

Luis Roberto García Noguez

Identificación de rasgos asociados a la depresión mediante técnicas de aprendizaje automático y procesamiento de lenguaje natural.

2023



Universidad Autónoma de Querétaro

Facultad de Ingeniería

Identificación de rasgos asociados a la depresión mediante técnicas de aprendizaje automático y procesamiento de lenguaje natural

Tesis

Que como parte de los requisitos para obtener el grado Maestro en Ciencias en Inteligencia Artificial

Presenta

Luis Roberto García Noguez

Dirigido por:

Dr. Saúl Tovar Arriaga



Dirección General de Bibliotecas y Servicios Digitales
de Información



Identificación de rasgos asociados a la depresión
mediante técnicas de aprendizaje automático y
procesamiento de lenguaje natural

por

Luis Roberto García Noguez

se distribuye bajo una [Licencia Creative Commons
Atribución-NoComercial-SinDerivadas 4.0
Internacional](#).

Clave RI: IGMAC-302395



Universidad Autónoma de Querétaro

Facultad de Ingeniería

Maestría en Ciencias en Inteligencia Artificial

Identificación de rasgos asociados a la depresión mediante técnicas de aprendizaje automático y procesamiento de lenguaje natural.

TESIS

Que como parte de los requisitos para obtener el grado de Maestro en Ciencias en Inteligencia Artificial

Presenta:

Luis Roberto García Noguez

Dirigido por:

Dr. Saúl Tovar Arriaga

SINODALES

Dr. Saúl Tovar Arriaga
Presidente

M.C. Wilfrido Jacobo Paredes García
Secretario

Dr. Juan Manuel Ramos Arreguin
Vocal

Dr. Sebastián Salazar Colores
Suplente

Med. Esp. Humberto Guendulain Arenas
Suplente

Centro Universitario, Querétaro, QRO. México.
Febrero 2023

A mis seres queridos, quienes me han acompañado a lo largo del camino.

Agradecimientos

Primero, me gustaría agradecer al Consejo Nacional de Ciencia y Tecnología (CONACYT) por brindarme el apoyo económico para la elaboración de este trabajo de investigación y mi formación académica. En este mismo rubro, quiero dar las gracias a la Universidad Autónoma de Querétaro y a la dirección de Investigación y Posgrado de la Facultad de Ingeniería por brindarme las herramientas y elementos necesarios a lo largo de mis estudios. Asimismo, quiero reconocer el constante apoyo de mis profesores que fueron guías en mi desarrollo académico. El trabajo realizado no sería posible sin su fructífero apoyo. Finalmente, agradezco a mi familia y compañeros, ya que sin su apoyo no hubiera sido posible concluir este objetivo.

Abstract

In recent years, the identification of mental illness traits on social media platforms has become popular with the increased use of these internet sites. However, detecting these features in depressed patients through the use of Artificial Intelligence tools is complex due to certain points that make it difficult to correctly detect this type of illness, such as the similarity of symptoms between similar mental illnesses. Working on this point, this study analyzed depressed users of social networks from the CLPsych 2015 database based on two of the most common types of depression: Major Depressive Disorder and Persistent Depressive Disorder. To achieve this, an analysis of temporality, symptomatology, and emotion was carried out to distinguish between these two types of depression. Likewise, a CNN model was used to classify the users. Better results were obtained based on the metrics of accuracy, F1 score, precision, and recall (85.72%, 85.45%, 85.82%, and 85.51%) for the group of proposed users with features of Major Depressive Disorder. In general, the classification provided higher results when users were divided into the two proposed groups compared to when classification included the two joined sets. This study is a positive attempt to analyze the traits associated with depression based on two of the most common types of this pathology using temporal criteria and distinctive emotions to reduce biases resulting from the lack of specificity observed in previous studies. This work represents an advance in the characterization of subjects with traits of depression who use the Twitter platform.

Resumen

En los últimos años, la identificación de rasgos de enfermedad mental en las plataformas de redes sociales se ha vuelto popular con el aumento en el uso de estos sitios de internet. Sin embargo, detectar estos rasgos en pacientes depresivos mediante el uso de herramientas de Inteligencia Artificial resulta complejo debido a ciertas características que dificultan la correcta detección de este tipo de enfermedades, como la similitud de síntomas entre enfermedades mentales similares. Trabajando en este punto, en este estudio se analizaron a los usuarios deprimidos de las redes sociales de la base de datos CLPsych 2015 en función de dos de los tipos más comunes de depresión: el Trastorno Depresivo Mayor y el Trastorno Depresivo Persistente. Para lograrlo, se llevó a cabo un análisis de temporalidad, sintomatológico y emocional para distinguir estos dos tipos de depresión. Asimismo, se usó un modelo CNN para clasificar a los usuarios. Se obtuvieron mejores resultados en base a las métricas de exactitud, puntaje F1, precisión y sensibilidad (85,72 %, 85,45 %, 85,82 % y 85,51 %) para el grupo de usuarios propuestos con características de Trastorno Depresivo Persistente. En general, el clasificar brindó resultados más altos cuando se clasificaron a los usuarios en los dos grupos propuestos en comparación a cuando se clasificaron los dos conjuntos unidos. Este estudio es un intento positivo de analizar los rasgos asociados a la depresión en función de dos de los tipos más comunes de esta patología utilizando criterios temporales y emociones distintivas para reducir los sesgos a consecuencia de la falta de especificidad observada en estudios anteriores, lo que supone un avance en la caracterización de sujetos con rasgos de depresión que usan la plataforma Twitter.

Abreviaturas y Siglas

TDM – Trastorno Depresivo Mayor

TDP – Trastorno Depresivo Persistente

CLPsych – Computational Linguistics and Clinical Psychology – Lingüística
Computacional y Psicología Clínica

TEPT – Trastorno de Estrés Post-Traumático

IA – Inteligencia Artificial

CNN – Convolutional Neural Network – Red Neuronal Convolutiva

EBN – Explainable Boosting Machine – Máquina Explicable de Boosting

RNA – Red neuronal artificial

ML – Machine Learning – Aprendizaje automático

PLN – Procesamiento de Lenguaje Natural

BERT – Bidirectional Encoder Representations from Transformers – Representaciones de
Codificadores Bidireccionales a partir de Transformadores

SVM – Support Vector Machine – Máquinas de Vectores de Soporte

KNN – K – Nearest Neighbors – K- vecinos más cercanos

Índice General

Agradecimientos	IV
Abstract	V
Resumen	VI
Abreviaturas y Siglas	VII
Índice General	VIII
Índice Tablas	X
Índice Figuras	XI
CAPÍTULO 1	1
Introducción	1
1.1 Descripción del Problema	3
1.2 Justificación.	4
Hipótesis	5
Objetivos	5
1.3.1 Objetivo general	5
1.3.2 Objetivos específicos	5
CAPÍTULO 2	7
Antecedentes	7
2.1 Antecedentes.....	7
2.2 Estado del Arte	8
2.3 Características en el lenguaje de pacientes con depresión.....	18
Marco Teórico	18
2.4 Procesamiento de lenguaje Natural	18
2.4.1 Lingüística	19
2.4.2 Lingüística Computacional	19

2.4.3	Actividades comúnmente desarrolladas en proyectos de PLN en IA	20
2.4.4	Aplicaciones del PLN	21
2.4.5	Análisis de emociones en el texto.....	22
2.4.6	GloVe.....	24
2.5	Aprendizaje profundo.....	25
2.5.1	CNN	25
2.5.2	Global Max Pooling	26
2.5.3	Capa densa	26
2.5.4	Dropout.....	27
2.6	Trastornos Depresivos.....	27
2.6.1	Trastorno Depresivo Mayor	28
2.6.2	Trastorno Depresivo Persistente	30
CAPÍTULO 3		32
Metodología		32
3.1	Conjunto de datos	32
3.2	Metodología propuesta	32
3.3	Pre-procesamiento	33
3.4	Obtención de la puntuación sintomática	34
3.4.1	Valor de tristeza	35
3.4.2	Actividad Nocturna.....	35
3.4.3	Palabras relacionadas con pensamientos suicidas.....	35
3.5	Separación del conjunto de datos	36
3.6	Modelo de clasificación	37
3.6.1	Selección de hiperparámetros.....	38
CAPÍTULO 4		40
Resultados		40
4.1	Métricas de evaluación.....	40
4.2	Entrenamiento.....	41
4.3	Clasificación de usuarios propuestos como TDM y TDP.....	43
4.4	Comparación entre TDM y TDP	44
4.5	Discusión.....	45

CAPÍTULO 5	47
Conclusiones	47
5.1 Trabajos Futuros	49
Referencias	50

Índice Tablas

Tabla 1. Estado del arte. Trabajos más relevantes usando la base de datos CLPsych 2015.....	17
Tabla 2. Ejemplo de las palabras y sus anotaciones en el léxico NRC.	24
Tabla 3. Características de la base de datos CLPsych 2015.....	32
Tabla 4. Ejemplo del preprocesamiento de Corpus.	34
Tabla 5. Resultados de los modelos propuestos y la comparación con los resultados del estado del arte	44
Tabla 6. Resultados de comparar los grupos de TDP y TDM usando tres síntomas presentes en pacientes con TDP	45

Índice Figuras

Figura 1. Ejemplo de matriz de co-ocurrencia.....	25
Figura 2. Operación realizada en una Red Neuronal Convolutiva. Un kernel de tamaño 3 recorre los datos de entrada y crea un mapa de características.	26
Figura 3. Operación Global Max Pooling. Esta operación selecciona el valor máximo de cada mapa de características.....	26
Figura 4. Ejemplo de capa Densa.....	27
Figura 5. Ejemplo de la operación Dropout. En este caso, una neurona está desactivada en la capa densa.....	27
Figura 6. Depresión crónica a consecuencia varios episodios depresivos. Los episodios depresivos tienen diferentes duraciones desde periodos que van desde las dos semanas hasta meses. También el periodo libre de síntomas puede tener una duración variable, desde semanas hasta meses.	30
Figura 7. Comportamiento de la sintomatología de pacientes con TDP a lo largo del tiempo. En esta patología existe una sintomatología más leve que la experimentada en un episodio depresivo pero un periodo más grande de tiempo (al menos dos años).....	31
Figura 8. Metodología propuesta.	33
Figura 9. La puntuación de tristeza por cada día se calcula promediando la puntuación de tristeza de cada mensaje.....	35
Figura 10. Pseudocódigo para obtener la ventana de 14 días con mayor puntuación sintomática.	37
Figura 11. Modelos CNN para clasificar a los usuarios.	38
Figura 12. Curva de entrenamiento para los datos de entrenamiento y validación utilizando la exactitud como métrica para el modelo TDM propuesto.....	42
Figura 13. Curva de entrenamiento para los datos de entrenamiento y validación utilizando la pérdida como métrica para el modelo TDM propuesto.....	42
Figura 14. Curva de entrenamiento para los datos de entrenamiento y validación utilizando la exactitud como métrica para el modelo TDP propuesto.....	43
Figura 15. Curva de entrenamiento para los datos de entrenamiento y validación utilizando la pérdida como métrica para el modelo TDP propuesto.	43

Introducción

Un trastorno mental puede definirse como un “síndrome caracterizado por una alteración clínicamente significativa en la cognición, la regulación emocional o el comportamiento de un individuo que refleja una disfunción en los procesos psicológicos, biológicos o de desarrollo que subyacen al funcionamiento mental” [1]. Aunque la salud mental es un problema global, son los países desarrollados y las economías emergentes los que presentan un mayor número de casos de enfermedades mentales. Asimismo, la depresión afecta a más personas que otros trastornos mentales en América Latina. De la misma manera, en casos avanzados la depresión puede provocar intentos de suicidio o suicidio [2]. Desafortunadamente, la familia y la comunidad de quienes padecen esta enfermedad también se ven afectados.

En lo que refiere a la cifras, de acuerdo a la Organización Mundial de la Salud [2], para el año 2021, la depresión afectó a más de 280 millones de personas en el mundo. Para el caso de México, conforme a datos del Instituto Nacional de Estadística y Geografía [3], el 15.4% de los habitantes mayores de 12 años tienen algún nivel de depresión ocasional. Asimismo, alrededor de 700,000 personas mueren cada año en todo el mundo a causa de los suicidios, siendo las enfermedades mentales la principal causa. Además, los suicidios representan la cuarta causa de muerte en la población de 15 a 29 años [2].

Considerando las cifras antes mencionadas, el estudio y análisis de enfermedades mentales tales como la depresión son de vital importancia para la sociedad. Por otro lado, el uso cada vez mayor de plataformas de redes sociales como Twitter, Facebook e Instagram, entre otras, ha abierto la pregunta de que si los datos recopilados pueden

usarse para encontrar tendencias en muchos aspectos de la salud humana. Por ejemplo, Twitter registra palabras y frases que reflejan estados emocionales, sentimientos y sucesos de la vida diaria. Estas condiciones se comparten con el público y tienen un registro de tiempo cuando surgen [4]. Esta información podría ser valiosa para la investigación de análisis de sentimientos y la detección de enfermedades mentales como, por ejemplo, la depresión.

La dificultad de detectar signos de enfermedad mental utilizando datos de redes sociales se ha analizado anteriormente [5]. Además, se han generado esfuerzos para etiquetar e identificar emociones y comportamientos en busca de identificar patrones que permita clasificar entre usuarios sanos y con depresión [6].

Una característica importante a examinar en la búsqueda de trastornos mentales es el análisis de los síntomas y la temporalidad para identificar sujetos con patologías [1]. El período en que una enfermedad mental está presente es esencial, ya que los médicos pueden diagnosticar su gravedad en función de la progresión de la enfermedad. Al analizar usuarios de las redes sociales para la detección de enfermedades, los enfoques de aprendizaje automático deben considerar el uso de las mismas funciones que usan los médicos. Solo así el software de diagnóstico asistido por ordenador puede ser más confiable.

Los algoritmos de aprendizaje automático que buscan rasgos de depresión en usuarios de la plataforma Twitter se basan en mensajes etiquetados de personas con depresión y sin la enfermedad [7]. Durante el entrenamiento, los modelos aprenden características para las dos clases. Aunque este enfoque parece tener resultados positivos, no está claro para un médico por qué el algoritmo tomó una decisión. La falta de explicabilidad en los modelos de aprendizaje automático hace que los profesionales de la salud desconfíen de tales algoritmos [8]. Tratando de hacer que los algoritmos sean más explicables, algunos trabajos analizan los síntomas de la depresión, como las emociones así como otras características sintomatológicas como la fatiga o los trastornos del sueño [9].

Trabajando en este problema, se propone un modelo convolucional para detectar signos de depresión en usuarios de redes sociales utilizando características adquiridas de texto y metadatos. Los datos usados en este trabajo pertenecen a la tarea compartida de la Lingüística Computacional y Psicología clínica (CLPsych) en su edición 2015 [10]. Esta base de datos cuenta con usuarios depresivos previamente caracterizados así como usuarios de control.

En el presente trabajo, se desarrolla un modelo de aprendizaje automático que clasifica a los usuarios deprimidos, considerando no solo síntomas característicos de la enfermedad, sino también los factores temporales de la patología, tratando de imitar lo que un médico haría para diagnosticar la enfermedad. Para ello, se usó la base de datos CLPsych 2015 [10], que contiene información temporal de tweets de usuarios, para diagnosticar dos de los tipos más comunes de depresión: Trastorno Depresivo Mayor (MDD) y Trastorno Depresivo Persistente (PDD).

Se desarrollaron dos modelos de aprendizaje automático que incorporan las características de cada trastorno mental para clasificar a los usuarios. Siguiendo este enfoque, buscamos las métricas obtenidas cuando los datos se clasificaron después de dividirlos en los dos grupos propuestos en comparación a cuando se clasificaron como un solo conjunto. Finalmente, se analizaron características propias de pacientes con Trastorno Depresivo Persistente entre ambos grupos, como la mayor frecuencia de tweets en tiempo pasado y futuro o palabras clave referente a situaciones sociales [11] con la intención de observar diferencia y similitudes entre ambos grupos.

1.1 Descripción del Problema

Las enfermedades mentales son aún un estigma social. Asimismo, los tratamientos para estas patologías requieren de atención médica especializada a la cual no toda la población tiene acceso debido a situaciones relacionadas con cuestiones socio-económicas del paciente, el estigma social y la falta de atención hacia estas enfermedades por parte del sector salud [12]. Además, habitualmente las personas se

atienden cuando están en una fase crónica de la enfermedad, por ello, detectar la enfermedad en los primeros episodios ayuda no sólo a que el tratamiento sea más breve sino además exitoso. Por otro lado, la desatención de estas patologías puede provocar severas consecuencias que van desde la discapacidad hasta intentos de suicidio o suicidios [13].

Con la intención de generar estrategias para la prevención e identificación de posibles personas deprimidas en los primeros estados (los cuales tienen un mejor pronóstico que los pacientes crónicos) han surgido investigaciones basadas en identificar rasgos asociados a la depresión en usuarios de redes sociales. La razón de utilizar datos de redes sociales es debido a que muchos usuarios deciden compartir sucesos y emociones en sus redes sociales que pueden brindar pistas sobre un posible padecimiento mental como la depresión. La Inteligencia Artificial (IA) ha tenido aproximaciones basadas en herramientas de Procesamiento de Lenguaje Natural (PLN) y Aprendizaje Automático (ML). No obstante, existen aún retos que superar como la identificación de características distintivas de enfermedades mentales similares, la explicabilidad de algunos modelos propuestos [9], el análisis temporal en dichos usuarios [14], la replicabilidad de modelos propuestos [15], la mejora en la calidad de la obtención de las bases de datos [16], el pre-procesamiento de datos de redes sociales, entre otros.

A lo largo de este trabajo se pretende abordar el problema del análisis temporal de algunos síntomas presentes en pacientes con depresión utilizando la información proporcionada por los metadatos de los tweets para, posteriormente, realizar una primera aproximación al análisis de las variantes de una enfermedad, como el caso de los dos tipos de depresión más comunes: Trastorno Depresivo Mayor y Trastorno Depresivo Persistente. Este es un primer acercamiento ya que en la literatura se observa principalmente el análisis únicamente del Trastorno Depresivo Mayor [7].

1.2 Justificación.

De acuerdo con la OMS [2], la depresión es la principal causa de discapacidad en

la población mundial y contribuye de forma importante a la carga mundial de morbilidad . Además, el grupo poblacional que tiene un mayor uso de redes sociales es el mismo grupo que es afectado por enfermedades mentales y suicidios, es decir, jóvenes entre 15 y 29 años [17]. Por ello, la detección de rasgos asociados a la depresión mediante el análisis del texto y los metadatos obtenidos de las redes sociales surgen como una alternativa para acceder a usuarios afectados con esta enfermedad para que puedan acercarse con especialistas para ser atendidos y con ello, disminuir las consecuencias que habitualmente están asociadas a la cronificación del padecimiento y a la múltiples recaídas de los pacientes.

Hipótesis

La clasificación de usuarios de la base de datos CLPsych 2015 usando criterios temporales y considerando a los pacientes bajo los dos rasgos depresivos más frecuentes (TDM y TDP) aumentará los resultados del clasificador en comparación con la clasificación de usuarios considerando únicamente criterios de pacientes con TDM.

Objetivos

1.3.1 Objetivo general

Analizar a los usuarios de la base de datos CLPsych 2015 considerando los dos tipos más comunes de la depresión mediante el análisis de emociones, temporalidad y sintomatología propias de cada patología usando herramientas de Inteligencia Artificial para mejorar la clasificación de usuarios

1.3.2 Objetivos específicos

- Desarrollar un modelo basado en características temporales, emocionales y sintomatológicas para dividir la base de datos CLPsych 2015 en dos grupos.
- Desarrollar un modelo de aprendizaje profundo para clasificar a los usuarios considerando los grupos propuestos
- Comparar entre ambos grupos para observar diferencias y similitudes entre estos dos conjuntos de usuarios.

Antecedentes

2.1 Antecedentes

Para analizar los rasgos asociados a la depresión en usuarios de redes sociales es necesario emplear distintas técnicas que permitan examinar de la mejor manera el contenido de los mensajes o el comportamiento de los usuarios. En base a ello, se destaca en la literatura algunos trabajos como el elaborado por Oxman *et al.* [18], quienes demostraron que el análisis lingüístico del habla podría ayudar a clasificar pacientes dentro de grupos de diagnóstico que sufren depresión y paranoia.

La investigación de la depresión mediante la recolección de información tanto en redes sociales como en otros sitios de internet se hizo más popular a medida que aumentaba el número de personas que compartían información en estos sitios. Por ejemplo, en la literatura se observa la elaboración de marcadores lingüísticos de personas con y sin depresión, recolectados a través de un foro en internet [19]. En este trabajo el autor empleó el software LIWC (Consulta Lingüística y Recuento de Palabras) [20] como herramienta para observar las diferencias en el vocabulario entre ambos grupos, lo que mostró que las personas depresivas usaban más pronombres singulares en primera persona.

Para el año 2010 empezaron a surgir trabajos para analizar enfermedades usando información de redes sociales, como el caso de Chew *et al.* [21], quienes presentaron una investigación donde se empleaba como base de datos *tweets* recolectados durante la pandemia H1N1. En este trabajo, se implementa el análisis de sentimientos a dos millones de *tweets* que contenían las palabras clave como “gripe porcina” o “H1N1”.

En este artículo, los autores mostraron que el análisis de sentimientos podría ser logrado permitiendo monitorear una gran cantidad de datos en tiempo mostrando a las redes sociales como una herramienta para obtener información sobre la proliferación y el estado de enfermedades como la influenza. Por otro lado, Aaron [5] aplicó técnicas clásicas de aprendizaje automático *como máquinas de vectores de soporte (SVM), regresión lineal y el clasificador ingenuo de Bayes* para clasificar *tweets* usando tasas de incidencia de la gripe (resfriado común), conduciendo a coeficientes de correlación de aproximadamente 0.95 para la técnica de regresión lineal entre las estimaciones obtenidas y las estadísticas de salud reportadas por los Estados Unidos. Dichos resultados muestran que la correcta aplicación de las técnicas de aprendizaje automático a información de redes sociales puede conducir a resultados prometedores en el análisis de enfermedades.

2.2 Estado del Arte

Con el aumento en la popularidad de las redes sociales surgieron más investigaciones orientadas a analizar información de redes sociales en busca de información relacionada a enfermedades mediante algoritmos de IA. Fue así que el año 2013 surgió el primer gran trabajo orientado a identificar características entre usuarios sanos y depresivos. En este artículo De Choudhury *et al.* [22], clasificaron *tweets* de pacientes diagnosticados con depresión y desarrollaron un modelo para conocer la proliferación de la depresión en las poblaciones. Los autores usaron SVM (Máquinas de Vecotores de Soporte) como clasificador y obtuvieron un 70% de exactitud y una precisión de 82%, mostrando que las redes sociales podrían ser una herramienta para analizar depresión en las poblaciones.

En 2014, Coppersmith *et al.* [23], obtuvieron datos de la red social Twitter para analizar cuatro enfermedades: depresión, Trastorno de Estrés Post-Traumático (TEPT), Trastorno Bipolar y Trastorno Afectivo Estacional (TAE). En este trabajo los autores

desarrollan su base de datos mediante el autodiagnóstico de los usuarios, es decir, los usuarios etiquetados con la enfermedad analizada eran sujetos que explícitamente habían sido diagnosticados con la enfermedad. Este método de construcción de base de datos tiene ventajas como la facilidad y el bajo costo para recopilar los datos, no obstante pueden existir sesgos debido a que son los usuarios quienes mencionaron haber sido diagnosticados con la enfermedad.

Para el siguiente año se desarrolla la tarea compartida de la CLPsych (Lingüística Computacional y Psicología Clínica) en su edición 2015 [10], donde participaron varios grupos de investigadores que tenían como misión modelar características relevantes del lenguaje observadas por usuarios de la red social Twitter que padecían enfermedades mentales tales como la depresión y el TEPT (Trastorno de Estrés Post-Traumático). A continuación se muestran algunos puntos interesantes de estas investigaciones.

En 2015, Resnik *et al.* [24], emplearon modelos probabilísticos supervisados en el análisis lingüístico para detectar signos de depresión. Los autores emplean las técnicas de LDA (Modelo de Asignación Latente de Dirichlet) supervisado y ANCHOR (algoritmo propuesto por Arora *et al.* [25], que identifica una sola palabra “ancla” asociada con cada tema) supervisado. Asimismo, los autores emplearon la base de datos CLPsych 2015, misma que contiene 477 usuarios que habían sido clínicamente diagnosticados con depresión. A manera de conclusión, se muestra que el modelo que presentó mejor desempeño con lenguaje relacionado a depresión, fue el algoritmo LDA. Para analizar los datos, los autores acomodaron los tweets por semana, y cada conjunto semanal contenía la etiqueta del autor. Para clasificar los datos, se empleó Máquina de Vectores de Soporte y se obtuvo un resultado de 0.86 usando la métrica de área bajo la curva (AUC).

En el año 2016, Nadeem *et al.* [26], buscan analizar la depresión en usuarios de redes sociales únicamente como un problema de clasificación de texto, es decir, no utilizaron metadatos de los tweets para clasificar. Utilizando la base de datos CLPsych 2015, un enfoque de bolsa de palabras y distintos algoritmos clásicos de aprendizaje máquina

como Regresión logística, Ingenuo de Bayes y Máquina de Soporte de Vectores, lograron un valor de 0.86 de exactitud y una precisión de 0.86 para los algoritmos de Ingenuo de Bayes y Regresión Logística respectivamente.

En este mismo año, Mowery *et al.* [27], propone un método distinto enfocado a analizar tres síntomas de pacientes deprimidos: estado de ánimo deprimido, problemas de sueño y fatiga. Para analizar dichos síntomas, los autores proponen varias características como el análisis de sentimientos, análisis de emociones y LIWC (Consulta Lingüística y Conteo de Palabras). Los mejores resultados de los experimentos se obtuvieron para el clasificador de Bosques aleatorios y el síntoma de fatiga.

También en 2016, Tao *et al.* [28], presentan una investigación donde emplearon las técnicas de análisis de sentimientos para examinar las contribuciones de los usuarios de la red social Twitter, en busca de detectar aspectos depresivos en el lenguaje. En este trabajo, los autores desarrollaron un prototipo que sirve para monitorear los *tweets* de los usuarios y notifica a una persona designada en caso de que se observen rasgos conductuales o emocionales relacionados a la depresión. Los autores mencionan que este tipo de herramientas pueden ofrecer a los trabajadores sociales o padres de familia la capacidad de acceder a personas potencialmente depresivas para tomar acciones preventivas.

Para el siguiente año, Deshpande y Rao [29] aplicaron el procesamiento del lenguaje natural a *tweets* para analizar las emociones centradas en la depresión. Su conjunto de datos eran 10,000 *tweets* obtenidos a partir de una lista de palabras seleccionadas para detectar tendencias de depresión como “deprimido”, “desesperado”, “suicidio”, etc. Después, para clasificar los *tweets* utilizaron los algoritmos de SVM (Máquina de Soporte Vectorial) y *Naive Bayes*. La precisión obtenida al realizar la clasificación fue del 83% para *Naive Bayes* y 79% para SVM. A manera de conclusión, los autores mencionan que un correcto pre-procesamiento de los datos y una mejor selección de las palabras clave podrían aumentar la eficiencia del modelo.

En ese mismo año, Shen *et al.* [30], generaron un grupo de seis conjuntos de datos de

usuarios depresivos basados en características como emociones negativas y comportamientos característicos de personas que sufren depresión. En base a estos grupos, los autores proponen un modelo de aprendizaje basado en un diccionario depresivo multimodal para detectar a usuarios deprimidos en la red social Twitter. Además, al analizar los datos obtenidos, los autores observaron rasgos interesantes relacionados a las personas con depresión como la hora de publicación de los *tweets* o el número de *tweets* que comparten con sucesos relacionados a su vida.

También en 2017, Singh y Wang [31] crearon un conjunto de datos de varias páginas de Twitter relacionadas a estos temas y los etiquetaron con una puntuación de 0 para los tweets de control y 1 para tweets depresivos. Posteriormente, construyeron modelos de aprendizaje profundo RNN (Redes Neuronales Recurrentes), GRU (Unidades Recurrentes Cerradas) y CNN (Redes Neuronales Convolucionales) para generar predicciones en el conjunto de datos generado. Para esos modelos, los autores examinaron los efectos basados en características (hora de publicación, tipo de contenido) contra los basados en palabras (vocabulario depresivo). Al final, después de comparar los resultados, los autores comentan que los mejores modelos fueron el GRU basados en palabras y el CNN también basado en palabras con una efectividad de 98% y 97% respectivamente. A manera de conclusión los usuarios mencionan la importancia de tener un conjunto de datos de personas que han sido diagnosticados con depresión, ya que este tipo de información les permitiría analizar otras tendencias como el hecho de que las personas con depresión no siempre están “tristes”.

Para el año 2018, Biradar y Torad [32] proponen un sistema que utiliza *tweets* como fuente de datos y emplean el software SentiStrenght para efectuar análisis de sentimientos al conjunto de datos. Una vez realizado el análisis de sentimientos, se destina un conjunto de datos como entrenamiento para un modelo de red neuronal del tipo perceptrón multicapa que clasifica los *tweets* dentro de las categorías deprimidas y no deprimidas. Empleando un set de datos de 6,908 *tweets* lograron obtener una precisión de 79.46%.

También en 2018, Islam *et al.* [33], deciden implementar datos (comentarios) de la red social Facebook y la técnica de clasificación KNN (K-vecinos más cercanos) para detectar síntomas de la depresión en usuarios. En este estudio, se consideran cuatro tipos de factores: proceso emocional, estilo lingüístico, proceso temporal y todas las características juntas. Por ello, desarrollan un modelo que analiza cada factor de forma independiente y conjunta. Las variables emocionales fueron: positivas, negativas, tristeza, enojo, ansiedad; mientras que para las categorías temporales utilizaron: enfoque presente, enfoque pasado y enfoque futuro; y por último, las dimensiones lingüísticas fueron: artículos, preposiciones, verbos auxiliares, adverbios, conjunciones, pronombres, verbos y negaciones. Finalmente, los autores muestran que los resultados varían entre 60 y 70% de precisión. A manera de conclusión, se menciona que sería útil emplear otra técnica para extraer mejor las características emocionales ya que no se obtuvieron buenos resultados en esta área.

En este mismo año, Orabi *et al.* [34], utiliza la base de datos CLPsych 2015 y probaron cuatro modelos de aprendizaje profundo (3 modelos convolucionales y uno recurrente). Asimismo, los autores proponen un embebido de palabras para analizar textos de usuarios depresivos. Los mejores resultados se obtienen para los modelos convolucionales implementando el embebido de palabras propuesto por los autores, logrando un 0.87 de exactitud. Logrando así, los mejores resultados para esta base de datos.

En 2019 Leis *et al.* [35], elaboraron una investigación para detectar depresión basándose en el comportamiento y en las características lingüísticas de usuarios de la red social *Twitter*. Su trabajo se dividió en dos partes; la primera, consistió en elaborar tres conjuntos de datos: el primer conjunto se basó en datos de 90 usuarios que explícitamente expresaron que tenían la enfermedad; el segundo grupo consistió en un conjunto de datos que contenían alguna de las palabras que los autores proponen como vocabulario depresivo; y el tercer conjunto, consistió en un grupo de datos de control de

450 usuarios seleccionados aleatoriamente. La segunda parte de la investigación involucró el análisis y la comparación de los tres conjuntos de datos, lo que brindó resultados relacionados a la hora de publicación de los *tweets* o a las características del vocabulario empleado. Cabe señalar que los autores recabaron alrededor de 100 millones de *tweets* y la selección de usuarios con características depresivas fue en base a un conjunto de 20 palabras que por lo que mencionan los autores, son las palabras más relacionadas a la depresión. Es importante mencionar que a diferencia de los trabajos previos, esta investigación se desarrolló con *tweets* en español y no en inglés.

También en 2019, Cacheda *et al.* [36], proponen un modelo para la detección previa de signos de depresión usando características textuales, semánticas y similitudes en la escritura. El conjunto de datos usado fue el *eRisk* 2017 [37]. Un aspecto interesante en este trabajo es que los autores consideraron dos semanas como periodo para dividir las contribuciones de cada usuario para tener un criterio similar al propuesto en clínica. Los mejores resultados se obtuvieron con el clasificador de Bosques aleatorios.

Para el año 2020, también Leis *et al.* [38], desarrollaron un trabajo para conocer los patrones de lenguaje y de comportamiento de usuarios de *Twitter* que había tomado algún medicamento para tratar la depresión. Para ello, seleccionaron un grupo de usuarios que mencionaron alguno de los medicamentos antidepresivos ISRS (Inhibidores selectivos de la receptación de serotonina) incluyendo los 128 nombres genéricos encontrados. Posteriormente, se crearon dos conjuntos de datos; el primero, compuesto por los *tweets* publicados durante los 30 días posteriores a la mención del medicamento depresivo; y el segundo, consistió en el conjunto de *tweets* publicados 90 días antes y después de cualquier *tweet* que mencione un medicamento antidepresivo. Se emplearon 186 usuarios y un total de 668, 842 *tweets*. Su investigación mostró que la cantidad de *tweets* generados por día y el número de palabras por *tweet* de los usuarios cuando estaban en tratamiento fue mayor que cuando estaban en periodo de tratamiento desconocido. Además, en cuanto a las características lingüísticas, los autores encontraron que el porcentaje de pronombres en primera persona del singular

fue mayor cuando los usuarios estaban en tratamiento.

Para el 2021, Zhou *et al.* [39], publicaron un artículo donde estudian la dinámica de la depresión comunitaria a consecuencia de la pandemia de COVID-19 a través del contenido generado por usuarios en Twitter. Se propone un nuevo enfoque basado en características multimodales (características emocionales y de comportamiento) y el método TF-IDF (Término de frecuencia-Frecuencia Inversa de Documento) mismo que expresa la importancia de una palabra para un texto. Con este par de técnicas, los autores detectan polaridades de depresión en el texto de Twitter. Los autores extraen tweets recientemente (enero a mayo de 2020) de usuarios que emanan del estado de Nueva Gales del Sur en Australia. Los resultados muestran que las personas se deprimieron más después del brote de COVID-19. Además, las medidas implementadas por el gobierno, como el bloqueo estatal, también aumentaron los niveles de depresión.

El progreso en el aprendizaje profundo ha facilitado la elaboración de modelos de detección exitosos. Una nueva tendencia en el procesamiento del lenguaje natural (NLP), los modelos de lenguajes pre-entrenados contextualizados, han mostrado un alto rendimiento en varias tareas de procesamiento de texto. Uno de estos modelos se llama BERT [40], que utiliza codificadores de texto basados en transformadores bidireccionales. Debido al buen desempeño de BERT, se han desarrollado muchos modelos de lenguaje pre-entrenados en diferentes dominios como el caso de BERT mental [41]. En este modelo, se analizaron enfermedades tales como la depresión, la ansiedad y el estrés. Para el caso de la depresión se utilizaron la base de datos CLPsych 2015 [10] y eRisk [37]. Aunque se obtuvieron resultados menores a los observados en el estado del arte en lo que refiere a la detección de usuarios depresivos, es preciso mencionar que la reducción de estos resultados pudo ser a causa de que la similitud en algunos síntomas de enfermedades tales como la depresión y la ansiedad.

También en el año 2021, Uban *et al.* [42], realiza un análisis en busca de marcadores lingüísticos de desórdenes a diferentes niveles de lenguaje (contenido, estilo y emociones). Ellos analizan su aproximación en base a estilos cognitivos y emociones,

con la intención de entender lo que es posible conectar entre los estilos cognitivos con la comunicación de las emociones a lo largo del tiempo. En este trabajo se analizaron trastornos como la depresión y la anorexia. Para clasificar a los usuarios se usaron modelos de aprendizaje profundo se observó que el aspecto temporal en los síntomas de las enfermedades mentales manifestados en el lenguaje es un punto importante para entender marcadores de enfermedades mentales con mayor profundidad.

Para el ese mismo año, Lee et al. [9], realizan un trabajo que busca reducir los sesgos de trabajos previos orientados a la falta de explicabilidad y replicabilidad de dichos modelos. Para luchar contra este problema, los usuarios usan micromodelos orientados a detectar desórdenes cognitivos mediante palabras clave y similaridad semántica. Como algoritmo de clasificación usan las Máquinas Explicables de Boosting (EBM), con la intención de desarrollar un modelo con mayor legibilidad que los observados en algoritmos de aprendizaje profundo. Con la intención de probar la replicabilidad de sus resultados en distintas bases de datos, los autores probaron su modelo clasificando usuarios con depresión, TEPT y evaluación de riesgo suicida. En lo referente a la depresión, reportaron un valor de 0.82 en la métrica de AUC.

Para el año 2022, Nguyen et al. [15], trabajaron también en el problema de la mala generalización y la falta de confianza de modelos automatizados para analizar condiciones de salud mental. La aproximación que proponen se basa en los síntomas descritos en PHQ9, que es un cuestionario utilizado por los médicos en el proceso de detección de la depresión. En este trabajo, los autores encontraron que basar el modelo en los síntomas de PHQ9 mejora sustancialmente su capacidad para generalizar datos fuera de distribución en comparación con un enfoque estándar basado en BERT. Asimismo, los autores mencionan que fundamentar las predicciones del modelo en síntomas clínicamente relevantes puede mejorar la capacidad de generalización al tiempo que produce un modelo que es más fácil de inspeccionar.

Las investigaciones orientadas a detectar rasgos asociados a la depresión muestran el uso de diferentes estrategias para extraer características de los usuarios de redes

sociales como la herramienta LIWC o las palabras clave. Así mismo, se han propuesto diferentes estrategias para codificar el texto que van desde métodos simples como el enfoque de bolsa de palabras hasta métodos más elaborados como los embebidos de palabras usados modelos como Word2Vec [43] o BERT. De la misma manera, se han empleado distintos clasificadores siendo los más comúnmente usados las Máquinas de Vectores de Soporte y la Regresión Logística para el caso de los clasificadores clásicos, mientras que en el caso de clasificadores usando aprendizaje profundo, podemos destacar a la arquitectura convolucional como la más usada.

A pesar de que se ha trabajado en los últimos años con problemas como la falta de generalización o la replicabilidad de los modelos, aún existen áreas que no han recibido tanta atención, como lo es el caso del análisis de la temporalidad o la caracterización de los diferentes tipos de depresión. Para el caso de la temporalidad se observan trabajos que han implementado características temporales como la implementación de un periodo de 14 días para dividir los datos de los usuarios, no obstante, no existe una razón sintomatológica para seleccionar el inicio y el final de dicho periodo [24], [36]. Asimismo, la mayor parte de los estudios se orientan a analizar Depresión Mayor [7], dejando a un lado otros tipos de depresiones que también una prevalencia significativa en la sociedad [44]. Trabajando en estos puntos, en el presente trabajo se realiza una primera aproximación de un análisis temporal para separar la base de datos CLPsych 2015 considerando los dos tipos de depresión más comunes: Trastorno Depresivo Mayor y el Trastorno Depresivo Persistente.

Dado que en el presente trabajo se usará la base de datos CLPsych 2015 [10], en la Tabla 1 se presentan los resultados más relevantes utilizando este conjunto de datos.

Tabla 1. Estado del arte. Trabajos más relevantes usando la base de datos CLPsych 2015

Artículo	Características	Modelo	Accuracy	Recall	Precision	F1-score	AUC	Explicable
Preotiuc-Pietro et al. Mental Illness Detection at the World Well-Being Project for the CLPsych 2015 Shared Task. 2015 [45].	Metadatos, unigramas	Ensamble de clasificadores y LR			0.867		0.904	✗
Resnik et al. The University of Maryland CLPsych 2015 Shared Task System. 2015 [24].	TF-IDF	SVM					0.86	✗
Orabi et al. Deep Learning for Depression Detection of Twitter Users. 2018 [34]	Embebidos de palabras	CNN	0.879	0.87	0.874	0.869		✗
Harrigian, et al. Do Models of Mental Health Based on Social Media Data Generalize? 2020 [46]	GloVe y LIWC	LR				0.78	0.79	✓
Lee, et al. Micromodel for Efficient, Explainable, and Reusable Systems: A Case Study on Mental Health.2021 [9]	N-gramas Consulta BERT	EBM					0.82	✓

2.3 Características en el lenguaje de pacientes con depresión

Los estudios sobre el efecto del estado mental en el lenguaje utilizado por una persona ya han mostrado varias conexiones, comenzando con observaciones de usos más frecuentes de pronombres personales singulares [47], [48]. Estudios más recientes encontraron, por ejemplo, un uso elevado de la palabra "yo" en particular y más palabras de emociones negativas en ensayos de estudiantes universitarios deprimidos [49], más verbos en tiempo pasado y pronombres en general hablados por pacientes rusos con depresión [50] y un uso más frecuente de palabras absolutistas (por ejemplo, absolutamente, completamente, todo, nada) en foros relacionados con la depresión, la ansiedad o la ideación suicida que en foros no relacionados o foros sobre asma, diabetes y cáncer [51].

En cuanto al comportamiento se ha observado principalmente que los usuarios depresivos tienden a tener más actividad nocturna [30], [35], [52].

Marco Teórico

2.4 Procesamiento de lenguaje Natural

Definimos el Procesamiento de Lenguaje Natural (PLN) como el campo de conocimiento de la Inteligencia Artificial que estudia la interacción del lenguaje humano con una computadora. Las técnicas de PLN permiten a las computadoras analizar, comprender y derivar significado del lenguaje humano [53]. Dichas técnicas son desarrolladas en base a modelos matemáticos elaborados por lingüistas computacionales e ingenieros, basándose en teorías de N. Chomsky [54].

2.4.1 Lingüística

La lingüística es el estudio científico del lenguaje, incluyendo su gramática, semántica y fonética. La lingüística clásica consiste en idear y evaluar reglas del lenguaje y aunque en se ha realizado un gran progreso en los métodos formales para la sintaxis y la semántica, existen problemas complejos referentes a la comprensión del lenguaje natural. Para tratar de resolver estos problemas, se emplean formalismos matemáticos discretos y teoría del lenguaje natural (por ejemplo, lenguajes formales y teoría de autómatas) para modelar algunos de los problemas de esta área [55].

2.4.2 Lingüística Computacional

La lingüística computacional es el estudio de la lingüística utilizando las herramientas de la informática. La gran cantidad de datos y los avances computacionales referentes al procesamiento de los mismos han permitido explorar puntos interesantes a partir de los grandes conjuntos de datos de texto al escribir y ejecutar software.

Uno de los grandes avances en esta área se dio gracias al enfoque estadístico que hasta el día de hoy tiene gran repercusión en el campo [56]. Dicho enfoque surgió gracias a los avances en los métodos estadísticos y al aprendizaje automático que brindaron mejores resultados, velocidad y solidez.

Al igual que el análisis desarrollado en el lenguaje natural, un sistema de PLN se sustenta en los niveles de análisis definidos como fonología, morfología, sintaxis, semántica y pragmática [57]. A continuación se describen brevemente cada uno de ellos.

- Análisis Fonológico: Implica analizar los sistemas de sonido de las lenguas.
- Análisis Morfológico: Refiere al análisis de una palabra basado en las partes significativas que contiene. Algunas palabras no se pueden descomponer en múltiples partes significativas, pero muchas palabras se componen de más de

una unidad significativa. La unidad más pequeña de significado en una palabra se llama morfema.

- **Análisis Sintáctico:** Se define como el análisis que nos dice el significado lógico de ciertas oraciones dadas o partes de esas oraciones considerando las reglas de la gramática para definir el significado lógico y la corrección de las oraciones.
- **Análisis Semántico:** Consiste en brindarle significado a las palabras o a una oración.
- **Análisis Pragmático:** Este análisis implica interpretar el discurso considerando el contenido e intencionalidad. Por ejemplo, en un discurso entre dos personas este análisis implicaría considerar factores contribuyentes como el contexto, los diálogos anteriores, los estados mentales de los individuos, etc.

Mediante estos distintos niveles de análisis el PLN se acerca a la comprensión de los medios humanos de comunicación.

2.4.3 Actividades comúnmente desarrolladas en proyectos de PLN en IA

Existen distintos procesos que se llevan a cabo a lo largo de un proyecto de PLN. A continuación, se muestran algunos de los pasos más frecuentemente usado en estas actividades:

- **Segmentación de oraciones:** Este paso puede realizarse cuando trabajamos con textos largos ya que consiste en dividir todo el párrafo en diferentes oraciones para una mejor comprensión.
- **Tokenización de palabras:** Este proceso divide la oración en palabras o tokens separados. Esto ayuda a entender el contexto del texto.

- Stemming: Ayuda en el preprocesamiento del texto. El modelo analiza las partes del discurso para descubrir de qué está hablando la oración. Este proceso normaliza las palabras en su forma base o raíz. En otras palabras, ayuda a predecir las partes del discurso para cada token.
- Lemmatización: Es una técnica que ayuda a reducir variantes morfológicas en las palabras a raíces comunes de las palabras o lemas. Por ejemplo, “jugando” y “juega” son formas de la palabra “jugar”. Por lo tanto, jugar es el lema de estas palabras.
- Análisis de palabras vacías: El siguiente paso es considerar la importancia de todas y cada una de las palabras en una oración dada. Por ejemplo, en el idioma inglés, algunas palabras aparecen con más frecuencia que otras como "is", "a", "the", "and". Como aparecen con frecuencia, en PLN se marcan como palabras vacías y se eliminan para centrarse en palabras más importantes para la tarea dada.
- Etiquetado de parte del discurso (POS): Las etiquetas POS contienen verbos, adverbios, sustantivos y adjetivos que ayudan a indicar el significado de las palabras de forma gramaticalmente correcta en una oración.

Éstas son algunas de los principales procesos que se llevan a cabo en un proyecto de PLN [57]. Dependiendo de la aplicación se pueden usar éstas u otras herramientas para lograr el objetivo principal que es dar sentido al texto para realizar ciertas tareas automáticamente, como la revisión ortográfica, la traducción, las herramientas de monitoreo de redes sociales, etc.

2.4.4 Aplicaciones del PLN

Las aplicaciones del PLN se utilizan ampliamente en la actualidad, incluso cuando las personas no son conscientes de ello. Por ejemplo, algunas de las aplicaciones más utilizadas son la traducción automática, los filtros de spam, las sugerencias de

autocompletar (en teléfonos celulares y motores de búsqueda). Algunas otras aplicaciones que no todos usan incluyen aplicaciones de resumen de texto, las técnicas de clasificación de texto que se utilizan para clasificar SMS (Servicio de Mensajes Cortos) o por ejemplo, para clasificar libros de acuerdo a su género [53].

Existen diversas tareas desarrolladas dentro del PLN, algunas de ellas son las siguientes:

- Inferencia: Consiste en predecir si existen relación, contradicción o neutralidad entre dos oraciones.
- Equivalencia semántica: Determinar si dos oraciones son semánticamente equivalentes, es decir, que tienen el mismo significado.
- Pregunta-respuesta: Esta tarea consiste en encontrar en un texto determinado la respuesta correspondiente dada una pregunta.
- Análisis de sentimientos: Implica identificar si el sentimiento que se está expresando en una oración es positivo, negativo o neutral
- Aceptabilidad lingüística: Esta tarea radica en identificar si una determinada oración está gramaticalmente correcta.
- Respuesta correcta: Consiste en determinar la continuación de una oración en base a posibles opciones.

2.4.5 Análisis de emociones en el texto

Otra tarea importante dentro del PLN es el análisis de emociones. Esta tarea es similar al análisis de sentimientos, sin embargo, existen algunas diferencias, ya que en el análisis de sentimientos se monitorean únicamente marcadores positivos, negativos o neutros, mientras que en el análisis de emociones se enfoca en un amplio espectro de emociones y sensibilidades humanas dependiendo el modelo de emociones utilizado [58]. El análisis de emociones en texto es particularmente importante para este trabajo ya que la tristeza es una emoción fundamental en pacientes que sufren depresión [59], por esta razón, se propone desarrollar un modelo que analice esta emoción.

Para realizar análisis de emociones existen distintos tipos de herramientas que van desde métodos sencillos basados en palabras clave como los léxicos hasta enfoques más complejos que involucran modelos pre-entrenados como BERT [58].

Para este trabajo se usó un enfoque basado en palabras llamado NRC [60], el cual consta de 14.182 palabras, asociadas con ocho emociones (ira, miedo, anticipación, confianza, sorpresa, tristeza, alegría y disgusto) y dos sentimientos (negativo y positivo). Creado por anotación manual en una plataforma de *crowdsourcing*. La plataforma usada para la anotación fue CrowdFlower, donde se desarrollaron las anotaciones de hablantes nativos de inglés que residen en los Estados Unidos de América. Para cada emoción, a los anotadores se les presentaron cuatro palabras a la vez (4 tuplas) y se le pidió que seleccionara la palabra que transmite la intensidad de emoción más alta y la palabra que transmite la intensidad de emoción más baja.

Los anotadores eran libres de proporcionar respuestas a tantas tuplas de 4 como quisieran. Alrededor del 5% de los datos fueron anotados internamente (por el autor). Estas preguntas fueron denominadas como preguntas etiqueta, misma que se intercalaron con otras preguntas. Si un anotador se equivoca en una pregunta etiqueta, se le notifica inmediatamente. Si la precisión de un anotador en las preguntas etiqueta cae por debajo del 70%, se le niegan más anotaciones y todas sus anotaciones se descartan. Los puntajes de intensidad se calcularon a partir de las respuestas mediante un procedimiento de conteo simple [61]: para cada elemento, el puntaje es la proporción de veces que se eligió el elemento por tener la mayor intensidad menos la proporción de veces que el elemento fue elegido por tener la menor intensidad. Las puntuaciones van desde -1 (menor intensidad de emoción) a 1 (mayor intensidad de emoción). Dado que el grado de emoción es una escala unipolar, se transformaron linealmente las puntuaciones de -1 a 1 en puntuaciones en el rango de 0 (menor intensidad de emoción) a 1 (máxima intensidad de emoción). La Tabla 2 muestra un ejemplo de las palabras y sus anotaciones en el léxico NRC.

Tabla 2. Ejemplo de las palabras y sus anotaciones en el léxico NRC.

Palabra	Enojo	Palabra	Miedo
outraged	0.964	horror	0.923
hate	0.828	hellish	0.829
violence	0.742	grenade	0.828
volatility	0.687	anguish	0.703

2.4.6 GloVe

Para que una computadora pueda trabajar con el lenguaje natural es necesario convertir las palabras a una representación numérica [62]. Existen distintas maneras de hacerlo, desde enfoques sencillos como la codificación “one-hot” hasta métodos más complejos como word2vec, GloVe [63] o fasttext [64].

Para el presente trabajo se optó por trabajar con el modelo GloVe. El motivo de seleccionar este embebido de palabras es debido a que ha mostrado buenos resultados en tareas similares [65]. Es un algoritmo de aprendizaje no supervisado para representar palabras usando un vector. El entrenamiento de este método se realiza sobre estadísticas globales agregadas de coocurrencia palabra a palabra basada en un corpus, y las representaciones obtenidas muestran subestructuras lineales del espacio vectorial de palabras.

La idea principal de GloVe es obtener la relación entre las palabras de las estadísticas. Para ello se utiliza una matriz de co-ocurrencia que muestra con qué frecuencia un par de palabras en particular aparecen juntas. La Figura 4 muestra un ejemplo de una matriz de co-ocurrencia para dos oraciones: “The fast dog wears no tie” and “The dog in the tie ran fast”.

	dog	fast	tie	in	no	ran	the	wears
dog	0							
fast	2	0						
tie	2	2	0					
in	1	1	1	0				
no	1	1	1	0	0			
ran	1	1	1	1	0	0		
the	3	3	3	2	1	2	1	
wears	1	1	1	1	1	0	1	0

Figura 1. Ejemplo de matriz de co-ocurrencia

La matriz de co-ocurrencia de palabras en la Figura 1 es una matriz cuadrada de tamaño $n \times n$, donde n es el vocabulario (el número de palabras únicas) en el corpus. Cada celda $W_{i,j}$ tiene el número de veces que ambas palabras w_i y w_j aparecen juntas en un documento.

GloVe combina características de métodos de ventana de contexto local como Skip-gram con enfoques de modelos de matriz global como Análisis Semántico Latente (LSA) [66].

2.5 Aprendizaje profundo

2.5.1 CNN

Las Redes Neuronales Convolucionales (CNN) [67] son un tipo de red neuronal que utiliza la operación de convolución en los datos de entrada para identificar características. Esta operación usa un filtro o kernel que se desliza sobre los datos de entrada para detectar patrones. A medida que el kernel pasa sobre dichos valores, se genera una multiplicación entre los valores del kernel y los datos de entrada elemento a elemento, generando un mapa de características con los resultados. La Figura 2 muestra un ejemplo de esta operación. En esta figura, un kernel de tamaño 3 realiza la operación de convolución con los datos de entrada, produciendo un mapa de características.

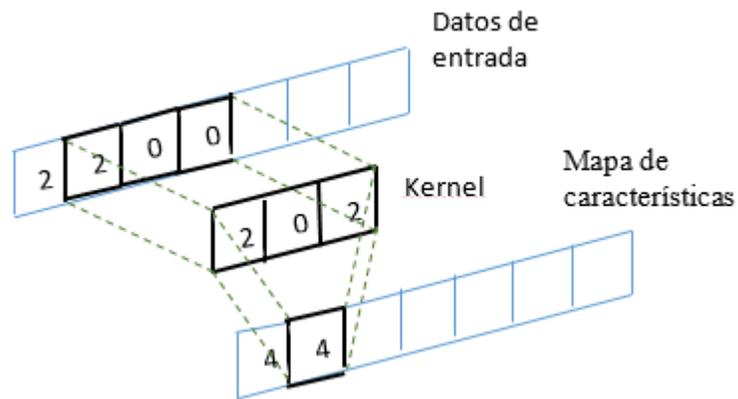


Figura 2. Operación realizada en una Red Neuronal Convolutiva. Un kernel de tamaño 3 recorre los datos de entrada y crea un mapa de características.

2.5.2 Global Max Pooling

Esta operación consiste en obtener el valor máximo del vector de características generado en la operación de convolución. La Figura 3 muestra un ejemplo de esta operación.

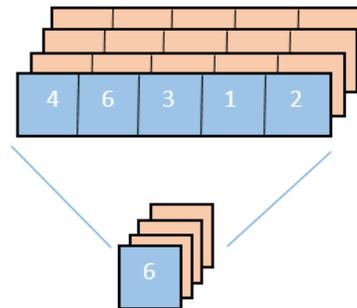


Figura 3. Operación Global Max Pooling. Esta operación selecciona el valor máximo de cada mapa de características.

2.5.3 Capa densa

Esta capa es una red neuronal que está profundamente conectada. En este caso, cada neurona recibe información de todas las neuronas de su capa anterior. La figura 4 muestra un ejemplo de este tipo de red neuronal.

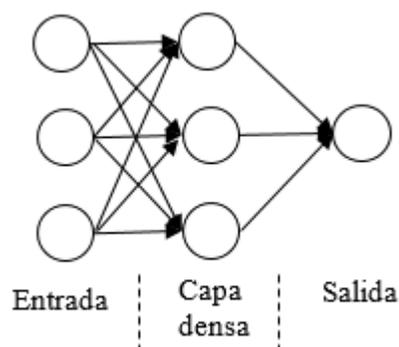


Figura 4. Ejemplo de capa Densa.

2.5.4 Dropout

Cuando tenemos un conjunto de datos de entrenamiento con pocas muestras, es probable que las redes neuronales de aprendizaje profundo se sobreajusten. Una alternativa para combatir el sobreajuste es la operación de abandono o *dropout*. El término "dropout" se refiere a la eliminación de unidades (ocultas y visibles) en una red neuronal [68]. Esta operación se puede utilizar en diferentes tipos de capas, como capas convolucionales, capas recurrentes, capas densas y capas de entrada. Sin embargo, esta operación no se utiliza en la capa de salida. La Figura 5 muestra un ejemplo de este modelo.

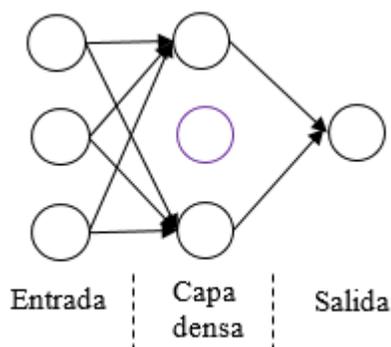


Figura 5. Ejemplo de la operación Dropout. En este caso, una neurona está desactivada en la capa densa.

2.6 Trastornos Depresivos

Identificar los trastornos mentales es una tarea clínica desafiante. Una de las razones es que algunos síntomas de la depresión son similares a otros trastornos mentales [1]. Por ejemplo, la fatiga también podría ser un síntoma de un trastorno de ansiedad. Por este motivo, al desarrollar modelos de IA que busquen identificar rasgos de enfermedades mentales, es crucial considerar los factores temporales y sintomáticos de cada enfermedad para que el algoritmo de detección sea lo más similar a lo que en el área médica.

Hay muchas formas de medir los rasgos de depresión en las personas a través de cuestionarios. Además, existen muchas escalas para estimar la intensidad de la depresión, como la Escala de depresión de Beck (BDI) [69], la Escala de calificación de depresión de Hamilton (HDRS) [70] y la Escala de depresión del Centro de Estudios Epidemiológicos [71].

Por otro lado, otro punto importante en el análisis de los usuarios deprimidos es el análisis de los diferentes tipos de depresión como el Trastorno Depresivo Mayor, el Trastorno Depresivo Persistente, el Trastorno Afectivo Estacional, etc. [1]. El análisis de los diferentes tipos de depresión puede aportar características importantes que ayuden a identificar mejor a los usuarios que padecen esta enfermedad mental. Sin embargo, en este documento solo se abordarán los dos tipos más comunes de depresión: TDM y TDP. Dado que estos dos tipos de depresión son los más comunes, es muy probable que en la base de datos CLPsych 2015 que contiene pacientes deprimidos, haya tantos usuarios con TDM como con TDP. A continuación, se describe brevemente algunas de las características esenciales de estas condiciones.

2.6.1 Trastorno Depresivo Mayor

El trastorno depresivo mayor (TDM) es una enfermedad debilitante que se caracteriza por un estado de ánimo triste, disminución del interés por actividades cotidianas, deterioro de la función cognitiva y síntomas vegetativos, como trastornos del sueño o del apetito. El TDM ocurre aproximadamente dos veces más en mujeres que en

hombres y afecta a uno de cada seis adultos en su vida [72]. De acuerdo al DSM-V los síntomas de esta enfermedad mental son:

1. Sentimientos de tristeza, llanto, vacío o desesperanza
2. Pérdida de interés por actividades que antes resultaban placenteras
3. Pérdida de peso significativa sin dieta ni aumento de peso
4. Insomnio o hipersomnía.
5. Agitación o retraso psicomotor.
6. Fatiga o pérdida de energía.
7. Sentimientos de inutilidad o culpa excesiva o inapropiada.
8. Disminución de la capacidad para pensar, concentrarse o tomar decisiones.
9. Pensamientos recurrentes de muerte.

Aunque muchos episodios depresivos duran más de dos semanas (algunos pueden durar meses), un mínimo de dos semanas con cinco síntomas (considerando los síntomas de sentimientos de tristeza o pérdida de interés) se considera un episodio depresivo.

En algunas ocasiones el paciente puede sufrir de un único episodio depresivo. No obstante, en otros casos el paciente con TDM puede tener más de un episodio depresivo, lo que podría conducir a una depresión crónica (Figura 6). En general, alrededor del 20 al 30% de los trastornos depresivos tienen vías crónicas [73] y están relacionados con resultados de salud más graves y peores consecuencias que la depresión no crónica [74].

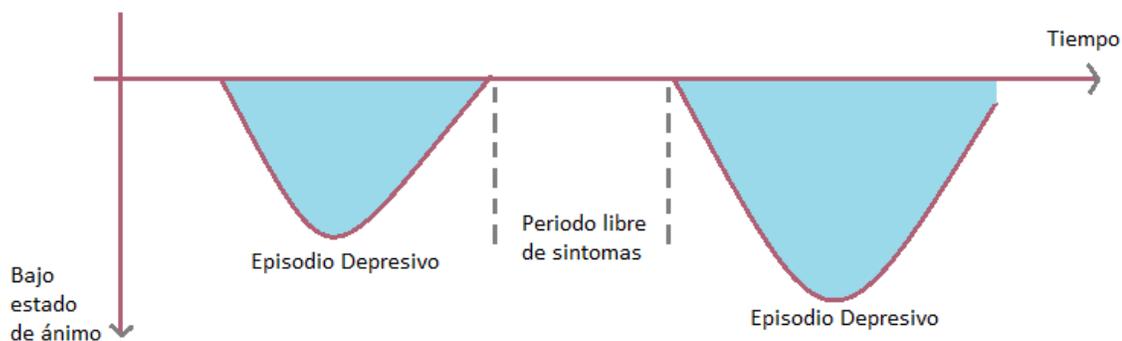


Figura 6. Depresión crónica a consecuencia varios episodios depresivos. Los episodios depresivos tienen diferentes duraciones desde periodos que van desde las dos semanas hasta meses. También el periodo libre de síntomas puede tener una duración variable, desde semanas hasta meses.

2.6.2 Trastorno Depresivo Persistente

El Trastorno Depresivo Persistente (TDP) se caracteriza por una sintomatología depresiva menos grave en comparación con la de un episodio depresivo. En muchos casos, el trastorno suele comenzar en la adolescencia y es persistente, con síntomas que suelen durar al menos dos años y en algunos casos, puede llegar a durar décadas. Algunas personas con este trastorno a veces sufren episodios depresivos adicionales. Según el DSM-V [1], los pacientes con TDP tienen un estado de ánimo deprimido la mayor parte del día, presente más días de los que está ausente, según lo determinado por información subjetiva u observación de otros, durante un mínimo de dos años. Asimismo, debe existir la presencia, durante la depresión, de dos (o más) de los siguientes síntomas:

1. Poco apetito o comer en exceso.
2. Insomnio o hipersomnia.
3. Baja energía o fatiga.
4. Baja autoestima.
5. Falta de concentración o dificultad para tomar decisiones.
6. Sentimientos de desesperanza.

Además, es importante considerar que durante los dos años (un año en niños y adolescentes) de la alteración, el individuo nunca ha estado sin los síntomas mencionados anteriormente durante más de dos meses seguidos.

Como se observa en la figura 7, la sintomatología de los pacientes con TDP es menos severa que en los pacientes con TDM pero tiene una duración mayor. No obstante, es importante que el paciente con esta patología reciba atención médica especializada debido a que los síntomas pueden empeorar o el paciente puede sufrir de episodios depresivos [44].

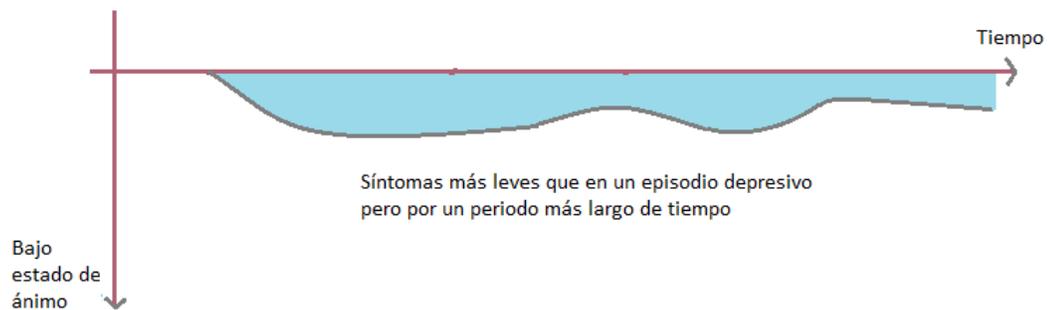


Figura 7. Comportamiento de la sintomatología de pacientes con TDP a lo largo del tiempo. En esta patología existe una sintomatología más leve que la experimentada en un episodio depresivo pero un periodo más grande de tiempo (al menos dos años)

Metodología

3.1 Conjunto de datos

Se utilizó la base de datos CLPsych 2015 [10], que contiene los últimos 3000 tweets públicos de cada usuario. Además, el conjunto de datos contiene la edad y el sexo de los usuarios. El conjunto de entrenamiento tiene 573 individuos etiquetados como usuarios de control, 327 usuarios con depresión y 246 usuarios con Trastorno de Estrés Post-Traumático. Para el conjunto de prueba, se tienen 300 usuarios control, 150 usuarios diagnosticados con depresión y 150 diagnosticados con Trastorno de Estrés Post-Traumático. Además, la base de datos cuenta con la información referente a la edad y sexo del usuario. La Tabla 3 muestra la información completa de la base de datos.

Tabla 3. Características de la base de datos CLPsych 2015.

Característica	Control	Depresión	TEPT
Número de usuarios	873	477	396
Cantidad de Tweets	1,905,373	1,100,349	877,019
Promedio de edad	24.4	21.7	27.9
Genero	74%F y 26%M	80% F y 20% M	67% F y 33% M

3.2 Metodología propuesta

Los modelos de aprendizaje profundo aplicados a analizar texto de usuarios con y sin

depresión han mostrado resultados prometedores [34], [75], [76]. No obstante, estos modelos carecen de explicabilidad en las características que estás extrayendo y clasificando [9]. Por ello, este trabajo se orienta a realizar una aproximación de algunos síntomas presentes en pacientes con depresión mediante la incorporación de un análisis temporal que permitió realizar una propuesta de clasificación de usuarios de la base de datos CLPsych 2015 en base a los dos tipos más comunes de depresión. Para lograr este objetivo, se propuso una metodología en busca de analizar síntomas de la depresión en base a criterios temporales. El primer paso consistió en pre-procesar los datos para disminuir el ruido que pueda existir dada la informalidad de la escritura de las redes sociales. El segundo paso implicó obtener una puntuación sintomática de cada uno de los usuarios, la cual consideró tres síntomas del Trastorno Depresivo Mayor (tristeza, actividad nocturna e ideación suicida). Este Puntaje se obtuvo considerando la ventana de tiempo de 14 días con el puntaje más alto. El tercer paso implicó la separación de la base de datos en dos grupos, es decir, bajo la consideración de estudios clínicos donde se muestra que la depresión crónica comprende alrededor del 33-50% de los pacientes con trastornos depresivos [77]–[79], se realizó una separación del conjunto de datos en dos grupos; es decir, la mitad de usuarios con el valor mayor sintomatológico fue propuesto como usuarios con TDM y la mitad de usuarios con el menor valor sintomatológico fue propuesto como TDP. Por último, el cuarto y último paso consiste en clasificar a los dos grupos propuestos usando un modelo convolucional. A continuación se explican más ampliamente cada uno de los pasos de la propuesta metodológica (Figura 6).

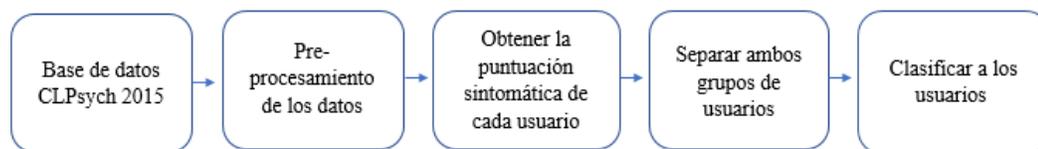


Figura 8. Metodología propuesta.

3.3 Pre-procesamiento

Dado que se trabaja con información de redes sociales, es preciso realizar una limpieza previa para mejorar es aspecto de los datos ya que, en muchas ocasiones, existen elementos en los mensajes que pueden provocar ruido al momento de clasificar a los usuarios. El proceso para la limpieza de los datos consistió en los siguientes puntos:

1. Convertir los emjis y emoticones a palabras mediante el módulo demoji de Python
2. Se normalizaron las palabras a minúsculas
3. Se eliminaron las url's, palabras de paro (excepto pronombres en primera persona) y caracteres no alfanuméricos.
4. Se descartaron las palabras y caracteres repetidos.

Tabla 4. Ejemplo del preprocesamiento de Corpus.

Ejemplo de tweet	
Tweet sin pre-procesar	HeeeeLLO Today I was diagnosed with depression ☹️ which is a very common mental illness: https://www.psychiatry.org/patients-families/depression/what-is-depression
Tweet pre-procesado	hello today i diagnosed depression sad common mental illness

3.4 Obtención de la puntuación sintomática

Con la intención de separar al conjunto de datos en los dos grupos propuestos se generó una puntuación por usuario, misma que consta de tres elementos: puntuación de tristeza, actividad nocturna e ideación suicida. Estos síntomas fueron seleccionados considerando los síntomas de TDM para filtrar este tipo de usuario. A continuación se muestra una breve descripción de cómo se obtuvieron estos valores.

3.4.1 Valor de tristeza

Dado que el estado de ánimo triste es un síntoma distintivo en las personas depresivas y considerando que la temporalidad es un aspecto importante para distinguir entre TDM y TDP, se ordenó el historial de mensajes del usuario por días según la fecha y se analizó cada publicación para detectar el nivel de tristeza utilizando el NRC Léxico de intensidad de emociones [60], el cual que es una lista de palabras en inglés con puntajes de intensidad para ocho emociones: ira, anticipación, disgusto, miedo, alegría, tristeza, sorpresa y confianza. Una vez obtenido el valor de tristeza de cada mensaje, se calcula el valor por día promediando los mensajes de cada día (Figura 7).

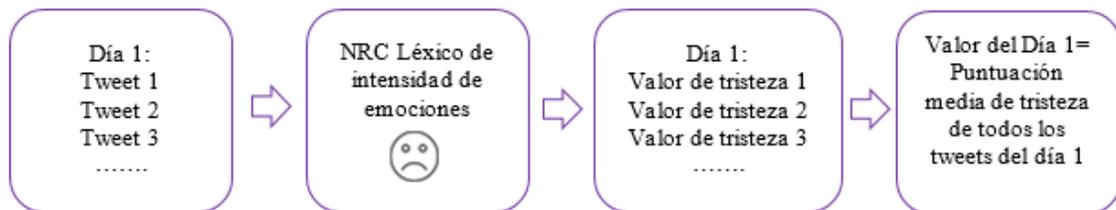


Figura 9. La puntuación de tristeza por cada día se calcula promediando la puntuación de tristeza de cada mensaje.

3.4.2 Actividad Nocturna

Los pacientes con depresión pueden experimentar trastornos del sueño [1]. Por ello, si hay un tweet publicado entre las 0 y las 6 horas, el día se puntúa con 1; si no hay publicaciones durante este tiempo, el día se puntúa como 0.

3.4.3 Palabras relacionadas con pensamientos suicidas

La ideación suicida puede ser un síntoma que se presenta en pacientes con TDM en etapas avanzadas [13]. Por ello, para tratar de distinguir entre estos dos grupos (TDM y TDP), se propusieron las siguientes palabras: suicida, intentos, pánico, autolesiones, matar, morir, odiar, enfermar, llorar y herir. Estas palabras fueron seleccionadas en base

a las proporcionadas por Benton *et al.* [80], El análisis de estas palabras clave en los usuarios implicó contar el número de palabras y dividir dicho valor entre el número de tweets. Este análisis se realizó por cada día y el score final de cada usuario consistió en promediar el valor de todos los días del sujeto.

Los valores de tristeza y suicidio se normalizaron en un rango de 0 a 1. Posteriormente se sumaron los tres valores (tristeza, actividad nocturna e ideación suicida) por día para obtener el valor de la puntuación sintomática.

3.5 Separación del conjunto de datos

Para detectar a los usuarios propuestos como TDM, se seleccionó el periodo de 14 días para cada usuario que tuviera el puntaje sintomático más alto (Figura 8). El motivo de seleccionar este período se debe a la temporalidad indicada en el DSM-V [1] para diagnosticar a un paciente con TDM. Una vez establecidos los 14 días con mayor puntaje de sintomático, se ordenaron los usuarios deprimidos por este valor y se obtuvo el score que dividía el conjunto en dos conjuntos de aproximadamente 50% y 50% por ciento de usuarios depresivos.

En consecuencia, el conjunto de control se dividió utilizando este valor; sin embargo, dado que el conjunto de control tiene una puntuación sintomática más baja en promedio, el conjunto de control no se dividió de manera uniforme. El grupo TDM propuesto (239 usuarios depresivos y 383 usuarios de control) tenía menos usuarios que el grupo TDP propuesto (238 usuarios depresivos y 490 usuarios de control).

```

Entrada= Historial del usuario con puntaje sintomático por día
Salida= Puntuación por usuario de la ventana de 14 días con el mayor
valor

1. For usuario in usuarios:
2.     Max_window=0
3.     Current_window=0
4.     While (Fin de la ventana ≤ Fin del historial del usuario)
5.         Current_window = sum (Puntuación de los 14 días)/ 14
6.         If (Current_window > Max_window):
7.             Max_window = Current_window
8.         end
9.         Desplazar la ventana un día
10.    end
11.    Guardar Max_window de cada usuario
12. end

```

Figura 10. Pseudocódigo para obtener la ventana de 14 días con mayor puntuación sintomática.

3.6 Modelo de clasificación

Se propuso un modelo de aprendizaje profundo para clasificar a los usuarios de ambos trastornos. Este modelo sigue de la capa de embebidos de palabras. En esta capa, se usó el modelo pre-entrenado de embebido de palabras GloVe, y se seleccionó la versión donde cada palabra está representada por un vector de 300 elementos.

Para los usuarios propuestos como TDP, después del embebido de palabras, el modelo tiene una capa convolucional con 300 filtros y un tamaño de kernel de 3. Luego, el modelo tiene una capa densa con 350 unidades. En ambas capas se usó una función de activación relu. Posteriormente, se usó una capa *dropout* con valor de 0.2 para reducir el sobreajuste. Asimismo, esta capa está conectada a la salida que es una neurona con una función de activación sigmoidea.

En el caso de TDM se utilizó un modelo similar y solo se cambió el número de filtros de 300 a 250 y las unidades en la capa densa de 350 a 250 unidades. La Figura 9 muestra el modelo propuesto

Se aplicó una tasa de aprendizaje con valor de 0.007 con un optimizador Adam y se utilizó una función de pérdida de entropía cruzada binaria.

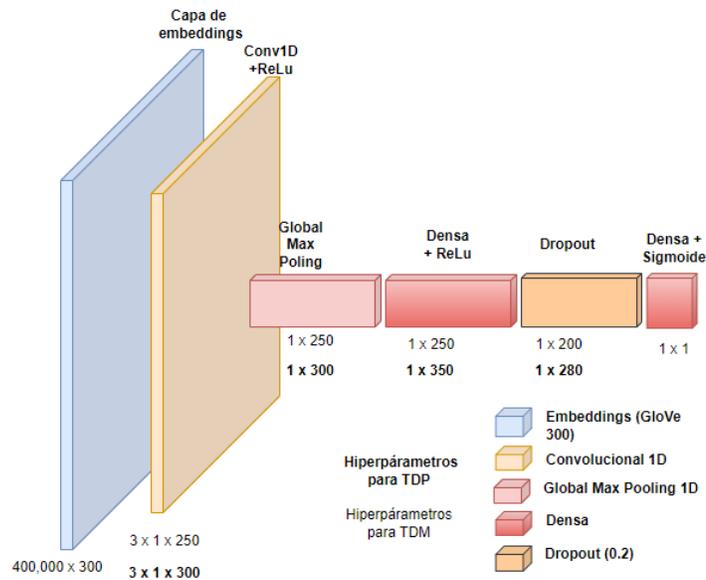


Figura 11. Modelos CNN para clasificar a los usuarios.

3.6.1 Selección de hiperparámetros

Los algoritmos de aprendizaje automático tienen hiperparámetros que permiten que los algoritmos se adapten a conjuntos de datos específicos.

Aunque el impacto de los hiperparámetros puede entenderse en general, es posible que no se conozca su efecto específico en un conjunto de datos y sus interacciones durante el aprendizaje.

Elegir los hiperparámetros correctos para los modelos de machine learning o deep learning es una de las mejores formas de mejorar el rendimiento de los modelos. Para ajustar los hiperparámetros de un modelo existen maneras automatizadas de hacerlo mediante algoritmos como la optimización Bayesiana, así como también se puede ir probando diferentes valores de manera manual [81]. Esta última manera de ajustar los hiperparámetros tiene ventajas como un mayor control sobre el proceso o en casos donde se está investigando o estudiando cómo afectan estos cambios en el modelo tiene

sentido hacerlo manualmente. No obstante, existen desventajas como el hecho de que la sintonización manual es un proceso tedioso, ya que puede haber muchas pruebas y realizar un seguimiento puede resultar costoso y consumir mucho tiempo. Además, este no es un enfoque muy práctico cuando hay muchos hiperparámetros a considerar.

Para seleccionar los hiperparámetros de la red, se seleccionó el método manual con la intención de ir observando los cambios que se observaba a medida que se cambiaban los parámetros. Los valores que se probaron se basaron únicamente en cambiar el valor del número de filtros en la capa convolucional y el número de neuronas en la capa densa. Los valores que se probaron van desde 100 hasta 400 dando pasos de 50 en cada prueba. Los mejores resultados fueron los propuestos para cada uno de los modelos. La futura implementación de modelos automáticos de ajuste de hiperparámetros podría ayudar a mejorar el rendimiento del modelo probando, por ejemplo, valores con un paso menor a 50.

Resultados

4.1 Métricas de evaluación

Para observar el desempeño de la metodología propuesta, se siguieron dos estrategias. La primera, fue el uso de métricas empleadas en estudios similares: exactitud, precisión, sensibilidad y puntuación F1 [82]. Estas métricas surgen a partir de los denominados verdaderos positivos (TP) y verdaderos negativos (TN) que son los valores predichos correctamente por el modelo. Asimismo, también se incluyen los denominados falsos positivos (FP) y falsos negativos (FN) que son las predicciones incorrectas del clasificador [83]. A continuación, se describen cada una de estas métricas:

- **Exactitud:** Expresa las predicciones correctas respecto al total de predicciones. En nuestro caso particular de este trabajo, la Exactitud se refiere a la relación entre las predicciones correctas de pacientes con y sin la patología respecto a todos los casos analizados.

$$Exactitud = \frac{TP+TN}{TP+FP+TN+FN} \quad (1)$$

- **Precisión:** Esta métrica muestra la relación entre los pacientes identificados correctamente con la enfermedad y el total de pacientes identificados con la patología. Es decir, cuando el modelo predice un paciente con la enfermedad, esta métrica nos muestra con qué porcentaje es correcta esta predicción.

$$Precisión = \frac{TP}{TP+FP} \quad (2)$$

- **Sensibilidad:** Esta métrica nos muestra el porcentaje de casos positivos de la

patología que se identificaron correctamente.

$$\text{Sensibilidad} = \frac{TP}{TP+FN} \quad (3)$$

- Puntuación F1: Se define como la media armónica entre sensibilidad y precisión. Esta métrica busca eliminar el sesgo provocado por tener datos desbalanceados, por ello, esta métrica resulta importante en el presente trabajo dado el desbalance que se tiene en el conjunto de datos entre las dos clases (Tabla 1).

$$\text{Puntuación F1} = \frac{2}{\frac{1}{\text{Recall}} + \frac{1}{\text{Precision}}} \quad (4)$$

La interpretación de estas métricas muestra resultados interesantes, por ejemplo, una baja precisión indica más falsos positivos (FP); es decir, los signos de depresión se predicen incorrectamente en los usuarios. Por otro lado, un valor bajo de sensibilidad muestra más falsos negativos (FN); es decir, no se detectan signos de depresión en los usuarios.

4.2 Entrenamiento

Para observar el comportamiento durante el entrenamiento de ambos modelos (TDM y TDP), se graficaron tanto la precisión como la pérdida durante las épocas. El conjunto de datos se dividió en 80% de datos para entrenamiento, 10% para validación y 10% para prueba. En general, el rendimiento del modelo entrenado es satisfactorio. La figura 10 muestra la curva de entrenamiento para los usuarios TDM según la exactitud. En esta figura, se observa la exactitud lograda a medida que aumentan las épocas. La exactitud aumenta a medida que se van incrementando las épocas hasta la número 7, donde comienza la etapa de sobreajuste. Por esta razón, el modelo se detuvo hasta este momento. Además, en la figura 11, se observa que a partir de la época 7, la pérdida en los datos de validación ya no disminuye.



Figura 12. Curva de entrenamiento para los datos de entrenamiento y validación utilizando la exactitud como métrica para el modelo TDM propuesto.

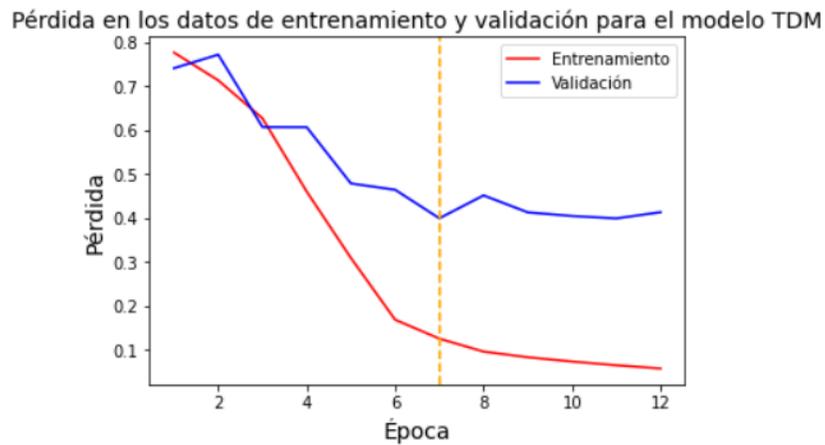


Figura 13. Curva de entrenamiento para los datos de entrenamiento y validación utilizando la pérdida como métrica para el modelo TDM propuesto.

En el caso de los usuarios propuestos como TDP (Figura 12), la exactitud aumenta a medida que se incrementan las épocas hasta llegar a la época número 4 donde la exactitud de la validación se mantiene constante. En consecuencia, ese fue el límite de entrenamiento del modelo. De la misma forma, en la figura 13, a partir de la Época 4 ya no disminuye la pérdida en los datos de validación.

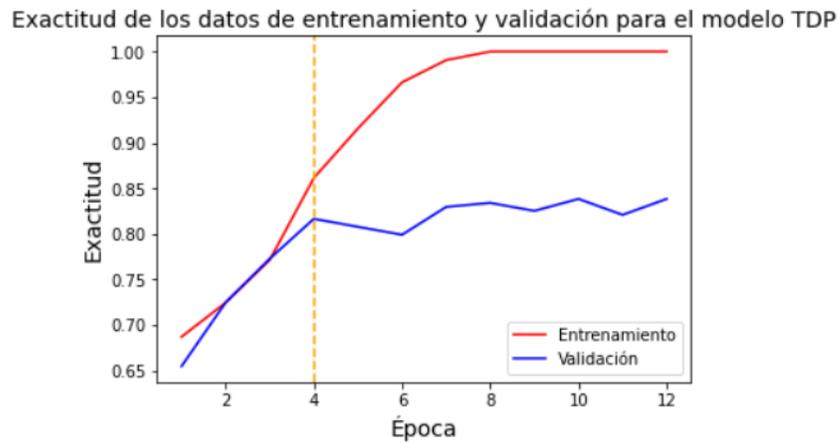


Figura 14. Curva de entrenamiento para los datos de entrenamiento y validación utilizando la exactitud como métrica para el modelo TDP propuesto.

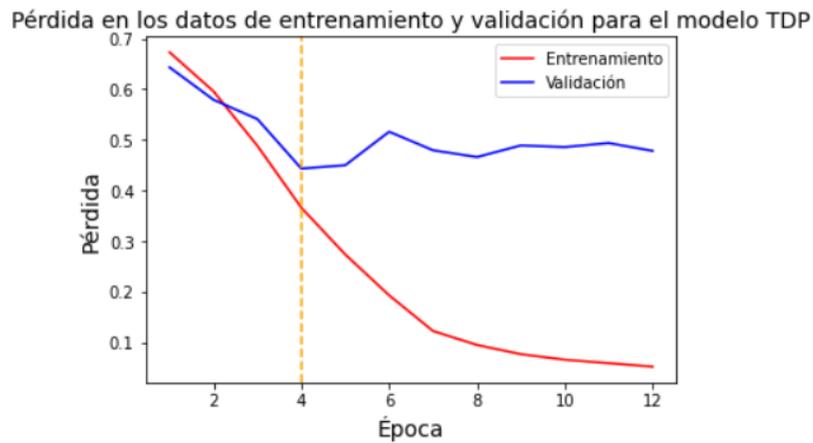


Figura 15. Curva de entrenamiento para los datos de entrenamiento y validación utilizando la pérdida como métrica para el modelo TDP propuesto.

4.3 Clasificación de usuarios propuestos como TDM y TDP

El conjunto de datos (477 usuarios depresivos y 873 de control) se dividió como se propuso anteriormente (TDM y TDP). Después de esta división, para disminuir el sesgo en la medición del rendimiento del modelo de clasificación a consecuencia de la selección del conjunto de datos para entrenamiento y prueba, se seleccionó la validación cruzada estratificada k-fold, seleccionando un valor de $k=5$. Para clasificar a los sujetos, se concatenaron todos los tweets a nivel de usuarios; sin embargo, dado que los vectores resultantes contenían una diferente longitud, se aplicó la normalización de la longitud del

vector del texto mediante relleno y truncamiento para secuencias cortas y largas respectivamente. Se usó una longitud promedio de 30.000 elementos para cada vector.

La primera parte de los resultados consistió en clasificar a los usuarios propuestos como TDM y TDP. Asimismo, se clasificaron ambos grupos juntos (TDM y TDP) utilizando el modelo para usuarios de TDM que es similar al propuesto por Orabi, *et al.* [34]. La Tabla 1 muestra los resultados.

Tabla 5. Resultados de los modelos propuestos y la comparación con los resultados del estado del arte

Modelo	Embebido de Palabras	Exactitud	Precisión	Sensibilidad	Puntuación F1
TDM	GloVe	0.85	0.85	0.85	0.85
TDP	GloVe	0.81	0.81	0.79	0.79
Todo el conjunto de datos (TDM+TDP)	GloVe	0.80	0.80	0.80	0.80
Todo el conjunto de datos (Orabi, <i>et al.</i> [34])	Propuesto por Orabi, <i>et al.</i> [34]	0.88	0.87	0.87	0.87

4.4 Comparación entre TDM y TDP

Con la intención de observar las diferencias entre los dos grupos, ambos grupos depresivos fueron analizados siguiendo tres características importantes observadas en pacientes con Trastorno Depresivo Persistente: sentimientos de culpa y preocupaciones sobre el pasado o el futuro, irritabilidad o enojo excesivo y evitar actividades sociales [1], [11]. Para intentar captar estos posibles síntomas en estos pacientes se propusieron algunas alternativas. Para el caso del primer síntoma, los tweets se analizaron con el etiquetador gramatical de NLTK. De esta manera, de acuerdo a la categoría de cada palabra se seleccionó si se trataba de una oración en presente, pasado o futuro. Así, el total de oraciones en tiempo pasado o futuro se dividió por el total de tweets.

El segundo síntoma consistió en analizar el enojo de la misma forma que se obtuvo la puntuación de tristeza; es decir, se obtuvo el puntaje de enojo por día. No obstante, en este caso se promediaron todos los días del usuario y se obtuvo un valor promedio por usuario. Finalmente, el tercer síntoma se analizó buscando palabras clave como popular, familia, amigos, gente, grupo, fiesta y social. En este punto, se usó el mismo procedimiento que el empleado para obtener los puntajes de las palabras relacionadas a obtener la ideación suicida. La Tabla 2 muestra los resultados de dividir los valores de TDM entre los valores de TDP.

Tabla 6. Resultados de comparar los grupos de TDP y TDM usando tres síntomas presentes en pacientes con TDP

Característica	TDP/TDM
Tweets en tiempo pasado	1.08
Tweets en tiempo futuro	1.07
Enojo	0.83
Palabras asociadas a situaciones sociales	0.95

Los usuarios propuestos como TDP muestran, en promedio, un mayor número de tweets en tiempo pasado y futuro, así como tweets con menor uso de palabras relacionadas con situaciones sociales. Tanto el mayor valor de los tweets en pasado y futuro, como el menor número de palabras asociadas con situaciones sociales, se relacionan con los síntomas de los usuarios diagnosticados con TDP [11]. Aunque el enojo es una emoción presente en ambas patologías, los usuarios propuestos como TDP tuvieron una puntuación más baja en comparación con los propuestos como TDM, lo que puede indicar que estos últimos pueden presentar síntomas más graves que los usuarios propuestos como TDP.

4.5 Discusión

La Tabla 4 muestra que se obtuvieron mejores resultados al clasificar a los usuarios con

TDM. Una explicación para esto podría ser que los síntomas en esos pacientes son más notorios que en los pacientes con TDP. Los pacientes con TDM tienen síntomas más significativos y, por lo tanto, tienden a ser más perceptibles que los pacientes con TDP, donde los síntomas pueden ser más confusos para distinguir entre dichos usuarios y los usuarios de control.

Nuestro enfoque de clasificación mediante la concatenación de todos los mensajes del usuario mostró resultados prometedores como en investigaciones anteriores [34], [84]. En este mismo rubro, en el caso de pacientes con TDP, se necesita la persistencia de los síntomas durante al menos dos años, por ello, es preferible utilizar todo el historial para observar mejor los posibles rasgos que ayudar a diferenciar entre usuarios depresivos y de control. Por otro lado, para el caso de los usuarios con TDM, se podría sugerir utilizar uno o varios períodos de 14 días, dada la temporalidad propuesta por el DSM-V. Sin embargo, en muchos casos, los síntomas de un episodio depresivo pueden durar más de 14 días o el paciente puede presentar más de un episodio depresivo y por ahora es complejo saber cuándo comienza y termina el periodo depresivo de los sujetos en redes sociales; por esta razón, en este caso la clasificación por usuario concatenando todos los tweets también brindó buenos resultados.

Conclusiones

Los resultados mostraron que es más difícil distinguir a los usuarios propuestos como TDP que a los propuestos como TDM. Esto sugiere que las características observadas en los pacientes propuestos como TDM fueron mayormente diferenciables con respecto a los usuarios de control, lo que muestra dos puntos interesantes; el primer punto positivo es que los pacientes con TDM son considerados pacientes de mayor gravedad (debido a que pueden estar asociados a intentos suicidad o suicidios), por lo que es un buen indicar que el clasificador los identifique mejor. Por otro lado, el obtener más bajos resultados para los pacientes propuestos como TDP, representa un reto para desarrollar un modelo que permita detectar esas características en estos sujetos que son, hasta ahora, menos diferenciables.

La distinción propuesta para sujetos con rasgos de TDM y TDP aún tiene puntos importantes a mejorar y verificar en una base de datos que tenga usuarios diagnosticados especificando el tipo de depresión que sufre el paciente ya que en este trabajo solo se realizó una primera aproximación considerando la prevalencia de ambos tipos de depresión en área clínica.

Otro punto importante a tener en cuenta respecto a la temporalidad es el análisis de los pacientes para saber si se trata de una depresión crónica o no crónica. Esto implicaría observar la aparición de episodios depresivos o la sintomatología presente por un periodo considerable de tiempo. Además, dado que la sintomatología propuesta estaba dividida por días en el historial del usuario, se podría analizar otro tipo de depresión como es el trastorno afectivo estacional.

Un aspecto positivo de este trabajo fue la clasificación de usuarios empleando como criterio fundamental la temporalidad de los mensajes, lo que ayudó a observar el problema del análisis de rasgos depresivos en usuarios de redes sociales como un problema que no solo

involucraba texto, sino como un problema más complejo que involucraba un aspecto fundamental en el diagnóstico de esta patológica: la temporalidad. Este aspecto brindó una visión más amplia de los síntomas involucrados en esta enfermedad y nos hizo cuestionar aspectos de vital importancia como la duración de los episodios depresivos, lo que nos llevó a ver este problema como uno donde los usuarios deprimidos en redes sociales no solo involucraban sujetos que podrían estar padeciendo de Trastorno de Depresivo Mayor sino que podrían estar sufriendo de otras variantes de la enfermedad. Aunque resulte complejo considerar estos aspectos que sin duda generan más dudas que respuestas, consideramos que es vital mencionarlas y contemplarlas para poder desarrollar un modelo que sea más confiable.

Por otro lado, aunque se generaron resultados cercanos a los observados en el estado del arte, aún hay puntos que desde la visión de la Inteligencia Artificial se pueden mejorar, como el caso de los hiperparámetros ya que, implementando un mejor modelo para optimizar estos valores se podrían alcanzar mejores resultados. Otro punto importante es la implementación de modelos del estado del arte en el Procesamiento de Lenguaje Natural, como el caso de BERT, que podría tener efectos positivos para analizar emociones. En general, se tiene que mejorar el análisis de síntomas.

5.1 Trabajos Futuros

La depresión es una enfermedad con un alto impacto en nuestra sociedad, que tienen repercusiones no sólo en quien padece dicho padecimiento sino en la sociedad en general. Además, las redes sociales pueden ser una fuente de datos para observar a pacientes con rasgos asociados a esta patología. Por ello, se planea continuar buscando herramientas que puedan contribuir a observar características que ayuden a identificar a estos usuarios, como el uso de BERT para analizar tristeza o posible similaridad semántica entre los tweets de los usuarios y tweets que muestren algunas características de síntomas asociados a la depresión.

Referencias

- [1] American Psychiatric association, *Diagnostic and statistical manual of mental disorders*. 2013.
- [2] R. Salas-Zárate, G. Alor-Hernández, M. D. P. Salas-Zárate, M. A. Paredes-Valverde, M. Bustos-López, and J. L. Sánchez-Cervantes, “Detecting Depression Signs on Social Media: A Systematic Literature Review,” *Healthc.*, vol. 10, no. 2, pp. 1–22, 2022, doi: 10.3390/healthcare10020291.
- [3] Instituto Nacional de Estadística y Geografía (INEGI), “Presenta Inegi Resultados De La Primera Encuesta Nacional De Bienestar Autorreportado (Enbiare) 2021,” *Comun. Prensa Núm. 772/21*, pp. 1–3, 2021.
- [4] X. Chen, M. D. Sykora, T. W. Jackson, and S. Elayan, “What about Mood Swings: Identifying Depression on Twitter with Temporal Measures of Emotions,” *Web Conf. 2018 - Companion World Wide Web Conf. WWW 2018*, vol. 2, pp. 1653–1660, 2018, doi: 10.1145/3184558.3191624.
- [5] A. Culotta, “Detecting influenza outbreaks by analyzing Twitter messages,” no. May, pp. 1–11, 2010, [Online]. Available: <http://arxiv.org/abs/1007.4748>.
- [6] R. A. Calvo, D. N. Milne, M. S. Hussain, and H. Christensen, “Natural language processing in mental health applications using non-clinical texts,” *Nat. Lang. Eng.*, vol. 23, no. 5, pp. 649–685, 2017, doi: 10.1017/S1351324916000383.
- [7] S. Chancellor and M. De Choudhury, “Methods in predictive techniques for mental health status on social media: a critical review,” *npj Digit. Med.*, vol. 3, no. 1, 2020, doi: 10.1038/s41746-020-0233-7.
- [8] R. Caruana, Y. Lou, J. Gehrke, P. Koch, M. Sturm, and N. Elhadad, “Intelligible Models for HealthCare,” pp. 1721–1730, 2015, doi: 10.1145/2783258.2788613.
- [9] A. Lee, J. K. Kummerfeld, L. C. An, and R. Mihalcea, “Micromodels for Efficient, Explainable, and Reusable Systems: A Case Study on Mental Health,” *Find. Assoc. Comput. Linguist. Find. ACL EMNLP 2021*, pp. 4257–4272, 2021, doi: 10.18653/v1/2021.findings-emnlp.360.
- [10] G. Coppersmith, M. Dredze, C. Harman, K. Hollingshead, and M. Mitchell, “CLPsych 2015 Shared Task: Depression and PTSD on Twitter,” *2nd Comput. Linguist. Clin. Psychol. From Linguist. Signal to Clin. Reality, CLPsych 2015 - Proc. Work.*, pp. 31–39, 2015, doi: 10.3115/v1/w15-1204.

- [11] M. W. Eysenck, S. Payne, and R. Santos, "Anxiety and depression: Past, present, and future events," *Cogn. Emot.*, vol. 20, no. 2, pp. 274–294, 2006, doi: 10.1080/02699930500220066.
- [12] J. Boardman, "Social exclusion and mental health - How people with mental health problems are disadvantaged: An overview," *Ment. Heal. Soc. Incl.*, vol. 15, no. 3, pp. 112–121, 2011, doi: 10.1108/20428301111165690.
- [13] H. Cai *et al.*, "Prevalence of suicidal ideation and planning in patients with major depressive disorder: A meta-analysis of observation studies," *J. Affect. Disord.*, vol. 293, no. May, pp. 148–158, 2021, doi: 10.1016/j.jad.2021.05.115.
- [14] F. T. Giuntini, M. T. Cazzolato, M. de J. D. dos Reis, A. T. Campbell, A. J. M. Traina, and J. Ueyama, "A review on recognizing depression in social networks: challenges and opportunities," *J. Ambient Intell. Humaniz. Comput.*, vol. 11, no. 11, pp. 4713–4729, 2020, doi: 10.1007/s12652-020-01726-4.
- [15] T. Nguyen, A. Yates, A. Zirikly, B. Desmet, and A. Cohan, "Improving the Generalizability of Depression Detection by Leveraging Clinical Questionnaires," pp. 8446–8459, 2022, doi: 10.18653/v1/2022.acl-long.578.
- [16] C. Aguirre, K. Harrigian, and M. Dredze, "Gender and racial fairness in depression research using social media," *EACL 2021 - 16th Conf. Eur. Chapter Assoc. Comput. Linguist. Proc. Conf.*, pp. 2932–2949, 2021, doi: 10.18653/v1/2021.eacl-main.256.
- [17] M. A. Villarroel and E. P. Terlizzi, "Symptoms of Depression Among Adults: United States, 2019," *NCHS Data Brief*, no. 379, pp. 1–8, 2020.
- [18] T. Thomas, "REGULAR ARTICLES The Language," no.23, March, 1982.
- [19] R. N. Ramirez-Esparza, C. K. Chung, E. Kacewicz, and J. W. Pennebaker, "The psychology of word use in depression forums in english and in spanish: Testing two text analytic approaches," *ICWSM 2008 - Proc. 2nd Int. Conf. Weblogs Soc. Media*, pp. 102–108, 2008.
- [20] J. W. Pennebaker, C. K. Chung, M. Ireland, A. Gonzales, and R. J. Booth, "The Development and Psychometric Properties of LIWC2007 The University of Texas at Austin," *Development*, vol. 1, no. 2, pp. 1–22, 2007, doi: 10.13140/RG.2.2.23890.43205.
- [21] C. Chew and G. Eysenbach, "Pandemics in the age of Twitter: Content analysis of tweets during the 2009 H1N1 outbreak," *PLoS One*, vol. 5, no. 11, pp. 1–13, 2010, doi: 10.1371/journal.pone.0014118.

- [22] M. De Choudhury, S. Counts, and E. Horvitz, “Social media as a measurement tool of depression in populations,” *Proc. 5th Annu. WebSci’13*, vol. volume, pp. 47–56, 2013, doi: 10.1145/2464464.2464480.
- [23] G. Coppersmith, M. Dredze, “Quantifying Mental Health Signals in Twitter,” no. 222886, pp. 2–5, 2014.
- [24] P. Resnik, W. Armstrong, L. Claudino, and T. Nguyen, “The University of Maryland CLPsych 2015 Shared Task System,” *2nd Comput. Linguist. Clin. Psychol. From Linguist. Signal to Clin. Reality, CLPsych 2015 - Proc. Work.*, no. c, pp. 54–60, 2015, doi: 10.3115/v1/w15-1207.
- [25] S. Arora *et al.*, “A practical algorithm for topic modeling with provable guarantees,” *30th Int. Conf. Mach. Learn. ICML 2013*, vol. 28, no. PART 2, pp. 939–947, 2013.
- [26] M. Nadeem, “Identifying Depression on Twitter,” pp. 1–9, 2016, [Online]. Available: <http://arxiv.org/abs/1607.07384>.
- [27] D. Mowery, A. Park, M. Conway, and C. Bryan, “Towards automatically classifying depressive symptoms from Twitter data for population health.,” *Work. Comput. Model. People’s Opin. Personal. Emot. Soc. Media*, pp. 182–191, 2016, [Online]. Available: <http://www.aclweb.org/anthology/W/W16/W16-43.pdf#page=196>.
- [28] X. Tao, X. Zhou, J. Zhang, and J. Yong, “Sentiment analysis for depression detection on social networks,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10086 LNAI, pp. 807–810, 2016, doi: 10.1007/978-3-319-49586-6_59.
- [29] M. Deshpande and V. Rao, “Depression detection using emotion artificial intelligence,” *Proc. Int. Conf. Intell. Sustain. Syst. ICISS 2017*, no. Iciss, pp. 858–862, 2018, doi: 10.1109/ISS1.2017.8389299.
- [30] G. Shen *et al.*, “Depression detection via harvesting social media: A multimodal dictionary learning solution,” *IJCAI Int. Jt. Conf. Artif. Intell.*, vol. 0, pp. 3838–3844, 2017, doi: 10.24963/ijcai.2017/536.
- [31] R. Singh *et al.*, *A Framework for Early Detection of Antisocial Behavior on Twitter Using Natural Language Processing*, vol. 993. Springer International Publishing, 2020.
- [32] A. Biradar and S. G. Totad, *Detecting Depression in Social Media Posts Using Machine Learning*, vol. 1037. Springer Singapore, 2019.

- [33] M. R. Islam, A. R. M. Kamal, N. Sultana, R. Islam, M. A. Moni, and A. Ulhaq, "Detecting Depression Using K-Nearest Neighbors (KNN) Classification Technique," *Int. Conf. Comput. Commun. Chem. Mater. Electron. Eng. IC4ME2 2018*, pp. 1–4, 2018, doi: 10.1109/IC4ME2.2018.8465641.
- [34] A. H. Orabi, P. Buddhitha, M. H. Orabi, and D. Inkpen, "Deep learning for depression detection of Twitter users," *Proc. 5th Work. Comput. Linguist. Clin. Psychol. From Keyboard to Clin. CLPsych 2018 2018 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol.*, pp. 88–97, 2018, doi: 10.18653/v1/w18-0609.
- [35] A. Leis, F. Ronzano, M. A. Mayer, L. I. Furlong, and F. Sanz, "Detecting signs of depression in tweets in Spanish: Behavioral and linguistic analysis," *J. Med. Internet Res.*, vol. 21, no. 6, 2019, doi: 10.2196/14199.
- [36] F. CACHEDA, D. Fernandez, F. J. Novoa, and V. Carneiro, "Early detection of depression: Social network analysis and random forest techniques," *J. Med. Internet Res.*, vol. 21, no. 6, 2019, doi: 10.2196/12554.
- [37] and J. P. David E. Losada, Fabio Crestani, "eRISK 2017: CLEF Lab on Early Risk Prediction on the Internet: Experimental Foundations," *Eval. Hierarchical Clust. via Markov Decis. Process. Effic. Navig. Search*, vol. 10456, pp. 125–131, 2017, doi: 10.1007/978-3-319-65813-1.
- [38] A. Leis, F. Ronzano, M. A. Mayer, L. I. Furlong, and F. Sanz, "Evaluating behavioral and linguistic changes during drug treatment for depression using tweets in Spanish: Pairwise comparison study," *J. Med. Internet Res.*, vol. 22, no. 12, 2020, doi: 10.2196/20920.
- [39] J. Zhou, H. Zogan, S. Yang, S. Jameel, G. Xu, and F. Chen, "Detecting Community Depression Dynamics Due to COVID-19 Pandemic in Australia," *IEEE Trans. Comput. Soc. Syst.*, vol. 8, no. 4, pp. 958–967, 2021, doi: 10.1109/TCSS.2020.3047604.
- [40] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," *NAACL HLT 2019 - 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - Proc. Conf.*, vol. 1, no. M1m, pp. 4171–4186, 2019.
- [41] S. Ji, T. Zhang, L. Ansari, J. Fu, P. Tiwari, and E. Cambria, "MentalBERT: Publicly Available Pretrained Language Models for Mental Healthcare," 2021, [Online]. Available: <http://arxiv.org/abs/2110.15621>.
- [42] A. S. Uban, B. Chulvi, and P. Rosso, "An emotion and cognitive based analysis of

- mental health disorders from social media data,” *Futur. Gener. Comput. Syst.*, vol. 124, pp. 480–494, 2021, doi: 10.1016/j.future.2021.05.032.
- [43] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient estimation of word representations in vector space,” *1st Int. Conf. Learn. Represent. ICLR 2013 - Work. Track Proc.*, pp. 1–12, 2013.
- [44] E. Schramm, D. N. Klein, M. Elsaesser, T. A. Furukawa, and K. Domschke, “Review of dysthymia and persistent depressive disorder: history, correlates, and clinical implications,” *The Lancet Psychiatry*, vol. 7, no. 9, pp. 801–812, 2020, doi: 10.1016/S2215-0366(20)30099-7.
- [45] D. Preoțiu-Pietro, M. Sap, H. A. Schwartz, and L. Ungar, “Mental Illness Detection at the World Well-Being Project for the CLPsych 2015 Shared Task,” *2nd Comput. Linguist. Clin. Psychol. From Linguist. Signal to Clin. Reality, CLPsych 2015 - Proc. Work.*, pp. 40–45, 2015, doi: 10.3115/v1/w15-1205.
- [46] K. Harrigian, C. Aguirre, and M. Dredze, “Do models of mental health based on social media data generalize?,” *Find. Assoc. Comput. Linguist. Find. ACL EMNLP 2020*, pp. 3774–3788, 2020, doi: 10.18653/v1/2020.findings-emnlp.337.
- [47] N. F. Wilma Bucci, “The language of depression.” 1981.
- [48] W. Weintraub, “Verbal Behavior: Adaptation and Psychopathology.,” vol. 46, no. 2, p. 55, 1982, [Online]. Available: <http://eprints.uanl.mx/5481/1/1020149995.PDF>.
- [49] S. S. Rude, E. M. Gortner, and J. W. Pennebaker, “Language use of depressed and depression-vulnerable college students,” *Cogn. Emot.*, vol. 18, no. 8, pp. 1121–1133, 2004, doi: 10.1080/02699930441000030.
- [50] D. Smirnova, E. Sloeva, N. Kuvshinova, A. Krasnov, D. Romanov, and G. Nosachev, “1419 – Language changes as an important psychopathological phenomenon of mild depression,” *Eur. Psychiatry*, vol. 28, p. 1, 2013, doi: 10.1016/s0924-9338(13)76454-x.
- [51] M. Al-Mosaiwi and T. Johnstone, “In an Absolute State: Elevated Use of Absolutist Words Is a Marker Specific to Anxiety, Depression, and Suicidal Ideation,” *Clin. Psychol. Sci.*, vol. 6, no. 4, pp. 529–542, 2018, doi: 10.1177/2167702617747074.
- [52] M. De Choudhury, Scott, S. Counts, M. Gamon, and E. Horvitz, “Predicting Depression via Social Media,” *Proc. Seventh Int. AAAI Conf. Weblogs Soc. Media*, vol. 128, pp. 109–127, 2013, doi: 10.1007/978-981-19-1724-0_6.
- [53] X. Lin, Y. Fu, Z. Yang, N. Lin, and S. Jiang, “Detecting Depression in Social Media

- using Prompt-Learning and Word-Emotion Cluster,” pp. 200–205, 2022, [Online]. Available: <https://pytorch.org/>.
- [54] P. H. Matthews, “N. Chomsky Aspects of the Theory of Syntax. Cambridge, Mass.: M.I.T. Press, 1965. Pp. x + 251.” *J. Linguist.*, vol. 3, no. 1, pp. 119–152, 1967, doi: 10.1017/s0022226700012998.
- [55] I. Bolshakov, “Computational linguistics,” *Love Lang. An Introd. to Linguist.*, pp. 582–602, 2019, doi: 10.1017/9781108568753.019.
- [56] M. Ruslan, “The Oxford Handbook of computational linguistics,” no. 1, pp. 1–9, 2020, [Online]. Available: <http://journal.um-surabaya.ac.id/index.php/JKM/article/view/2203>.
- [57] K. R. Chowdhary, *Natural Language Processing BT - Fundamentals of Artificial Intelligence*. 2020.
- [58] P. Nandwani and R. Verma, “A review on sentiment analysis and emotion detection from text,” *Soc. Netw. Anal. Min.*, vol. 11, no. 1, pp. 1–19, 2021, doi: 10.1007/s13278-021-00776-6.
- [59] J. Rottenberg, “Emotions in Depression: What Do We Really Know?,” *Annu. Rev. Clin. Psychol.*, vol. 13, pp. 241–263, 2017, doi: 10.1146/annurev-clinpsy-032816-045252.
- [60] S. M. Mohammad, “Word affect intensities,” *Lr. 2018 - 11th Int. Conf. Lang. Resour. Eval.*, pp. 174–183, 2022.
- [61] T. N. Flynn and A. A. J. Marley, “Worst Scaling : Theory and Methods,” pp. 178–201, 1992.
- [62] R. P. LEBRET, “Words Embeddings for Natural Language Processing,” *Graph Learn. Netw. Sci. Nat. Lang. Process.*, vol. 7148, pp. 77–96, 2022, doi: 10.1201/9781003272649-4.
- [63] C. D. M. Jeffrey Pennington, Richard Socher, “GloVe: Global Vectors for Word Representation Jeffrey,” *Br. J. Neurosurg.*, vol. 31, no. 6, pp. 682–687, 2014, doi: 10.1080/02688697.2017.1354122.
- [64] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, “Enriching Word Vectors with Subword Information,” *Trans. Assoc. Comput. Linguist.*, vol. 5, pp. 135–146, 2017, doi: 10.1162/tacl_a_00051.
- [65] N. Sekar, S. Chandrakala, and G. Prakash, “Analysis of global word representations

- for depression detection,” *CEUR Workshop Proc.*, vol. 2786, pp. 136–148, 2021.
- [66] S. T. Dumais, G. W. Furnas, T. K. Landauer, S. Deerwester, and R. Harshman, “Using latent semantic analysis to improve access to textual information,” *Conf. Hum. Factors Comput. Syst. - Proc.*, vol. Part F1302, pp. 281–285, 1988, doi: 10.1145/57167.57214.
- [67] P. H. YANN LECUN, LEON BOTTOU, YOSHUA BENGIO, “Gradient-Based Learning Applied to Document Recognition,” *Proc. IEEE*, vol. 86, no. 11, 1998, doi: 10.1016/j.bbrc.2005.03.111.
- [68] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: A simple way to prevent neural networks from overfitting,” *J. Mach. Learn. Res.*, vol. 15, pp. 1929–1958, 2014.
- [69] A. T. Beck, C. H. Ward, M. Mendelson, J. Mock, and J. Erbaugh, “An Inventory for Measuring Depression The difficulties inherent in obtaining,” pp. 561–571, 1960.
- [70] M. Hamilton, “Development of a rating scale for depressive illness.,” *Mod. Probl. Pharmacopsychiatry*, vol. 7, no. 0, pp. 139–150, 1974, doi: 10.1159/000395073.
- [71] L. S. Radolff, “The CES-D scale: a self-report depression scale for research in the general population,” *Appl. Psychol. Meas.*, vol. 1, no. 3, pp. 385–401, 1977.
- [72] C. Otte *et al.*, “Major depressive disorder,” *Nat. Rev. Dis. Prim.*, vol. 2, no. Mdd, pp. 1–21, 2016, doi: 10.1038/nrdp.2016.65.
- [73] J. A. Murphy and G. J. Byrne, “Prevalence and correlates of the proposed DSM-5 diagnosis of Chronic Depressive Disorder,” *J. Affect. Disord.*, vol. 139, no. 2, pp. 172–180, 2012, doi: 10.1016/j.jad.2012.01.033.
- [74] S. Satyanarayana, M. W. Enns, B. J. Cox, and J. Sareen, “Prevalence and correlates of chronic depression in the Canadian Community Health Survey: Mental health and well-being,” *Can. J. Psychiatry*, vol. 54, no. 6, pp. 389–398, 2009, doi: 10.1177/070674370905400606.
- [75] G. Rao, Y. Zhang, L. Zhang, Q. Cong, and Z. Feng, “MGL-CNN: A Hierarchical Posts Representations Model for Identifying Depressed Individuals in Online Forums,” *IEEE Access*, vol. 8, pp. 32395–32403, 2020, doi: 10.1109/ACCESS.2020.2973737.
- [76] H. Kour and M. K. Gupta, *An hybrid deep learning approach for depression prediction from user tweets using feature-rich CNN and bi-directional LSTM*, vol. 81, no. 17. Multimedia Tools and Applications, 2022.

- [77] F. Benazzi, “Chronic depression: A case series of 203 outpatients treated at a private practice,” *J. Psychiatry Neurosci.*, vol. 23, no. 1, pp. 51–55, 1998.
- [78] J. C. E. Moran, H. Kocsis, and A. J. Frances, “John C. arkowitz, my E. Moran, Jmes H. Kocsis and Allen J. Frances a,” *Psychiatry Interpers. Biol. Process.*, vol. 24, pp. 63–71, 1992.
- [79] S. Ildirli, Y. B. Şair, and F. Dereboy, “Persistent depression as a novel diagnostic category: Results from the menderes depression study,” *Noropsikiyatri Ars.*, vol. 52, no. 4, pp. 359–366, 2015, doi: 10.5152/npa.2015.7589.
- [80] A. Benton, M. Mitchell, and D. Hovy, “Multi-Task Learning for Mental Health using Social Media Text,” 2017, [Online]. Available: <http://arxiv.org/abs/1712.03538>.
- [81] L. Yang and A. Shami, “On hyperparameter optimization of machine learning algorithms: Theory and practice,” *Neurocomputing*, vol. 415, pp. 295–316, 2020, doi: 10.1016/j.neucom.2020.07.061.
- [82] M. Sokolova, N. Japkowicz, and S. Szpakowicz, “Beyond accuracy, F-score and ROC: A family of discriminant measures for performance evaluation,” *AAAI Work. - Tech. Rep.*, vol. WS-06-06, pp. 24–29, 2006, doi: 10.1007/11941439_114.
- [83] C. Aloysius and P. Tamilselvan, “A Novel Method to Reduce False Positives and Negatives in Sentiment Analysis,” *Int. J. Intell. Syst. Appl. Eng.*, vol. 10, no. 3, pp. 365–373, 2022.
- [84] Y. Zhang, H. Lyu, Y. Liu, X. Zhang, Y. Wang, and J. Luo, “Monitoring Depression Trend on Twitter during the COVID-19 Pandemic: Observational Study (Preprint),” *JMIR Form. Res.*, 2020, doi: 10.2196/26769.