% de exactitud, precisión, sensibilidad y puntuación F1. Posteriormente el autor indico que la combinación de los valores generados de las LSTM más una red relacional y junto con las características de sentimientos y tópicos utilizados en las publicaciones los resultados mejoran ligeramente.

Detection of suicide-related posts in Twitter data streams (Vioules, Moulahi, Aze, & Bringay, 2018)

El trabajo presentó un nuevo enfoque para detectar publicaciones relacionadas con el suicidio en Twitter utilizando técnicas de procesamiento del lenguaje natural y el *framework* de Martingale para detectar cambios el comportamiento de una serie de publicaciones de un usuario. El enfoque incluyó la creación de una puntuación de textos (nivel de angustia) para cuantificar el nivel de riesgo de un individuo en función de su comportamiento en línea. El método consistió en crear un lexicón de síntomas (expresiones usadas por usuarios deprimidos o con ideas suicidas) y utilizar un clasificador automático para identificar tweets de personas con una clase de angustia. El *framework* de Martingale se utilizó para procesar estas características de comportamiento y detectar cambios emocionales en el flujo de datos. La metodología se probó con un conjunto de 5.446 tuits y una serie de datos de 10 usuarios. Los resultados mostraron una puntuación de precisión máxima del 66%, sensibilidad de 43% y puntuación F1 de 33% una primera etapa de clasificación multiclase utilizando diversos métodos de aprendizaje automáticos. Posteriormente se realizó una

clasificación binaria para determinar textos con o sin angustia, dando como mejor resultado una sensibilidad del 71% y posteriormente una segunda clasificación binaria para determinar el nivel de angustia en el mensaje, dando como resultado una sensibilidad de 82%.

Multi-task learning to detect suicide ideation and mental disorders among social media users. (Buddhitha & Inkpen, 2023)

Esta investigación utilizó tres corpus de datos para investigar la relación entre desórdenes mentales e ideación suicida; estos datos se encuentran “The University of Maryland Reddit Suicidality dataset” (UMD dataset), Self-Reported Mental Health Diagnoses dataset (SMHD dataset) y el Computational Linguistics and Clinical Psychology dataset (CLPsych dataset). Una red neuronal convolucional multicanal fue utilizado como el modelo principal. La arquitectura empleó múltiples tareas de aprendizaje compartiendo parámetros suaves y duros en la capa de concatenación. Para cada canal de la red neuronal se utilizaron diferentes procesos para determinar los desórdenes mentales o ideación suicida, por ejemplo, el uso de dos tamaños de kernel diferentes con diferentes N-Gramas. La investigación también identificó limitaciones como lo es la falta de datos y el no poder usar transformadores mencionados en el estado del arte debido a limitaciones de tamaño en las secuencias de entrada.

Proactive Suicide Prevention Online (PSPO): Machine Identification and Crisis Management for Chinese Social Media Users with Suicidal Thoughts and Behaviors (Liu et al., 2019)

Esta investigación usó aprendizaje automático para identificar individuos en riesgo, para lo cual se construyó un modelo de clasificación binario para determinar si un comentario indicaba tendencia suicida. El estudio uso datos públicos obtenidos de un microblog de un usuario llamado Zoufan, el cual murió de suicidio debido a una depresión. El microblog contiene más de 1.3 millones de comentarios, muchos de los cuales están relacionados con el suicidio. Se utilizó máquinas de soporte vectorial, arboles de decisión, bosques aleatorios y regresión logística como algoritmos de clasificación con una validación cruzada de 10 dobleces. Se demostró que incorporando conocimiento del dominio y características relevantes es posible mejorar el desempeño de los algoritmos, siendo que en precisión, sensibilidad y exactitud alcanzaron valores mayores de 80%. Se resalto también que es importante considerar factores aparte del texto, como lo pueden ser las interacciones con otros usuarios o frecuencia de publicaron para construir modelos efectivos para la detección suicida en redes sociales.

A hybrid deep learning approach for depression prediction from user tweets using feature-rich CNN and bi-directional LSTM (Kour & Gupta, 2022)

Esta investigación discutió el uso de una solución de aprendizaje profundo híbrida para predecir la depresión de tweets usuarios. El conjunto de datos utilizado para la detección de depresión fue extraído de Twitter y está dividido en 3 conjuntos: depresion dataset (2558 datos), non-depression dataset (5304 datos) y finalmente el grupo de control o candidato a tener depresión (58810 datos). Los tweets son etiquetados de forma binaria por expertos en salud mental y se clasifican en: depresivo o no depresivo. Se propuso el modelo de clasificación de CNN-biLSTM y se comparó con otros modelos mencionados en el estado del arte como lo puede ser CNN o RNN. El modelo propuesto obtuvo una puntuación F1 de 0.9478 y una exactitud de 0.9428, superando a los modelos RNN (puntuación F1 de 0.9178 y exactitud de 0.9066) y el CNN (puntuación F1 de 0.9264 y exactitud de 0.9173).

2.2.1 Discusión del estado del arte

Con base en los trabajos investigados en el estado del arte, se puede concluir que el proceso clasificación de textos suicidas es similar entre sí. Algunos trabajos varían en los algoritmos de clasificación, probando cuál es más eficiente y registrando sus resultados, pero manteniendo la fase de preprocesamiento de una manera similar. En inglés existen herramientas más desarrolladas, mientras que en español las herramientas pueden tener un menor desempeño; por ejemplo, la herramienta NLTK no permite lematizar en español. Sin embargo, lo más importante, la base de todos los trabajos y lo que en gran medida determina la efectividad del modelo a implementar son los datos utilizados y su representación numérica. Algunos trabajos recopilan datos de redes sociales, y con equipos de profesionales etiquetan la información para determinar cuáles textos son suicidas y cuáles no. Estos datos también pueden ser obtenidos de estudios clínicos previos de personas ingresadas en instituciones de salud mental. Desafortunadamente, los datos, al ser contenido sensible, no están a libre disposición y, pese a que pueden ser solicitados a los autores originales llenando algunos formatos, no es un proceso sencillo, pues no consideran el uso por parte de países extranjeros. Una posible solución para conseguir datos relacionados con el suicidio es explorar plataformas *web* que contengan *blogs* o foros relacionados al tema, siendo *Reddit* un ejemplo de ello. Esta red social permite tener foros de conversación sobre temas específicos, siendo el foro (también conocido como *subreddit*) *“r/SuicideWatch”* un lugar donde las personas escriben publicaciones relacionadas con el suicidio. Muchas investigaciones optan por utilizar esta opción para obtener datos ya que están dentro del contexto del suicidio. En caso de que el estudio enfoque los datos al idioma español, suelen hacerse traducciones a estas publicaciones, ya que no hay muchos datos disponibles para su uso en la investigación. Por otra parte, se puede apreciar que las técnicas de representaciones numéricas no suelen variar mucho, teniendo a Word2Vec o TF-IDF como las más utilizadas. Esto dejar por fuera técnicas que pueden tener mucha más representatividad de los textos al transformarse, perdiendo contexto importante.

En la tabla comparativa 1 se explica de forma reducida y concreta los puntos clave considerados en la selección de los artículos, así como los principales factores relevantes tomados en consideración para la selección de estos, facilitando la comparación de cantidad de datos, técnicas de representación y clasificadores con su respectiva métrica.

Tabla 1 Tabla comparativa del estado del arte

Autor	Datos	Representaciones numéricas	Clasificadores – Mejores resultados	
Wang, Yu, & Tian, 2019	4485 comentarios	--	--	--
García-Martínez et al., 2022	2059 tweets	--	--	--
Yao et al., 2020	--	TF-IDF, Word2Vec	CNN SVM LR RF	F1 96% F1 94.1% F1 94% F1 92%
Metzler, Baginski, Niederkrotenthaler, & Garcia, 2022	3202 tweets	TF-IDF	SVM XLNet BERT	ACC 66% ACC 74% ACC 73%
Li, Zhou, An, Cheng, & Hu, 2022	164854 publicaciones	--	DHE-SD TextCNN	ACC 95% – F1 93% ACC 90% – F1 91%
Luo, Du, Tao, Xu, & Zhang, 2020	716899 tweets	--	--	--
Castillo-Zúñiga, Luna-Rosas, & López-Veyna, 2022	--	--	Seq – RF Seq – NN Seq – LR Par – RF Par – LR Par – DT Par – RNN	PR 96.84% PR 96.84% PR 97.41% PR 94.98% RP 94.52% PR 95.11% PR 98%
Acuña Caicedo, Gómez Soriano, & Melgar Sasieta, 2020	Life corpus 102 textos	Bag of Words, Bag of Stems, Bag of Lemmas, Bag of SYSNET, Bag of POS	Weka software (28 clasificadores)	KStar F1 71% RandomCommittee F1 70%
Tadesse, Lin, Xu, & Yang, 2020	3549 post suicidas 3652 post no suicidas	Word2Vec, TF-IDF, Bag of Words	LSTM-CNN LSTM CNN XGBOOST	R 94% R 90% R 93% R 84%
Renjith, Abraham, Jyothis,	14849 sin riesgo 13691 riesgo bajo	Word2Vec TF-IDF	LSTM-Attention-CNN SVM LSTM	F1 92% F1 82% F1 88%

Chandran, & Thomson, 2022	13462 riesgo moderado 13678 riesgo severo		CNN LSTM-CNN	F1 88% 86%
Ramírez-Cifuentes et al., 2020	252 usuarios (sin datos)	Bag of Words, Bag of NGrams Word embeddings, características sociales, psicológicos.	RF SVM LR CNN MLP	F1 85% F1 86% F1 85% F1 82% F1 88%
Valeriano, Condori-Larico, & Sulla-Torres, 2020	2068 textos	TF-IDF Word2Vec	SVM LR	R 74% R 79%
Ji, Li, Huang, & Cambria, 2022	UMD Reddit dataset (865 usuarios) Reddit SWHM (54412 publicaciones) Twitter (4800 tweets)	Embeddings Lexicon Modelado de Tópicos	fastText CNN LSTM BiLSTM RCNN SSA RN	R 53-57-79% R 55-56-78% R 56-59-80% R 56-61-82% R 55-60-80% R 56-62-81% R 56-64-83%
Vioules, Moulahi, Aze, & Bringay, 2018	500 tweets	n-gramas (uni-gramas, bi-gramas)	RF-SMO SL-SMO J48-SMO	R 67%-81% R 70%-76% R 71%-79%
Buddhitha & Inkpen, 2023	UMD Reddit dataset (1732 usuarios) Reddit SWHM (306793 usuarios) Twitter (1145 usuarios)	fastText, GloVe, Byte-Pair embeddings, character embeddings, stacked embeddings	Modelo propuesto (MTL)	F1 88%
Liu et al., 2019	27007 comentarios	n-gramas	SVM DT RF LR	R 78% R 68% R 75% R79%
Kour & Gupta, 2022	2558 muestras depresivas, 5304 no depresivas	Embedding 2D del autor	CNN-biLSTM RNN CNN	R 92% R 93% R 92%

3 Metodología de solución

Se realizaron las siguientes tareas para crear una solución al problema planteado. En la Imagen 1 se presenta la metodología de solución completa, así como una descripción sobre las tareas realizadas y la ruta de solución:

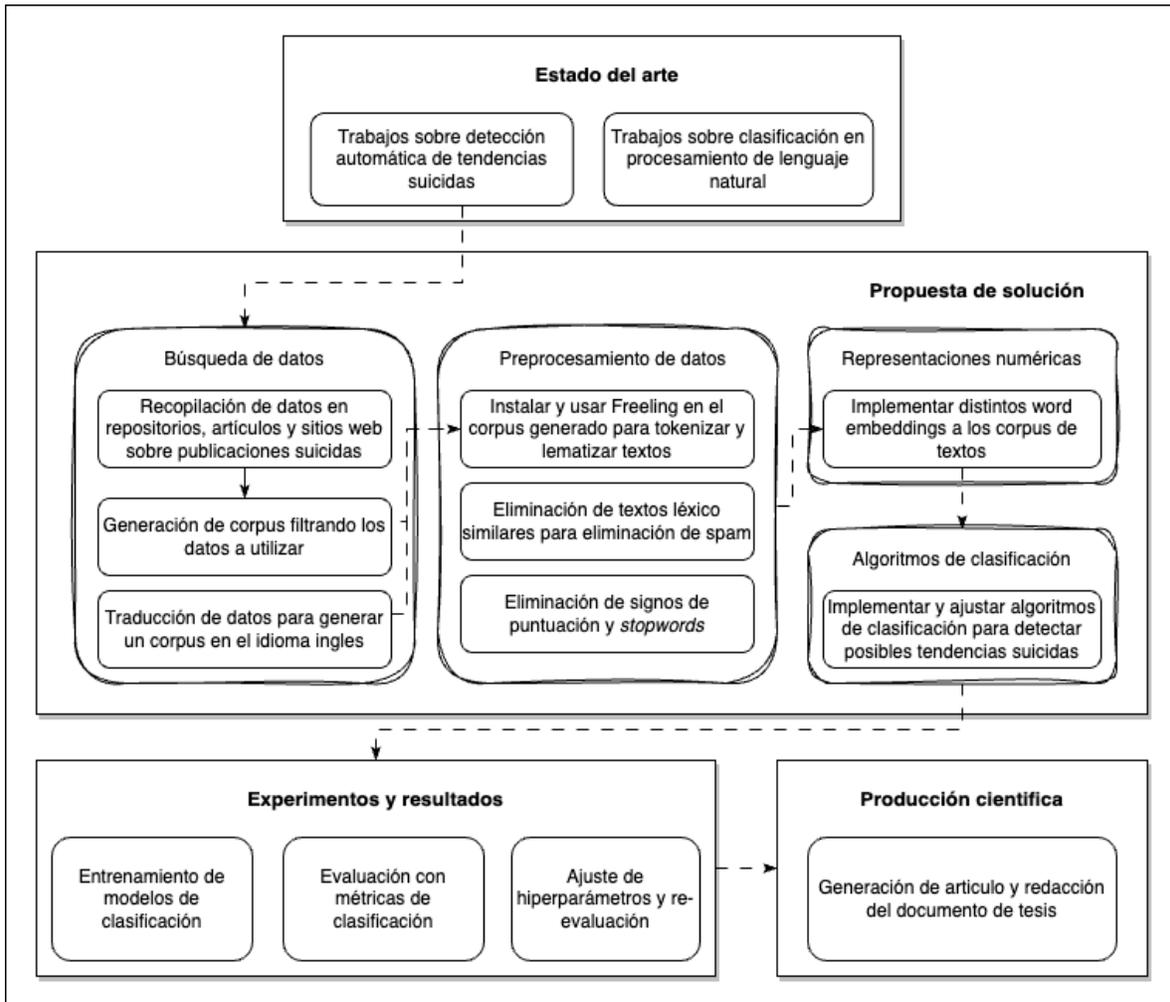


Imagen 1 Metodología de solución

3.1 Búsqueda y recopilación de datos, generación y limpieza de CSV

Esta sección cubre todo el proceso relacionado con la búsqueda, recopilación y generación de conjuntos de datos (*datasets*). El objetivo principal de esta actividad fue obtener datos que permitan identificar a individuos que expresen posibles ideaciones suicidas en publicaciones en redes sociales. Por lo tanto, es esencial que los textos recopilados sean claros y concisos, además de estar relacionados con el tema, para posteriormente extraer el contexto e información necesaria para la detección de dichas ideaciones.

Inicialmente, se realizó una recopilación de publicaciones con indicios de tendencias suicidas, con el propósito de identificar patrones que puedan servir para la clasificación automática de estas publicaciones. Uno de los principales problemas de los datos obtenidos fue la falta de detalle respecto a su validez médica. Como solución parcial, se recurre a la obtención de datos provenientes de la literatura especializada en suicidio o mediante búsquedas en repositorios de datos reconocidos, como Hugging Face o Kaggle, que son ampliamente utilizados en el campo de la inteligencia artificial como fuentes de datos para artículos y experimentos.

Para esta investigación, se recopilaron un total de 175,010 textos del repositorio de Hugging Face (Joheras, n.d.), de los cuales 77,223 textos fueron clasificados como suicidas y 97,787 como no suicidas. En la Imagen 2 se muestra la distribución de los datos que componen el repositorio, junto con los nombres específicos de los conjuntos de datos utilizados.

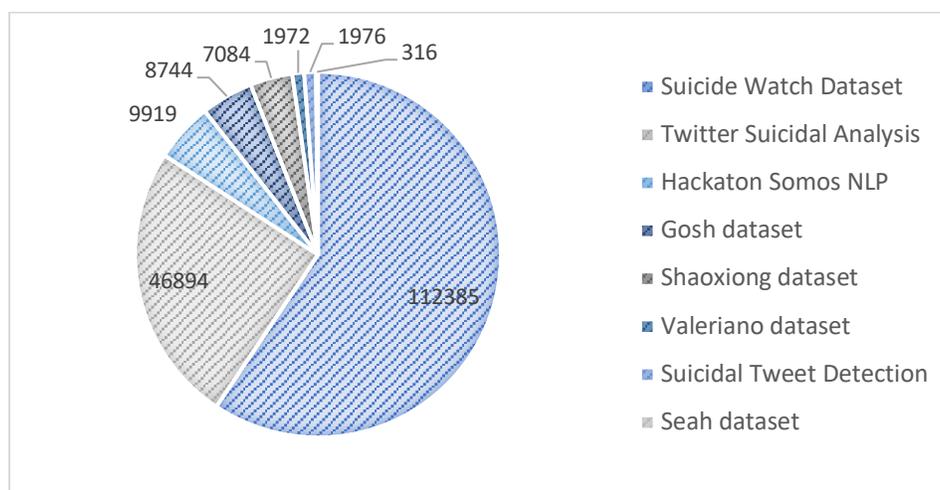


Imagen 2 Tamaño por conjunto de datos del repositorio

La tarjeta informativa del repositorio proporciona detalles sobre la fuente de cada conjunto de datos que lo integra, así como enlaces de referencia que conducen a los orígenes de la información. Para los experimentos realizados, se seleccionaron aleatoriamente 50,000 textos, distribuidos equitativamente entre las dos clases: un 50% correspondientes a publicaciones con tendencias suicidas y un 50% sin tendencias suicidas.

Se estableció como criterio que los textos seleccionados tuvieran una longitud máxima de 300 caracteres, con el objetivo de priorizar textos cortos. Una vez definidos los datos a utilizar, se almacenaron los textos seleccionados en archivos CSV. Posteriormente, se tradujeron al español, con la excepción del conjunto de datos del Hackaton Somos NLP, cuyos textos estaban originalmente en español y fueron traducidos al inglés. Para realizar las traducciones, se empleó la herramienta Google Translate debido a su facilidad de uso, rapidez de implementación y la posibilidad de realizar un gran número de traducciones de forma gratuita.

Después de la traducción, se conformaron los corpus en inglés y español, con el propósito de comparar la efectividad de los *embeddings* y evaluar la viabilidad de utilizar datos traducidos para el entrenamiento de los modelos.

3.2 Preprocesamiento de datos

El preprocesamiento de datos es una etapa fundamental en el procesamiento de lenguaje natural (PLN), ya que los textos en su estado original suelen contener ruido, como errores ortográficos, signos de puntuación innecesarios y palabras irrelevantes que pueden confundir a los modelos y afectar su rendimiento. Además, este proceso normaliza las palabras, permitiendo tratar diferentes formas de una misma palabra de manera uniforme, lo que facilita la generalización de los modelos. Otra ventaja importante del preprocesamiento es la reducción de la dimensionalidad de los datos, eliminando características innecesarias, lo cual permite a los modelos concentrarse en los aspectos más relevantes del texto. Dado que los textos son no estructurados, es necesario convertirlos en representaciones numéricas para que los algoritmos de aprendizaje automático puedan procesarlos.

Entre las principales técnicas de preprocesamiento se encuentran la tokenización, la lematización, la eliminación de stopwords, la normalización y la vectorización de textos. Herramientas como NLTK, spaCy, Freeling, Gensim, OpenNLP, TextBlob, Pandas y TF-IDF se utilizan comúnmente para estos fines, permitiendo realizar tareas como dividir el texto en tokens, eliminar palabras irrelevantes, obtener la forma base de las palabras mediante la lematización, y convertir el texto en vectores numéricos aptos para ser procesados por modelos de aprendizaje automático. A través de estas técnicas, el preprocesamiento no solo limpia y organiza los datos textuales, sino que también optimiza el análisis en tareas de PLN.

Para el preprocesamiento de datos en esta investigación, se empleó la librería Freeling (Atserias et al., 2006; Carreras, Chao, Padró, & Padró, 2004; Padró, 2011; Padró, Collado, Reese, Lloberes, & Castellón, 2010; Padró & Stanilovsky, 2012), una herramienta de análisis lingüístico especialmente útil para lenguas romances, que ofrece funcionalidades como análisis morfológico, reconocimiento de entidades y etiquetado gramatical (POS tagging), entre otros. Este proceso permitió llevar a cabo tareas clave para la limpieza de los datos, tales como la eliminación de categorías gramaticales innecesarias, la tokenización y la eliminación de *stopwords* y signos de puntuación, reduciendo así la longitud de los textos y conservando únicamente las palabras relevantes para el análisis.

Otra acción importante fue la eliminación de oraciones repetidas, una práctica crucial en tareas de clasificación de textos para mejorar la precisión y eficiencia de los modelos. Las oraciones duplicadas pueden generar sesgos en los datos,

ya que los algoritmos podrían considerar su frecuencia como indicativo de relevancia, lo que distorsiona la distribución de las clases y puede llevar a un sobreajuste. Este fenómeno afecta la capacidad del modelo para generalizar de manera efectiva con nuevos datos. Asimismo, las oraciones repetidas incrementan innecesariamente la carga computacional, prolongando los tiempos de entrenamiento y procesamiento. Eliminar estas redundancias asegura que los modelos se entrenen con datos diversos y representativos, mejorando su rendimiento en la clasificación de diferentes contextos.

Para identificar y eliminar oraciones similares, se emplearon técnicas que permiten medir la similitud léxica o semántica entre textos, como la distancia de Levenshtein, que evalúa el número mínimo de cambios necesarios para transformar una oración en otra, y la similitud coseno. Esta última técnica es ampliamente utilizada en PLN y se basa en representar las oraciones como vectores en un espacio multidimensional, donde cada dimensión corresponde a una palabra del vocabulario. La similitud se calcula tomando el coseno del ángulo entre los vectores; un coseno cercano a 1 indica alta similitud, mientras que un valor cercano a 0 sugiere baja similitud. Esta técnica es útil porque no se ve afectada por la longitud de las oraciones y se centra en la dirección de los vectores, capturando la similitud relativa entre las palabras.

Cuando se combina la similitud coseno con técnicas de representación de palabras como TF-IDF o *embeddings* (por ejemplo, Word2Vec o BERT), es posible realizar una evaluación eficaz de la similitud semántica entre oraciones, permitiendo identificar y eliminar aquellas redundantes sin comprometer la información clave. Esta metodología optimiza la calidad de los datos y maximiza la capacidad de los modelos para generalizar y clasificar con precisión en tareas relacionadas con la detección de ideación suicida en redes sociales.

3.3 Representaciones numéricas

Existen diversas técnicas para transformar textos en formatos comprensibles para una máquina, como números o vectores. Entre las representaciones numéricas tradicionales se incluyen *One-Hot Encodings*, *Bag of Words* (BOW) y *Term Frequency – Inverse Document Frequency* (TF-IDF). Asimismo, los *word embeddings* como Word2Vec, GloVe y los modelos basados en *Transformers* se destacan entre los enfoques más recientes. Todas estas técnicas buscan modelar el lenguaje en un formato numérico; sin embargo, es importante señalar que algunas son más eficientes que otras.

Tabla 2 Comparativa de técnicas de representaciones numéricas.

Característica	Tradicionales	Word Embeddings
Semántica	No captura relaciones semánticas	Captura similitudes semánticas
Dimensionalidad	Alta y dispersa	Baja y densa
Eficiencia computacional	Fácil de implementar, pero con matrices dispersas	Entrenamiento complejo, pero eficiente después
Generalización	Limitada a palabras específicas	Generaliza mejor a sinónimos y polisemia
Adaptabilidad a nuevos datos	Requiere recalcular matrices	Puede ajustarse con fine-tuning
Contexto	No tiene en cuenta el contexto	Modelos avanzados como BERT o GPT capturan contexto

En la Tabla 2 se ofrece un análisis concreto de las propiedades de cada método de representación numérica, evidenciando por qué los *word embeddings* resultan más eficientes en tareas de procesamiento de lenguaje natural.

Entre las características específicas de los *word embeddings* se destacan las siguientes:

1. **Captura de semántica:** Los *word embeddings* logran captar similitudes semánticas entre palabras, lo que permite que términos relacionados, como "rey" y "reina" o "perro" y "gato", estén próximos en el espacio vectorial, reflejando sus relaciones semánticas.

2. **Dimensionalidad reducida:** A diferencia de los modelos tradicionales como *Bag of Words*, que generan representaciones de alta dimensionalidad y dispersas, los *embeddings* reducen significativamente esta dimensionalidad, facilitando un procesamiento más eficiente del texto.
3. **Generalización y transferencia:** Los *word embeddings* pueden ser entrenados en grandes corpus de texto y reutilizados en diversas tareas de procesamiento de lenguaje natural (PLN), como la clasificación de textos, la traducción automática o el reconocimiento de entidades. Este enfoque fomenta una mayor capacidad de generalización, ya que las representaciones aprendidas pueden ser aplicadas a distintos dominios.
4. **Resolución de sinónimos y polisemia:** Las representaciones vectoriales permiten una identificación más precisa de sinónimos. Aunque algunos modelos tradicionales enfrentan dificultades con la polisemia (palabras con múltiples significados), los modelos más avanzados logran diferenciar mejor los contextos en que estas palabras aparecen.

Existen diversos métodos para la creación de *word embeddings*, cada uno con características, fortalezas y limitaciones particulares. A continuación, se presenta una comparativa entre algunos de los enfoques más relevantes, como *Word2Vec*, *GloVe*, *MPNet* y *GPT-3*:

- **Word2Vec (Mikolov, Chen, Corrado, & Dean, 2013):** Este método, desarrollado por Google en 2013, es uno de los primeros y más populares para la generación de *embeddings* de palabras. Ofrece dos variantes principales: *skip-gram* y *continuous bag of words* (CBOW), las cuales buscan predecir palabras a partir de su contexto. *Word2Vec* es eficiente y rápido en su entrenamiento, además de capturar de manera efectiva las relaciones semánticas entre palabras. Para esta investigación, se utilizó un modelo preentrenado, específicamente el “word2vec-google-news-300” (Google, n.d.), entrenado con aproximadamente 100 millones de palabras de noticias de Google, cuyo vector resultante tiene 300 dimensiones.
- **GloVe (Pennington, Socher, & Manning, 2014):** Desarrollado en la Universidad de Stanford, GloVe combina las ventajas de los modelos basados en conteo de palabras con los modelos predictivos, como *Word2Vec*. Este enfoque utiliza matrices de co-ocurrencias globales de palabras para capturar relaciones semánticas. Aunque aprovecha mejor la información global del corpus en comparación con otros métodos, sigue enfrentando limitaciones en el manejo de palabras polisémicas, ya que cada palabra tiene una representación vectorial única.

- **MPNet (Song, Tan, Qin, Lu, & Liu, 2020):** Desarrollado por Microsoft, MPNet es una evolución de los modelos de enmascaramiento de palabras como *BERT* y *XLNet*. MPNet enmascara palabras dentro de un texto y también permuta las secuencias, lo que mejora su capacidad para capturar relaciones semánticas en diferentes posiciones del texto. Aunque presenta un rendimiento superior en tareas de comprensión y generación de texto, es más costoso computacionalmente que modelos como *Word2Vec* o *GloVe*, y requiere mayor cantidad de datos para su entrenamiento. Existen modelos preentrenados como el disponible en Hugging Face, “all-mpnet-base-v2,” que genera vectores de 768 dimensiones.
- **GPT-3 (OpenAI, n.d.-b, n.d.-a):** Este modelo, desarrollado por OpenAI, incluye variantes como “text-embedding-3-small” y “text-embedding-3-large”, basadas en la arquitectura *transformer*. Los embeddings de GPT-3 muestran un rendimiento superior respecto a generaciones anteriores debido a su entrenamiento en vastos corpus de datos, lo que permite realizar múltiples tareas de PLN como búsqueda y similitud de textos. El “text-embedding-3-small” mantiene las 1536 dimensiones del modelo anterior “text-embedding-ada-002”, pero introduce mejoras en métricas como MIRACL y MTEB, a un costo reducido. El “text-embedding-3-large” incrementa la dimensionalidad a 3072, lo que le permite un mayor rendimiento en la comprensión y representación de información, superando a sus predecesores en métricas clave sin aumentar significativamente el costo.

3.4 Algoritmos de clasificación empleados

En esta investigación, los algoritmos de clasificación se utilizaron para identificar y detectar patrones que indicaran ideación suicida en redes sociales. El objetivo principal fue generar etiquetas que señalaran la probabilidad de que los usuarios estuvieran expresando pensamientos suicidas en sus publicaciones.

Para este fin, fueron empleados cinco algoritmos de clasificación: Árboles de Decisión (DT), Bosques Aleatorios (RF), Regresión Logística (LR), Máquinas de Soporte Vectorial (SVM) y Perceptrón Multicapa (MLP). Estos algoritmos se aplicaron sobre los datos previamente procesados mediante representaciones vectoriales utilizando *word embeddings*. Antes de su aplicación, los datos pasaron por un riguroso proceso de limpieza y preprocesamiento para maximizar la eficiencia de los resultados. Asimismo, la información fue dividida en conjuntos de entrenamiento y prueba, lo cual es fundamental para evaluar adecuadamente el desempeño de los modelos.

En este experimento, se asignó el 75% de los datos para el entrenamiento de los modelos de clasificación, mientras que el 25% restante se destinó a la evaluación a través de pruebas. Esta división se estableció con el propósito de asegurar un equilibrio adecuado que permita a los modelos aprender a partir de una amplia porción de los datos, mientras que un subconjunto significativo se reserva para pruebas objetivas en datos no vistos. De esta forma, se minimiza el riesgo de sobreajuste, garantizando una evaluación robusta de la capacidad de generalización del modelo.

Una vez completada la fase de entrenamiento, la evaluación del desempeño de los modelos se realizó mediante métricas de clasificación clave. Estas métricas son esenciales para analizar y comparar la eficacia de cada clasificador y sus resultados respectivos. En este caso, se utilizaron las métricas de Exactitud (Accuracy), Precisión (Precision), Sensibilidad (Recall) y Puntuación F1 (F1-Score). Sin embargo, se puso especial énfasis en la métrica de Sensibilidad, debido a la relevancia de identificar correctamente los casos de ideación suicida, minimizando los falsos negativos. Dado el carácter delicado de este proyecto, es crucial priorizar la detección precisa de textos con alta probabilidad de contener ideación suicida, ya que un error en este sentido podría tener graves consecuencias. Estas métricas proporcionan una evaluación integral y diversa, permitiendo valorar el desempeño de los modelos desde varias perspectivas y garantizando la robustez del análisis.

4 Experimentos y resultados

En los experimentos realizados, se emplearon cinco modelos de *word embeddings* para representar los textos, los cuales fueron luego utilizados para entrenar cinco algoritmos de clasificación automática. Esto resultó en un total de 25 combinaciones entre embeddings y clasificadores, como se detalla en la Tabla 3.

Este enfoque permitió comparar el rendimiento de diversas combinaciones de técnicas de representación y clasificación, proporcionando una visión integral de cómo interactúan estos métodos y cuál es su efectividad en la tarea de detección de ideación suicida en redes sociales. El análisis de estas combinaciones es crucial para identificar las mejores estrategias, maximizando el desempeño de los modelos de clasificación mediante el uso adecuado de embeddings que capturen de manera eficiente la semántica de los textos

Tabla 3 Puntajes de la validación cruzada entre embedding y clasificador

Models	DT	RF	LR	SVM	MLP
Word2Vec	0.72	0.79	0.82	0.65	0.83
GloVe	0.68	0.76	0.81	0.61	0.81
MPNet	0.74	0.82	0.86	0.71	0.86
GPT-3-Small	0.75	0.84	0.88	0.73	0.89
GPT-3-Large	0.75	0.85	0.88	0.77	0.90

Cada clasificador genera una salida binaria: “Posible ideación suicida” (representada por el valor 1) y “Sin ideación suicida” (valor 0).

Los resultados obtenidos para cada modelo de *embedding* fueron graficados tomando en cuenta los valores de exactitud del modelo, así como las métricas específicas de la clase suicida, tales como precisión, sensibilidad y puntuación F1. Estas métricas permiten evaluar de manera más precisa el rendimiento del modelo al detectar casos de ideación suicida, que es el objetivo principal de este estudio.

Adicionalmente, se llevó a cabo un ajuste de hiperparámetros para cada clasificador utilizando la técnica de *GridSearch*, disponible en la librería *ScikitLearn* de Python. Este ajuste tuvo como objetivo optimizar el rendimiento de los clasificadores, buscando obtener resultados equilibrados en las distintas métricas evaluadas. En la Tabla 4 presenta la configuración de los

hiperparámetros óptimos encontrados para cada combinación de *embedding* y clasificador, lo que permitió afinar el desempeño general de los modelos en este experimento.

Tabla 4 Hiperparámetros implementados por *embedding* – clasificador

Models	Hyperparameters used
GloVe – DT	'criterion': 'entropy', 'max_depth': 500, 'max_features': None, 'min_samples_leaf': 61, 'min_samples_split': 88, 'splitter': 'best'
GloVe – LR	'solver': 'liblinear', 'penalty': 'l2', 'max_iter': 100
GloVe – MLP	'activation': 'identity', 'alpha': 0.012834151269284711, 'hidden_layer_sizes': [50, 30], 'learning_rate': 'invscaling', 'max_iter': 300, 'solver': 'lbfgs'
GloVe – RF	'criterion': 'entropy', 'max_depth': None, 'max_features': None, 'min_samples_leaf': 8, 'min_samples_split': 17, 'n_estimators': 300
GloVe – SVM	'max_iter': 900, 'kernel': 'sigmoid', 'gamma': 'auto', 'degree': 5
GPT-3 – DT	'criterion': 'gini', 'max_depth': None, 'max_features': None, 'min_samples_leaf': 53, 'min_samples_split': 71, 'splitter': 'best'
GPT-3 – LR	'solver': 'newton-cg', 'penalty': 'l2', 'max_iter': 700
GPT-3 – MLP	'activation': 'relu', 'alpha': 0.004390090923029554, 'hidden_layer_sizes': [1500, 600], 'learning_rate': 'adaptive', 'max_iter': 400, 'solver': 'lbfgs'
GPT-3 – RF	'criterion': 'entropy', 'max_depth': 500, 'max_features': 'log2', 'min_samples_leaf': 18, 'min_samples_split': 85, 'n_estimators': 300
GPT-3 – SVM	'max_iter': 700, 'kernel': 'rbf', 'gamma': 'auto', 'degree': 4
GPT-3-L – DT	'criterion': 'entropy', 'max_depth': 500, 'max_features': None, 'min_samples_leaf': 61, 'min_samples_split': 88, 'splitter': 'best'
GPT-3-L – LR	'solver': 'newton-cg', 'penalty': 'l2', 'max_iter': 700
GPT-3-L – MLP	'activation': 'relu', 'alpha': 0.004390090923029554, 'hidden_layer_sizes': [1500, 600], 'learning_rate': 'adaptive', 'max_iter': 400, 'solver': 'lbfgs'
GPT-3-L – RF	'criterion': 'entropy', 'max_depth': 300, 'max_features': 'sqrt', 'min_samples_leaf': 70, 'min_samples_split': 29, 'n_estimators': 100
GPT-3-L – SVM	'max_iter': 900, 'kernel': 'sigmoid', 'gamma': 'auto', 'degree': 5
MPNet – DT	'criterion': 'gini', 'max_depth': None, 'max_features': None, 'min_samples_leaf': 53, 'min_samples_split': 71, 'splitter': 'best'
MPNet – LR	'solver': 'sag', 'penalty': 'l2', 'max_iter': 900
MPNet – MLP	'activation': 'relu', 'alpha': 0.004390090923029554, 'hidden_layer_sizes': [1500, 600], 'learning_rate': 'adaptive', 'max_iter': 400, 'solver': 'lbfgs'
MPNet – RF	'criterion': 'entropy', 'max_depth': 500, 'max_features': 'log2', 'min_samples_leaf': 18, 'min_samples_split': 85, 'n_estimators': 300

MPNet – SVM	'max_iter': 700, 'kernel': 'rbf', 'gamma': 'auto', 'degree': 4
Word2vec – DT	'criterion': 'entropy', 'max_depth': 500, 'max_features': None, 'min_samples_leaf': 61, 'min_samples_split': 88, 'splitter': 'best'
Word2vec – LR	'solver': 'newton-cholesky', 'penalty': 'l2', 'max_iter': 100
Word2vec – MLP	'activation': 'relu', 'alpha': 0.004390090923029554, 'hidden_layer_sizes': [1500, 600], 'learning_rate': 'adaptive', 'max_iter': 400, 'solver': 'lbfgs'
Word2vec – RF	'criterion': 'entropy', 'max_depth': 500, 'max_features': 'log2', 'min_samples_leaf': 18, 'min_samples_split': 85, 'n_estimators': 300
Word2vec – SVM	'max_iter': 900, 'kernel': 'sigmoid', 'gamma': 'auto', 'degree': 5

Basándose en los resultados del experimento, se seleccionaron y graficaron únicamente aquellos modelos que obtuvieron un rendimiento sobresaliente en la validación cruzada, alcanzando puntajes superiores al 85%. En la Imagen 3 se muestran los resultados específicos obtenidos utilizando el *embedding* Word2Vec. Los clasificadores más balanceados para la detección de la clase suicida fueron Regresión Logística (LR), Perceptrón Multicapa (MLP) y Bosques Aleatorios (RF), con puntajes que oscilaron entre el 79% y el 84% en las métricas clave (precisión, sensibilidad y puntuación F1). El clasificador Árboles de Decisión (DT) obtuvo un rendimiento ligeramente inferior, con un promedio de 71.5% en cada métrica, mientras que el modelo de Máquinas de Soporte Vectorial (SVM) fue el menos efectivo, destacándose por su baja precisión (62%) y una sensibilidad del 78%. Aunque los valores generales fueron relativamente bajos, la métrica de sensibilidad mostró una notable eficiencia en la identificación de textos con ideación suicida, lo que subraya la capacidad del modelo para minimizar los falsos negativos en esta clase crítica.

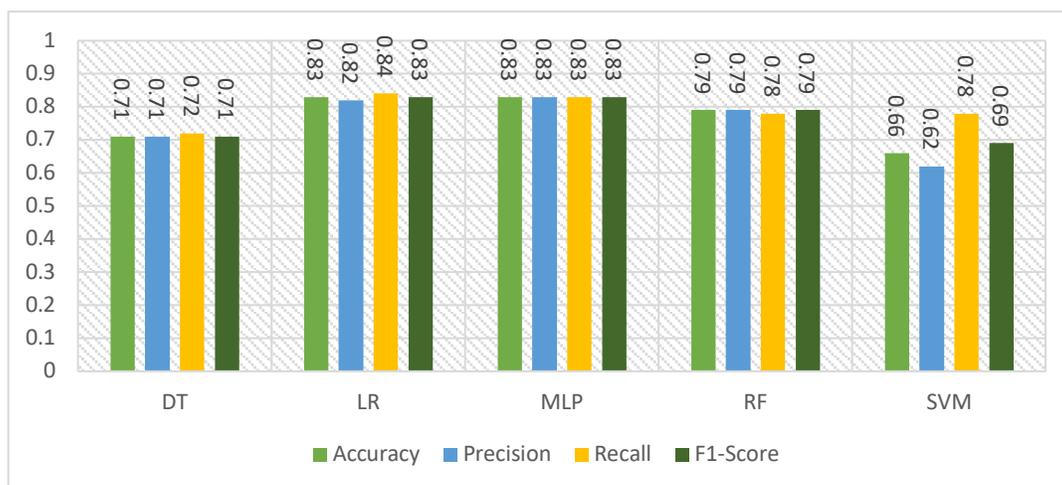


Imagen 3 Resultados con Word2Vec

Por su parte, el *embedding* GloVe, que comparte una dimensionalidad similar a Word2Vec, mostró resultados comparables, con diferencias marginales de entre 1% y 2% en cada puntaje, como se ilustra en la Imagen 4. Estas variaciones mínimas sugieren que el rendimiento de los *embeddings* podría estar influido por los datos utilizados para su entrenamiento. En este caso, el *embedding* desarrollado por Google, Word2Vec, parece ser ligeramente óptimo, lo que podría explicarse por la especificidad del corpus utilizado en su entrenamiento, proporcionando una mejor adaptación a las características del conjunto de datos evaluado.

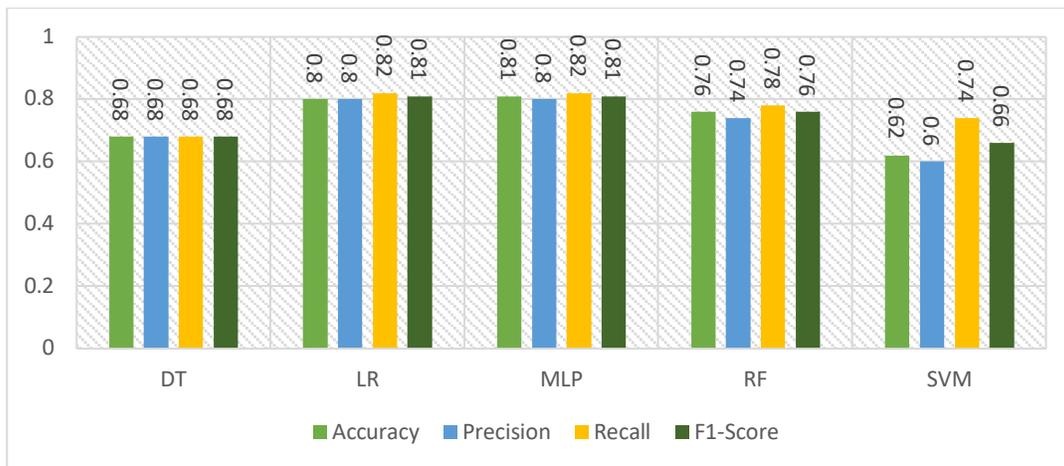


Imagen 4 Resultados con GloVe

En la Imagen 5 los mejores resultados obtenidos con el *embedding* MPNet correspondieron a los modelos de Regresión Logística (LR) y Perceptrón Multicapa (MLP), con puntajes entre 85% y 86%, superando ligeramente a los *embeddings* previos. Los modelos de Árboles de Decisión (DT) y Máquinas de Soporte Vectorial (SVM) presentaron los rendimientos más bajos, sin alcanzar el 80% en ninguna de las métricas evaluadas. Por otro lado, el modelo de Bosques Aleatorios (RF) obtuvo un 77% en Sensibilidad y más de un 80% tanto en Precisión como en la Puntuación F1, destacándose como un clasificador balanceado en este contexto.

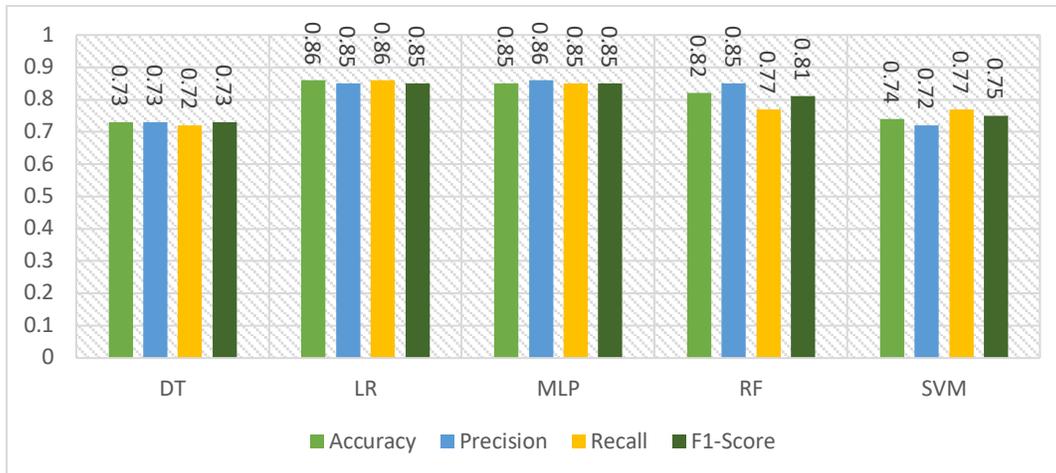


Imagen 5 Resultados con MPNet

Finalmente, en la Imagen 6 y en la Imagen 7 los puntajes obtenidos con los *embeddings* GPT-3 demostraron ser los más estables y eficientes, destacándose los clasificadores Regresión Logística (LR) y Perceptrón Multicapa (MLP) como los más efectivos. Con GPT-3 Small, el modelo MLP alcanzó un 89% en exactitud, precisión, sensibilidad y puntuación F1, seguido de cerca por LR, que obtuvo un 88% en cada métrica.

Asimismo, clasificadores como Árboles de Decisión (DT), Bosques Aleatorios (RF) y Máquinas de Soporte Vectorial (SVM) también mostraron un incremento en sus puntajes al utilizar estos *embeddings*, confirmando la eficiencia de GPT-3 en la mejora del desempeño de los modelos de clasificación

Por último, GPT-3 Large presentó las mejores métricas entre todos los *embeddings* evaluados, destacándose el clasificador Perceptrón Multicapa (MLP), que alcanzó un 90% en todas las métricas, con una sensibilidad particularmente alta del 91%. El segundo mejor clasificador fue Regresión Logística (LR), que logró un puntaje del 89% en todas las métricas.

Sin embargo, se observó una diferencia notable en el desempeño de Máquinas de Soporte Vectorial (SVM). A diferencia de GPT-3 Small, este clasificador mostró una mayor variabilidad en sus resultados, con una disminución del 12% en precisión, pero un incremento significativo en la sensibilidad, alcanzando un 89%. Este comportamiento refleja los desafíos que los *embeddings* con altas dimensionalidades pueden presentar para algunos clasificadores, lo que puede traducirse en una reducción drástica de los puntajes o un aumento en el tiempo de entrenamiento.

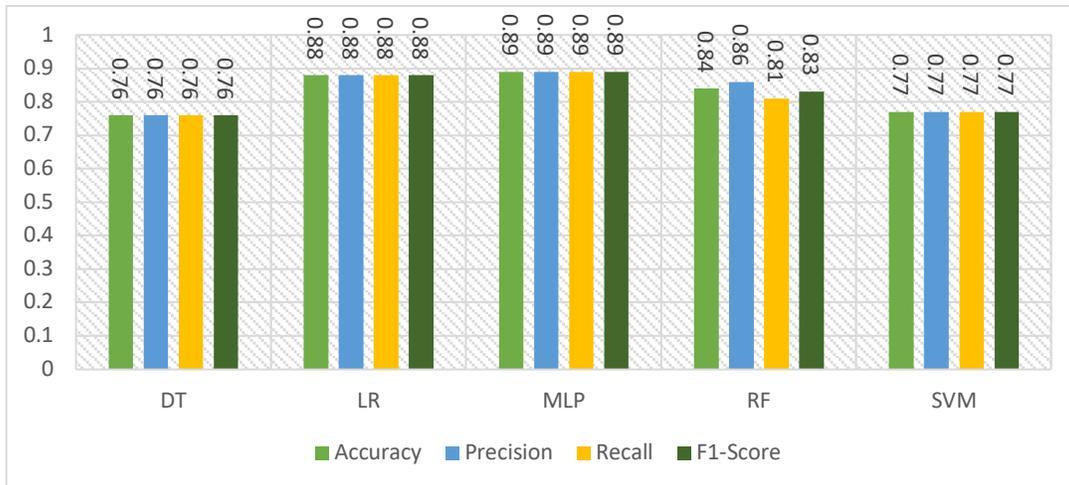


Imagen 6 Resultados con GPT-3 Small

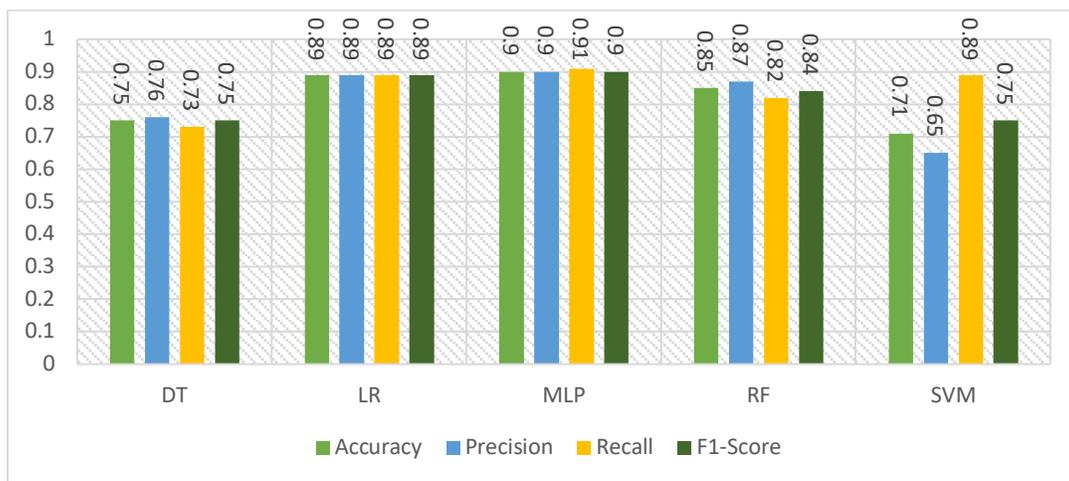


Imagen 7 Resultados con GPT-3 Large

De todas las combinaciones evaluadas, únicamente aquellas que superaron un umbral del 85% en la validación cruzada fueron seleccionadas para un análisis más detallado. Este criterio fue establecido para asegurar que solo los modelos más confiables y precisos sean considerados en la interpretación de resultados, garantizando así un enfoque riguroso y centrado en el desempeño óptimo de los algoritmos de clasificación y *embeddings* en la detección de ideación suicida.

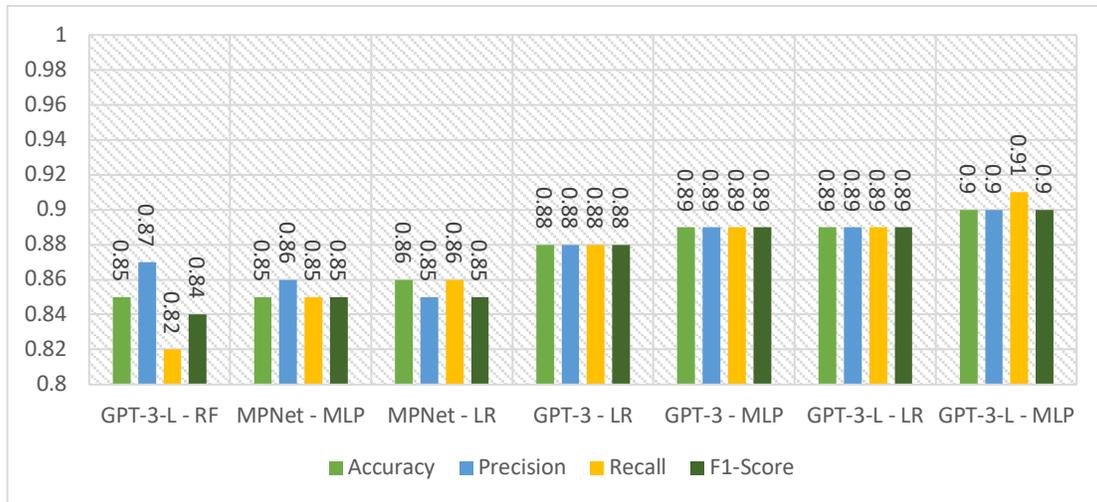


Imagen 8 Mejores combinaciones basándose en la puntuación de la validación cruzada, umbral 85%+

En la Imagen 8 se destacan siete combinaciones de *embeddings* y clasificadores que superaron el umbral del 85% en la validación cruzada. Estas combinaciones lograron un mínimo en la métrica de Sensibilidad del 82% y un máximo del 91%. Los puntajes obtenidos reflejan un equilibrio en todas las métricas, con la excepción de la combinación de GPT-3 Large y RF, donde se observó cierta varianza. El uso de *embeddings* de alta dimensionalidad demostró ser altamente eficaz para la clasificación de ideación suicida, con GPT-3 Small y Large emergiendo como las representaciones más destacadas, seguidos por los *embeddings* de MPNet.

Estos *embeddings* mostraron su capacidad para capturar de manera efectiva la información contextual en los textos, lo que resultó en puntajes significativos en las métricas evaluadas. Por otro lado, los clasificadores MLP y LR alcanzaron los puntajes más altos cuando se combinaron con los *embeddings*, destacando especialmente la combinación de GPT-3 Large con MLP, que logró resultados sobresalientes, obteniendo un 90% en cada una de las métricas evaluadas.

Como detalle adicional, es importante destacar que el uso de *embeddings* como GPT-3 Small puede generar resultados favorables a un costo significativamente menor en comparación con GPT-3 Large. Esta diferencia en costos puede ser un factor relevante en escenarios donde los recursos son limitados. Por otro lado, MPNet demostró ser una opción efectiva, logrando puntajes superiores al 85% en todas las métricas según el clasificador utilizado. A diferencia de los *embeddings* de GPT-3, MPNet no conlleva gastos

monetarios adicionales, lo que lo convierte en una alternativa viable y más económica para tareas de clasificación de ideación suicida en textos.

Una vez realizados los experimentos con los datos traducidos al español, se procedió a comparar la efectividad de los modelos respecto a los resultados obtenidos con los datos en inglés.

4.1 Clasificación con textos traducidos

La experimentación con embeddings en textos permite representar el significado de palabras y frases en un espacio vectorial, capturando relaciones semánticas y contextuales de manera eficiente. Al utilizar técnicas como *word embeddings* se logra que el modelo comprenda y conserve la esencia del texto, independientemente del idioma en que esté escrito. Este enfoque es especialmente útil en traducción y procesamiento de lenguajes naturales multilingües, ya que facilita que las representaciones semánticas de un texto en un idioma se mantengan consistentes al traducirlo a otro, lo cual optimiza el rendimiento en tareas de clasificación, búsqueda y análisis semántico entre idiomas.

Además de los experimentos anteriores, se llevó a cabo una prueba adicional con textos traducidos para evaluar su relevancia en la contextualización mediante *embeddings* en tareas de clasificación. El principal objetivo de este experimento fue comprobar que los *embeddings*, independientemente del idioma, mantienen una efectividad equivalente al representar y clasificar una misma oración traducida.

Los datos utilizados para este experimento fueron los siguientes:

Tabla 5 Sets de datos para traducción

Datos	Suicidas	No suicidas
Github (inglés) ¹	3990 textos	5121 textos
Life Corpus (inglés) ²	183 textos	90 textos
HuggingFace (español) ³	777 textos	9273 textos
Despues de limpieza (GH + LC)	3390 textos	4851 textos

¹ https://github.com/RezwanM/Suicide_Tweets

² <https://github.com/PlataformaLifeUA/corpus>

³ <https://huggingface.co/datasets/hackathon-somos-nlp-2023/suicide-comments-es>

Los datos están compuestos por dos idiomas: inglés y español. El conjunto final de datos a utilizar se conformó mediante la unión de los datos en inglés obtenidos de Github y el Life Corpus, así como del único conjunto de datos en español extraído de HuggingFace. Cabe señalar que los datos presentan un notable desbalance hacia la clase no suicida; sin embargo, el objetivo en este caso no es lograr una clasificación efectiva, como en estudios previos, sino comprobar que las traducciones mantienen valores consistentes respecto a sus textos originales.

Los experimentos realizados se conforman de la siguiente manera:

1. Entrenamiento y pruebas con datos en español nativo.
2. Entrenamiento y pruebas con datos en español traducido al inglés.
3. Entrenamiento y pruebas con datos en inglés nativo.
4. Entrenamiento y pruebas con datos en inglés traducido al español.

Tabla 6 Efectividad en textos traducidos con algoritmos de clasificación

Experimento	Modelo – Embedding	Accuracy	Clase 0 (No suicida)			Clase 1 (Suicida)		
			PR	R	F1	PR	R	F1
1	SVM: GPT-3	0.97	0.97	0.99	0.98	0.90	0.73	0.81
	<u>MLP: GPT-3</u>	0.97	0.98	0.98	0.98	0.84	0.78	0.81
2	SVM: GPT-3	0.97	0.97	0.99	0.98	0.89	0.71	0.79
	<u>MLP: GPT-3</u>	0.97	0.98	0.98	0.98	0.86	0.79	0.82
	SVM: MPNET	0.97	0.97	0.99	0.98	0.87	0.72	0.79
3	SVM: GPT-3	0.93	0.91	0.96	0.94	0.95	0.89	0.92
	<u>MLP: GPT-3</u>	0.93	0.92	0.95	0.94	0.94	0.90	0.92
4	<u>SVM: GPT-3</u>	0.92	0.90	0.95	0.92	0.94	0.87	0.90
	<u>MLP: GPT-3</u>	0.90	0.90	0.92	0.91	0.90	0.87	0.89

Los resultados muestran que la traducción de un conjunto de datos no afecta significativamente los resultados de su contraparte original, lo cual permite concluir que los *embeddings* constituyen la opción más adecuada para trabajar de manera multilingüe con un mismo conjunto de datos, facilitando así la transferencia de conocimiento de un idioma a otro.

4.2 Clasificación con textos traducidos y redes neuronales

Tomando en cuenta la comparación anterior, el objetivo de esta comparación fue determinar si es adecuado o inadecuado trabajar con traducciones en tareas de clasificación de ideación suicida. Los resultados se presentan en la

Imagen 9 donde se utilizó el mismo umbral del 85% en la validación cruzada para identificar los *embeddings* más representativos y los clasificadores más eficientes.

Al igual que en los experimentos en inglés, este umbral se estableció para garantizar que solo los modelos más confiables y precisos fueran considerados. Además, se observó una diferencia en los puntajes obtenidos con los datos traducidos, lo que permitió evaluar la viabilidad de utilizar traducciones en modelos de clasificación de ideación suicida. Cabe señalar que, aunque en algunos trabajos el mínimo aceptable para las métricas de clasificación suele ser del 80%, debido a la naturaleza delicada de este proyecto es fundamental aplicar un criterio más estricto para asegurar la confiabilidad del sistema y su capacidad para detectar correctamente posibles tendencias suicidas.

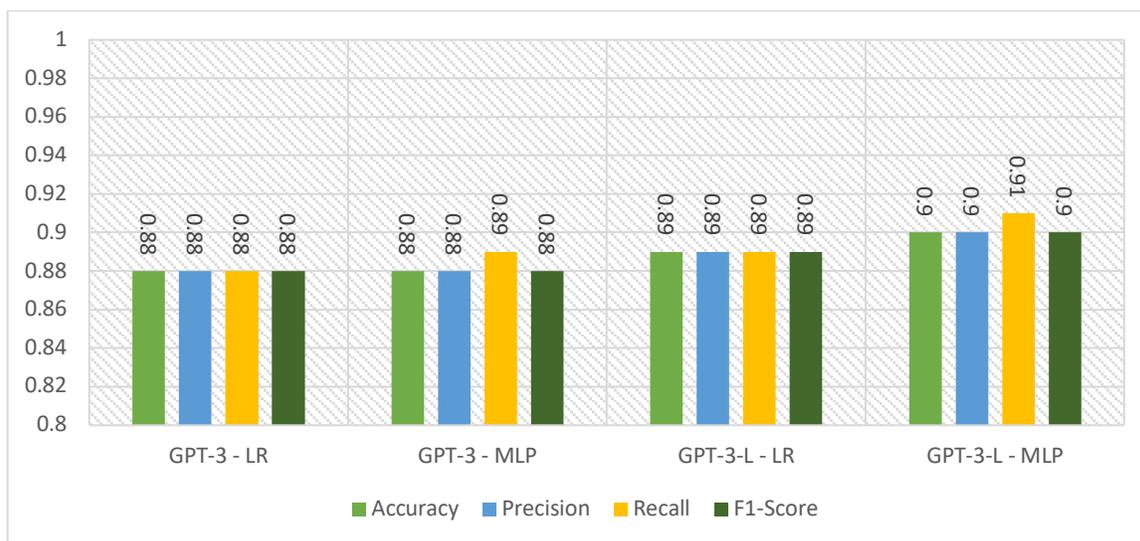


Imagen 9 Mejores combinaciones basándose en la puntuación de la validación cruzada, umbral 85%+ en idioma español

En la Imagen 9 los resultados muestran que solamente cuatro modelos superaron el umbral del 85% en la validación cruzada con los datos traducidos al español, en comparación con los siete modelos que alcanzaron este umbral en el experimento con los datos en inglés. A pesar de la reducción en el número de modelos sobresalientes, los *embeddings* de GPT-3 Small y Large siguen destacando como los más relevantes, combinados con los clasificadores MLP y LR. Estos clasificadores, al igual que en el experimento en inglés, demostraron ser los más adecuados para la detección de tendencias suicidas en redes sociales.

Los puntajes obtenidos por estos modelos varían entre el 88% y el 91%, con la combinación de GPT-3 Small y LR en el extremo inferior (88%) y la combinación de GPT-3 Large y MLP en el extremo superior (91%). Esto refuerza la conclusión de que GPT-3 Large y MLP son los más eficientes y confiables para la tarea de clasificación en ambas lenguas. Además, es importante resaltar que los modelos que superaron el umbral en español son los mismos que los cuatro primeros modelos en inglés, lo que implica que los *embeddings* de GPT-3 mantienen una contextualización coherente y efectiva tanto en inglés como en español.

Este comportamiento similar entre ambos idiomas se refleja también en los puntajes, que son prácticamente equivalentes, lo que sugiere que los *embeddings* de GPT-3 permiten una representación robusta del contexto, independientemente del idioma del texto procesado. Esto confirma que el uso de traducciones no tiene un impacto negativo significativo en los resultados cuando se emplean estos *embeddings* avanzados

Por último, se llevó a cabo un experimento final utilizando el mejor *embedding*, GPT-3 Large, implementado en un modelo de aprendizaje más profundo, con el objetivo de aprovechar plenamente su alta dimensionalidad y extraer mayor cantidad de información en comparación con los modelos tradicionales. Para este experimento, se diseñó una red neuronal aplicada tanto a los datos en inglés como en español, con el fin de comparar su efectividad frente a modelos menos complejos. En la Imagen 10 se presenta el diseño completo de la red neuronal.

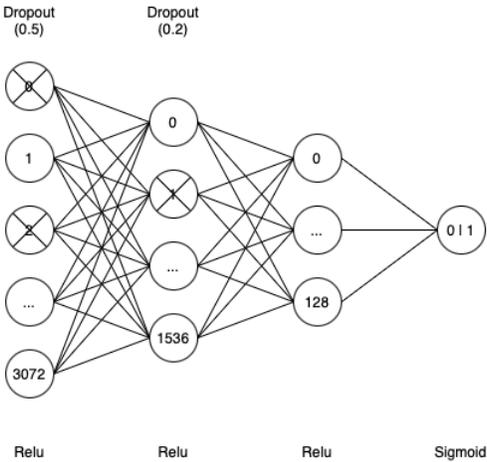


Imagen 10 Diseño de red neuronal

En este diseño, se observa que la red neuronal está compuesta por cuatro capas: una capa de entrada con 3072 neuronas, dos capas ocultas con 1536 y 128 neuronas, respectivamente, y una capa de salida. Todas las capas, excepto la capa de salida, emplean la función de activación *ReLU*, mientras que en la capa de salida se utilizó una función sigmoidea, lo que permite obtener resultados en un rango de 0 a 1. El proceso de entrenamiento se llevó a cabo durante 20 épocas, con un tamaño de *batch* de 32, utilizando el optimizador Adam y la función de pérdida logarítmica (*binary_crossentropy*). Los resultados obtenidos tanto con los datos en inglés como en español se presentan en la Imagen 11.

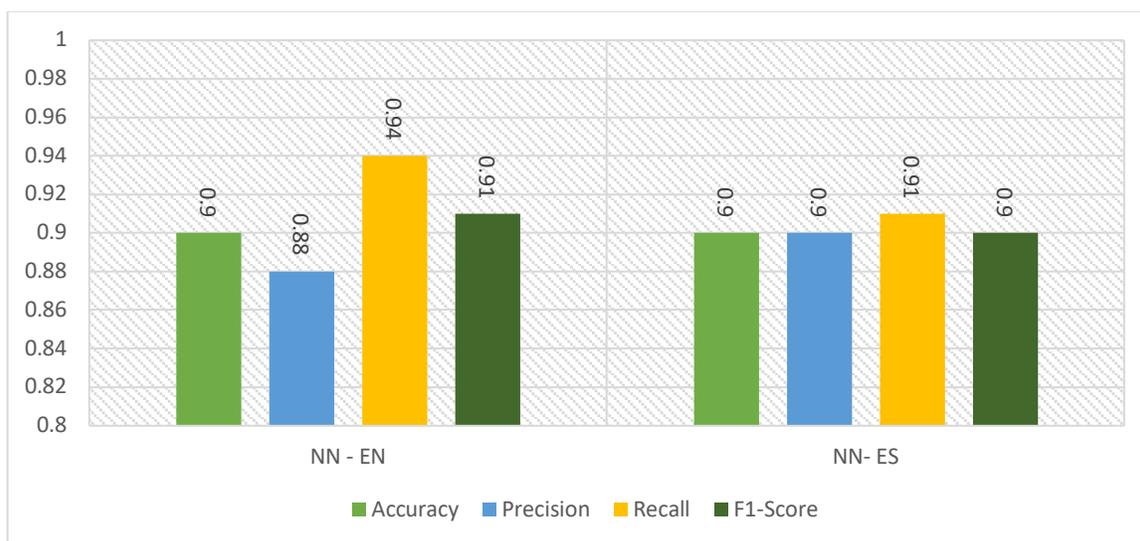


Imagen 11 Resultados inglés – español de la red neuronal

Estos resultados indican que, al emplear un modelo más optimizado para trabajar con altas dimensionalidades, es posible superar los resultados obtenidos por los modelos clásicos. Además, el uso de tecnologías más recientes permite aprovechar mejor los recursos tecnológicos, reduciendo el tiempo de procesamiento. Los resultados obtenidos con los datos en inglés muestran que las puntuaciones alcanzaron un 90%, destacando especialmente la métrica de Sensibilidad, que superó a todos los modelos probados anteriormente con un 94%, mientras que la Precisión fue la única métrica ligeramente inferior, con un 88%. En cuanto a los datos en español, aunque los resultados no fueron sobresalientes, todas las métricas alcanzaron al menos un 90%.

5 Conclusiones

Se llevaron a cabo una serie de experimentos con el propósito de comparar la efectividad de diversas combinaciones entre embeddings y algoritmos de clasificación. La fase inicial del proyecto consistió en la recopilación de datos provenientes de estudios previos, bases de datos en línea y repositorios especializados. A partir de los datos obtenidos, se conformó un corpus de 50,000 textos, los cuales fueron sometidos a un proceso de preprocesamiento que incluyó la eliminación de *stopwords* y signos de puntuación.

Posteriormente, se aplicaron cinco técnicas distintas de embeddings a los textos, transformándolos en vectores numéricos. Estas técnicas de embeddings varían en cuanto a su dimensionalidad, con el objetivo de extraer la mayor cantidad de contexto posible de los textos, permitiendo representar la información de manera que sea interpretable por un sistema computacional. Finalmente, los clasificadores automáticos fueron implementados y alimentados con la información generada por los embeddings, para llevar a cabo la tarea de clasificación de manera precisa y eficiente.

Los resultados del experimento arrojaron las siguientes conclusiones:

- **Embeddings GPT-3:** Los embeddings de GPT-3 mostraron el mejor rendimiento general, destacándose **GPT-3 Large** con los puntajes más altos en todas las métricas (90% de Exactitud, Sensibilidad, Precisión y Puntuación F1). Los mejores resultados se lograron utilizando modelos basados en redes neuronales, específicamente con un **Perceptrón Multicapa (MLP)** y una **red neuronal profunda** que incluía una capa de entrada y tres capas ocultas. Otros modelos como la **Regresión Logística (LR)** también mostraron un desempeño destacado.
- **MPNet:** El modelo de embeddings **MPNet** tuvo un rendimiento ligeramente superior en comparación con los embeddings **Word2Vec** y **GloVe**, obteniendo puntajes que oscilaron entre el 85% y 86%, especialmente en combinación con los clasificadores **LR** y **MLP**.
- **Word2Vec y GloVe:** Estos dos embeddings presentaron un rendimiento similar, con puntajes en el rango del 79% al 84% en todas las métricas cuando se utilizaron con modelos como **LR**, **MLP**, y **Bosques Aleatorios (RF)**.

En general, los resultados sugieren que el uso de embeddings más avanzados como GPT-3 puede proporcionar un desempeño superior en tareas de clasificación de textos complejos como la detección de ideación suicida.

En cuanto al rendimiento de los clasificadores automáticos, los resultados fueron los siguientes:

- **Perceptrón Multicapa (MLP):** Este clasificador fue el que obtuvo los mejores puntajes en todos los embeddings, alcanzando valores entre 83% y 91% en las métricas de Exactitud, Precisión, Sensibilidad y Puntuación F1. **Regresión Logística (LR)** también mostró resultados notables, con valores cercanos a los de MLP, fluctuando entre 82% y 89% en cada métrica.
- **Árboles de Decisión (DT):** Este clasificador presentó un desempeño moderadamente bueno, con un promedio de 71.5% en el embedding **Word2Vec**, aunque sus resultados fueron menores en **GloVe**, con puntajes alrededor del 68%. Por otro lado, **Bosques Aleatorios (RF)** mostró cierta variación, alcanzando una menor Sensibilidad (77%) pero logrando más de 80% en Exactitud, Precisión y Puntuación F1 cuando se usaron embeddings de alta dimensionalidad.
- **Máquina de Soporte Vectorial (SVM):** Este clasificador presentó el peor desempeño general. Su puntaje más bajo fue del 60% en Precisión utilizando **GloVe**, y experimentó una disminución significativa del 12% en Precisión cuando se utilizó **GPT-3 Large**, en comparación con otros modelos. En general, SVM fue el clasificador más inestable, con una considerable variación en sus métricas en comparación con otros clasificadores.

Los resultados obtenidos a partir de las traducciones mostraron conclusiones similares a las evidenciadas en los datos en inglés, lo que indica que es factible trabajar con traducciones y alcanzar resultados con diferencias mínimas en las métricas, con un máximo de variación del 1%. Además, los resultados derivados de las redes neuronales más profundas evidencian que es posible incrementar el valor de ciertas métricas, como es el caso de la Sensibilidad, la cual logró aumentar hasta un 94%, aunque conllevando una reducción de la Precisión en un 2%. En este contexto, el límite observable en los modelos tradicionales puede ser superado mediante el uso de métodos más robustos, capaces de identificar patrones significativos en tareas de procesamiento del lenguaje natural (NLP).

6 Producción científica

Participación en Feria de Ciencia 2023: Se dio una plática a chicos de preparatoria sobre inteligencia artificial y procesamiento de lenguaje natural.



Participación en el Coloquio de Lingüística Computacional: Se presento y acepto un trabajo en el coloquio de la UNAM donde se habló sobre la eliminación automática de datos léxico similares.



participación en el Segundo Simposio Iberoamericano de Maestría y Doctorado en Inteligencia Artificial: Se envió un artículo corto de la tesis actual, reportando los primeros resultados obtenidos.



SIMDIA2023 <simdia2023@easychair.org>

para mí ▾

lun, 23 oct 2023, 16:53



Estimado(a) autor(a),

Gracias por su envío a SIMDIA 2023. Nos complace informarle que su artículo fue aceptado para su inclusión en el programa y presentación en SIMDIA 2023.

Adjuntamos los informes de los revisores para que incorpore en su ponencia todas las mejoras posibles.

Sugerimos preste atención a la revisión que ha sido más crítica con su artículo.

Publicación de artículo titulado “Word Embeddings and Machine Learning Classifiers Applications for Automatic Detection of Suicide Tendencies in Social Media” en Programming and Computer Software FI 0.7 (2023) Q4 (Scopus: Q3) – Volumen 50, paginas 612–620.

ISSN 0361-7688, Programming and Computer Software, 2024, Vol. 50, No. 8, pp. 612–620. © Pleiades Publishing, Ltd., 2024.

Word Embeddings and Machine Learning Classifiers Applications for Automatic Detection of Suicide Tendencies in Social Media

Juan-Pablo Álvarez-López^{a,*} (ORCID: 0009-0005-5127-6191)
and Noé-Alejandro Castro-Sánchez^{a,**} (ORCID: 0000-0002-8083-3891)

^aTecnológico Nacional de México (TecNM), Centro Nacional de Investigación y Desarrollo Tecnológico (CENIDET), Morelos, México

*e-mail: m22ce042@cenidet.tecnm.mx

**e-mail: noe.cs@cenidet.tecnm.mx

Received May 7, 2024; revised August 16, 2024; accepted September 12, 2024

Abstract—This study presents an innovative and comprehensive model for the automatic detection of suicidal ideation in social media posts. Through an in-depth analysis of 50000 posts and the combination of four word embedding techniques (Word2Vec, GloVe, MPNet, and GPT-3) with five advanced classifiers, we have achieved an accuracy of over 90% in identifying users who may be at risk. Our results suggest that the integration of large language models like GPT-3 with deep neural network architectures offers a promising tool for suicide prevention in the digital realm, contributing to the development of automated screening systems capable of alerting mental health professionals to potential cases of risk.

Keywords: social media, suicide tendencies, natural language processing, text embeddings, classification algorithms

DOI: 10.1134/S0361768824700658

7 Referencias

- Acuña Caicedo, R. W., Gómez Soriano, J. M., & Melgar Sasieta, H. A. (2020). Assessment of supervised classifiers for the task of detecting messages with suicidal ideation. *Heliyon*, *6*(8), e04412. <https://doi.org/10.1016/j.heliyon.2020.e04412>
- Atserias, J., Casas, B., Comelles, E., González, M., Padró, L., & Padró, M. (2006). FreeLing 1.3: Syntactic and semantic services in an open-source NLP library. *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC 2006)*. Genoa, Italy.
- Augusto Cortez Vásquez, M., Hugo Vega Huerta, M., Jaime, L., & Quispe, P. (2009). *Procesamiento de lenguaje natural*.
- Baron, O. P. (2000). *Adolescencia y suicidio*. Barranquilla. Retrieved from <http://www.redalyc.org/articulo.oa?id=21300605>
- Barrionuevo, J. (2009). Suicidio e intentos de suicidio. *Buenos Aires: UBA, Facultad de Psicología*.
- Buddhitha, P., & Inkpen, D. (2023). Multi-task learning to detect suicide ideation and mental disorders among social media users. *Frontiers in Research Metrics and Analytics*, *8*. <https://doi.org/10.3389/frma.2023.1152535>
- Carreras, X., Chao, I., Padró, L., & Padró, M. (2004). FreeLing: An Open-Source Suite of Language Analyzers. *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC'04)*.
- Castillo-Zúñiga, I., Luna-Rosas, F. J., & López-Veyna, J. I. (2022). Detection of traits in students with suicidal tendencies on Internet applying Web Mining. *Comunicar*, *30*(71), 1–13. <https://doi.org/10.3916/C71-2022-08>
- Castro Sánchez, N. A. (2021). Procesamiento de diccionarios en la lingüística computacional. *Sociedad Mexicana de Inteligencia Artificial*.
- Coseriu, E. (1951). *Introducción a la lingüística*. Retrieved from www.artnovela.com.ar
- Durkheim, E. (1989). *El suicidio* (Vol. 37). Ediciones Akal.
- García-Martínez, C., Oliván-Blázquez, B., Fabra, J., Martínez-Martínez, A. B., Pérez-Yus, M. C., & López-Del-Hoyo, Y. (2022). Exploring the Risk of Suicide in Real Time on Spanish Twitter: Observational Study. *JMIR Public Health and Surveillance*, *8*(5). <https://doi.org/10.2196/31800>
- Google. (n.d.). Word2Vec – Google News. Retrieved from <https://code.google.com/archive/p/word2vec/>
- Ji, S., Li, X., Huang, Z., & Cambria, E. (2022). Suicidal ideation and mental disorder detection with attentive relation networks. *Neural Computing and Applications*, *34*(13), 10309–10319. <https://doi.org/10.1007/s00521-021-06208-y>

- Joheras. (n.d.). PrevenirA/spanish–suicide–intent · Datasets at Hugging Face. Retrieved from <https://huggingface.co/datasets/PrevenirA/spanish–suicide–intent>
- Kour, H., & Gupta, M. K. (2022). An hybrid deep learning approach for depression prediction from user tweets using feature–rich CNN and bi–directional LSTM. *Multimedia Tools and Applications*, *81*(17), 23649–23685. <https://doi.org/10.1007/s11042–022–12648–y>
- Li, Z., Zhou, J., An, Z., Cheng, W., & Hu, B. (2022). Deep Hierarchical Ensemble Model for Suicide Detection on Imbalanced Social Media Data. *Entropy*, *24*(4). <https://doi.org/10.3390/e24040442>
- Liu, X., Liu, X., Sun, J., Yu, N. X., Sun, B., Li, Q., & Zhu, T. (2019). Proactive suicide prevention online (PSPO): Machine identification and crisis management for Chinese social media users with suicidal thoughts and behaviors. *Journal of Medical Internet Research*, *21*(5). <https://doi.org/10.2196/11705>
- Luo, J., Du, J., Tao, C., Xu, H., & Zhang, Y. (2020). Exploring temporal suicidal behavior patterns on social media: Insight from Twitter analytics. *Health Informatics Journal*, *26*(2), 738–752. <https://doi.org/10.1177/1460458219832043>
- Ma, X., Hancock, J., & Naaman, M. (2016). Anonymity, intimacy and self–disclosure in social media. *Conference on Human Factors in Computing Systems – Proceedings*, 3857–3869. Association for Computing Machinery. <https://doi.org/10.1145/2858036.2858414>
- Márquez, N. P. (2008). Diferentes aproximaciones al concepto de equivalencia en traducción y su aplicación en la práctica profesional. *Tonos Digital*, *15*.
- MedlinePlus en español [Internet]. (2022, July 26). Suicidio.
- Metzler, H., Baginski, H., Niederkrotenthaler, T., & Garcia, D. (2022). Detecting Potentially Harmful and Protective Suicide–Related Content on Twitter: Machine Learning Approach. *Journal of Medical Internet Research*, *24*(8). <https://doi.org/10.2196/34705>
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *ArXiv Preprint ArXiv:1301.3781*.
- Moutier, C. (2021). Conducta suicida.
- OpenAI. (n.d.–a). Embeddings. Retrieved from <https://openai.com/index/new–embedding–models–and–api–updates>
- OpenAI. (n.d.–b). *Embeddings Platform*. Retrieved from <https://platform.openai.com/docs/guides/embeddings>
- Orihuela, J. L. (2008). *Internet: la hora de las redes sociales*. 57–62.
- Padró, L. (2011). Analizadores Multilingües en FreeLing. *Linguamatica*, *3*(2), 13–20.

- Padró, L., Collado, M., Reese, S., Lloberes, M., & Castellón, I. (2010). FreeLing 2.1: Five Years of Open-Source Language Processing Tools. *Proceedings of 7th Language Resources and Evaluation Conference (LREC'10)*. La Valletta, Malta.
- Padró, L., & Stanilovsky, E. (2012). FreeLing 3.0: Towards Wider Multilinguality. *Proceedings of the Language Resources and Evaluation Conference (LREC 2012)*. Istanbul, Turkey.
- Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543.
- Ramírez-Cifuentes, D., Freire, A., Baeza-Yates, R., Puntí, J., Medina-Bravo, P., Velazquez, D. A., ... González, J. (2020). Detection of suicidal ideation on social media: Multimodal, relational, and behavioral analysis. *Journal of Medical Internet Research, 22*(7).
<https://doi.org/10.2196/17758>
- Renjith, S., Abraham, A., Jyothi, S. B., Chandran, L., & Thomson, J. (2022). An ensemble deep learning technique for detecting suicidal ideation from posts in social media platforms. *Journal of King Saud University – Computer and Information Sciences, 34*(10), 9564–9575.
<https://doi.org/10.1016/j.jksuci.2021.11.010>
- Rivera-Rivera, L., Fonseca-Pedrero, E., Séris-Martínez, M., Vázquez-Salas, A., & Reynales-Shigematsu, L. M. (2020). Prevalencia y factores psicológicos asociados con conducta suicida en adolescentes. Ensanut 2018–19. *Salud Pública de México, 62*(6, Nov-Dic), 672–681.
<https://doi.org/10.21149/11555>
- Secretaría de Salud. (2022). Programa Nacional para la Prevención de Suicidio. In <https://www.gob.mx/salud/documentos/programa-nacional-para-la-prevencion-de-suicidio?tab=>. Retrieved from <https://www.gob.mx/salud/documentos/programa-nacional-para-la-prevencion-de-suicidio?tab=>
- Sierra, G., Andrade-Palos, P., Bel-Enguix, G., Osornio-Arteaga, A., Cabrera-Mora, A., García-Nieto, L., & Sierra-Aparicio, T. (2022). Suicide Risk Factors: A Language Analysis Approach in Social Media. *Journal of Language and Social Psychology, 41*(3), 312–330.
<https://doi.org/10.1177/0261927X211036171>
- Song, K., Tan, X., Qin, T., Lu, J., & Liu, T.-Y. (2020). Mpnet: Masked and permuted pre-training for language understanding. *Advances in Neural Information Processing Systems, 33*, 16857–16867.
- Tadesse, M. M., Lin, H., Xu, B., & Yang, L. (2020). Detection of suicide ideation in social media forums using deep learning. *Algorithms, 13*(1).
<https://doi.org/10.3390/a13010007>

- Valdez–Santiago, R., Villalobos Hernández, A., Arenas–Monreal, L., Benjet, C., & Vázquez García, A. (2023). Conducta suicida en México: análisis comparativo entre población adolescente y adulta. *Salud Pública de México*, *65*, s110–s116. <https://doi.org/10.21149/14815>
- Valeriano, K., Condori–Larico, A., & Sulla–Torres, J. (2020). Detection of suicidal intent in Spanish language social networks using machine learning. *International Journal of Advanced Computer Science and Applications*, *11*(4), 688–698. <https://doi.org/10.14569/IJACSA.2020.0110489>
- Vioules, M. J., Moulahi, B., Aze, J., & Bringay, S. (2018). Detection of suicide–related posts in Twitter data streams. *IBM Journal of Research and Development*, *62*(1). <https://doi.org/10.1147/JRD.2017.2768678>
- Wang, Z., Yu, G., & Tian, X. (2019). Exploring behavior of people with suicidal ideation in a Chinese online suicidal community. *International Journal of Environmental Research and Public Health*, *16*(1). <https://doi.org/10.3390/ijerph16010054>
- WHO. (2021, June 17). Suicidio. Retrieved October 31, 2024, from World Health Organization website: <https://www.who.int/es/news-room/fact-sheets/detail/suicide>
- Yao, H., Rashidian, S., Dong, X., Duanmu, H., Rosenthal, R. N., & Wang, F. (2020). Detection of suicidality among opioid users on reddit: Machine learning–based approach. *Journal of Medical Internet Research*, *22*(11). <https://doi.org/10.2196/15293>