



# Universidad Autónoma del Estado de Hidalgo

Instituto de Ciencias Básicas e Ingeniería

**Detección de usuarios con depresión a través del análisis de sus  
publicaciones en redes sociales**

T E S I S

Que para obtener el grado de

Licenciada en Ciencias Computacionales

**presenta**

Camila Fernanda Cienfuegos Velázquez

**Directores de Tesis:**

Dra. Rosa María Ortega Mendoza

Dr. Félix Agustín Castro Espinoza

Mineral de la Reforma, Hidalgo. Abril 2026



Mineral de la Reforma, Hgo., a 9 de abril de 2026

Número de control: ICBI-D/630/2026  
 Asunto: Autorización de impresión.

**MTRA. OJUKY DEL ROCÍO ISLAS MALDONADO**  
**DIRECTORA DE ADMINISTRACIÓN ESCOLAR DE LA UAEH**

Con Título Quinto, Capítulo II, Capítulo V, Artículo 51 Fracción IX del Estatuto General de nuestra Institución, por este medio, le comunico que el Jurado asignado a la egresada de la Licenciatura en Ciencias Computacionales **Camila Fernanda Cienfuegos Velázquez**, quien presenta el trabajo de titulación "**Detección de usuarios con depresión a través del análisis de sus publicaciones en redes sociales**", ha decidido, después de revisar fundamento en lo dispuesto en el Título Tercero, Capítulo I, Artículo 18 Fracción IV; dicho trabajo en la reunión de sinodales, **autorizar la impresión del mismo**, una vez realizadas las correcciones acordadas.

A continuación, firman de conformidad los integrantes del Jurado:

**Presidente:** Dra. Guadalupe Carmona Arroyo

**Secretario:** Dr. Ricardo Calderón Suárez

**Vocal:** Dra. Rosa María Ortega Mendoza

**Suplente:** Dr. Félix Agustín Castro Espinoza

1961 - 2026

Sin otro particular por el momento, reciba un cordial saludo.

Atentamente  
 "Amor, Orden y Progreso"

Dr. Gabriel Vergara Rodríguez  
 Director del ICBI



GVR/MMM

"Amor, Orden y Progreso"

Ciudad del Conocimiento, Carretera Pachuca-Tulancingo  
 Km. 4.5 Colonia Carboneras, Mineral de la Reforma, Hidalgo,  
 México. C.P 42184  
 Teléfono: 52 (771) 71 720 00 Ext. 40001  
 direccion\_icbi@uaeh.edu.mx, vergarar@uaeh.edu.mx



---

# Dedicatoria

*A mi madre y a mi padre, por su amor incondicional, su apoyo constante y por enseñarme, con su ejemplo, el valor del esfuerzo.*

*A mi hermano, por ser mi compañero de vida y mi confidente*

*A mis abuelos, por inculcarme la importancia de la perseverancia y el trabajo duro.*

*A mi mejor amiga, por estar a mi lado en cada paso del camino, por su cariño, su fe en mí y por ser un pilar en mi vida durante todos estos años.*

*Gracias a todos ustedes por sus enseñanzas en este camino llamado vida.*

---

# Agradecimientos

Quiero expresar mi más sincero agradecimiento a mi familia, cuyo inquebrantable apoyo y estímulo constante han sido pilares fundamentales para la realización de este trabajo.

Mi profundo reconocimiento a la Dra. Rosa María Ortega Mendoza y al Dr. Félix Agustín Castro Espinoza, asesores de esta tesis, por su paciencia, dedicación y valiosas enseñanzas, las cuales han enriquecido de manera significativa mi formación académica. De igual forma, agradezco al Dr. Ricardo Calderón Suárez y a la Dra. Guadalupe Carmona Arroyoal, sinodales de este trabajo, por sus observaciones, comentarios y aportaciones que contribuyeron al fortalecimiento de esta investigación.

Asimismo, agradezco al Mtro. Gamaliel Alberto López Reyes, compañero académico, por su orientación experta y generosa disposición para compartir sus conocimientos.

Extiendo mi gratitud a los docentes de la licenciatura, cuya entrega y experiencia me proporcionaron una sólida base teórica y práctica indispensable para el desarrollo de esta investigación.

Agradezco también al Ing. José Luis Castañeda Villamil por la confianza brindada y por su estímulo profesional.

También deseo agradecer a mis amigos —tanto a los compañeros que conocí durante la carrera como a quienes me han acompañado desde etapas anteriores— por sus consejos, su apoyo moral y los momentos que hicieron más llevadero este proceso.

Finalmente, agradezco a la Universidad Autónoma del Estado de Hidalgo por ofrecer un entorno académico de excelencia y por los recursos que han contribuido al éxito de este estudio.

---

# Resumen

En la actualidad, las redes sociales son espacios donde las personas comparten pensamientos, emociones, experiencias y opiniones. El texto generado constituye una fuente valiosa de conocimiento para estudiar el comportamiento y la salud mental de las personas. Particularmente, en este trabajo se analizan publicaciones de usuarios en distintas plataformas, principalmente X (antes *Twitter*) y *Reddit*, con el objetivo de identificar patrones lingüísticos asociados con rasgos de depresión y comparar su manifestación en redes sociales.

Para realizar esta investigación, se utilizaron tres conjuntos de datos de referencia, los cuales contienen publicaciones provenientes de tales plataformas. Utilizando tales colecciones, se crearon modelos computacionales para detectar usuarios con depresión mediante el uso de técnicas de procesamiento de lenguaje natural y aprendizaje automático. Se entrenaron varios modelos de clasificación para distinguir entre textos etiquetados como casos de depresión y aquellos que no lo son, priorizando la puntuación F1 en la clase positiva.

Los resultados indican que, a pesar de las diferentes características de las publicaciones cada plataforma social (e.g., longitud máxima de las publicaciones), existen similitudes en el uso del lenguaje asociado a depresión. Además, se observó que el desempeño de los modelos de clasificación depende en cierta medida de la colección de datos utilizada en el entrenamiento.

**Palabras clave:** procesamiento de lenguaje natural, redes sociales, depresión, aprendizaje automático, bolsa de palabras.

---

# Abstract

Nowadays, social media platforms are spaces where people share thoughts, emotions, experiences, and opinions. The text generated there constitutes a valuable source of knowledge for studying people's behavior and their mental health. Specifically, this study analyzes user posts on diverse platforms, primarily X (formerly Twitter) and Reddit, with the aim of identifying linguistic patterns associated with signs of depression and comparing their manifestation on social media.

To conduct this research, three reference datasets were used, which contain posts from social media platforms. Using these collections, computational models were created to detect users with depression through the use of natural language processing and machine learning techniques. Several classification models were trained to distinguish between texts labeled as cases of depression and those that are not, prioritizing the F1 score in the positive class.

The results indicate that, despite the different characteristics of posts on each social platform (e.g., maximum post length), there are similarities in the language use associated with depression. Furthermore, it was observed that the performance of the classification models depends to some extent on the dataset used in training.

**Keywords:** Natural Language Processing, social media, depression, machine learning, bag-of-words.

# Índice general

<b>1</b>	<b>Introducción</b>	<b>1</b>
1.1	Planteamiento del problema . . . . .	2
1.2	Pregunta de investigación . . . . .	3
1.3	Justificación . . . . .	3
1.4	Motivación . . . . .	4
1.5	Objetivo . . . . .	4
1.5.1	Objetivo general . . . . .	4
1.5.2	Objetivos específicos . . . . .	4
1.6	Organización de la tesis . . . . .	5
<b>2</b>	<b>Marco Teórico</b>	<b>7</b>
2.1	Aprendizaje automático . . . . .	7
2.1.1	Aprendizaje Supervisado . . . . .	7
2.1.2	Clasificadores . . . . .	8
2.1.2.1	Bosques aleatorios . . . . .	8
2.1.2.2	Máquinas de vectores de soporte . . . . .	9
2.1.2.3	Regresión Logística . . . . .	9
2.1.2.4	Vecinos más cercanos . . . . .	11
2.1.3	Desempeño de modelos de clasificación supervisada . . . . .	12
2.1.3.1	Matriz de confusión . . . . .	12
2.1.3.2	Exactitud . . . . .	13
2.1.3.3	Precisión . . . . .	14

2.1.3.4	Exhaustividad . . . . .	14
2.1.3.5	Macro F1 . . . . .	14
2.1.3.6	Validación cruzada . . . . .	15
2.1.4	Selección de atributos . . . . .	15
2.1.4.1	Chi cuadrada . . . . .	15
2.2	Procesamiento del Lenguaje Natural . . . . .	16
2.2.1	Clasificación automática de texto . . . . .	16
2.2.1.1	Bolsa de palabras . . . . .	17
2.2.1.2	Pesado de términos . . . . .	18
<b>3</b>	<b>Estado del Arte</b>	<b>21</b>
3.1	La depresión en entornos digitales . . . . .	21
3.2	Enfoques computacionales para la detección de depresión . . . . .	26
3.3	Conjuntos de datos para detección de depresión . . . . .	31
3.3.1	<i>Reddit Self-reported Depression Diagnosis</i> . . . . .	31
3.3.2	Dreddit . . . . .	32
3.3.3	<i>eRisk</i> . . . . .	33
3.3.4	<i>Binary Labeled Depression Corpus of 60,000 English tweets</i> . . . . .	35
<b>4</b>	<b>Experimentos y resultados</b>	<b>37</b>
4.1	Metodología . . . . .	37
4.2	Experimentos basados en clasificación . . . . .	39
4.2.1	Configuración experimental . . . . .	39
4.2.2	Experimento 1: El impacto del preprocesamiento . . . . .	41
4.2.3	Experimento 2: Evaluación de diferentes algoritmos de aprendizaje . . . . .	43
4.2.4	Experimento 3: Evaluando diferentes esquemas de pesado . . . . .	45
4.2.5	Experimento 4: Aplicando selección de características . . . . .	47
4.2.6	Evaluación de los modelos: matrices de confusión bajo las mejores configuraciones . . . . .	50

4.2.7	Análisis comparativo entre plataformas sociales . . . . .	52
<b>5</b>	<b>Conclusiones y trabajo futuro</b>	<b>57</b>
5.1	Conclusiones . . . . .	57
5.2	Limitaciones . . . . .	58
5.3	Trabajo futuro . . . . .	59

# Índice de figuras

2.1	Representación del algoritmo de Bosques Aleatorios . . . . .	8
2.2	Máquina de vectores de soporte . . . . .	9
2.3	Representación de Regresión Logística . . . . .	10
2.4	Clasificación KNN con $K = 1$ . . . . .	11
2.5	Clasificación KNN con $K = 6$ . . . . .	12
2.6	Representación de una matriz de confusión . . . . .	13
2.7	Proceso de clasificación de textos . . . . .	17
2.8	Ejemplo de representación binaria en bolsa de palabras . . . . .	18
2.9	Representación de una bolsa de palabras . . . . .	18
4.1	Metodología adoptada para la clasificación de textos . . . . .	38
4.2	Análisis del impacto del preprocesamiento en la colección <i>eRisk 2018</i> . . . . .	41
4.3	Análisis del impacto del preprocesamiento en la colección de <i>Twitter</i> . . . . .	42
4.4	Comparación del desempeño de diferentes clasificadores en el conjunto de datos <i>eRisk 2017</i> . . . . .	43
4.5	Comparación del desempeño de diferentes clasificadores en el conjunto de datos <i>eRisk 2018</i> . . . . .	44
4.6	Comparación del desempeño de diferentes clasificadores en el conjunto de datos de <i>Twitter</i> . . . . .	44
4.7	Selección de atributos en <i>eRisk 2017</i> . . . . .	47
4.8	Selección de atributos en <i>eRisk 2018</i> . . . . .	48
4.9	Selección de atributos en <i>Twitter</i> . . . . .	49
4.10	Matriz de confusión del mejor modelo en <i>eRisk 2017</i> . . . . .	50

4.11 Matriz de confusión del mejor modelo en <i>eRisk 2018</i> . . . . .	51
4.12 Matriz de confusión del mejor modelo en <i>Tweets</i> . . . . .	52
4.13 Nube de palabras más discriminativas en <i>eRisk 2017</i> . . . . .	53
4.14 Nube de palabras más discriminativas en <i>eRisk 2018</i> . . . . .	54
4.15 Nube de palabras más discriminativas en el conjunto de datos <i>Twitter</i> . . .	55

# Índice de tablas

3.1	Distribución de clases en el conjunto de datos <i>RSDD</i> . . . . .	32
3.2	Distribución de clases en el conjunto de datos <i>Dreaddit</i> . . . . .	33
3.3	Distribución de clases en <i>eRisk 2017</i> y <i>eRisk 2018</i> . . . . .	35
3.4	Distribución de clases en <i>Binary Labeled Depression Corpus</i> . . . . .	36
4.1	Comparación del desempeño al aplicar el esquema de pesado binario . . . . .	45
4.2	Comparación del desempeño al aplicar el esquema de pesado TF . . . . .	46
4.3	Comparación del desempeño al aplicar el esquema de pesado TF-IDF . . . . .	46

---

# Capítulo 1

## Introducción

La depresión es un trastorno mental que se caracteriza por un estado de ánimo bajo, así como por la pérdida de interés o placer en actividades cotidianas [1]. Puede causar síntomas graves, tales como sentimientos persistentes de tristeza, dificultad para concentrarse y fatiga, los cuales interfieren en la realización de actividades y causan un grado considerable de angustia a quienes la padecen [2].

Según la Organización Mundial de la Salud (OMS), la depresión afecta a cerca de 332 millones de personas en el mundo, alcanzando un 5.7% en adultos, siendo más elevada en mujeres (6.9%) que en hombres (4.6%) [1]. En México, de acuerdo con boletines emitidos por el Instituto Nacional de Estadística y Geografía (INEGI), cerca de 34.85 millones de personas han presentado síntomas depresivos, aunque sólo 1.63 millones han recibido tratamiento farmacológico con antidepresivos [3]. Tras la pandemia por COVID-19, debido a las condiciones de aislamiento y contención a las que se enfrentó la población, la depresión aumentó considerablemente en comparación con lo registrado antes de la crisis [4]. En este contexto, también se ha identificado que los jóvenes y estudiantes universitarios presentan un riesgo considerable de desarrollar ideación y conductas suicidas, el cual se incrementa ante la presencia de depresión o el consumo de sustancias psicoactivas y alcohol [5].

Ante este panorama, se ha comenzado a explorar cómo las personas comunican su estado emocional fuera del entorno clínico, por ejemplo, a través de redes sociales como *Face-*

*book*<sup>1</sup>, una vía cada vez más habitual para compartir emociones y experiencias personales [6]. Según el Fondo de las Naciones Unidas para la Infancia (UNICEF) el 70 % declaró haber usado redes sociales para expresarse sobre temas de salud mental [7], lo cual abre la posibilidad de emplear herramientas computacionales que permitan analizar el lenguaje utilizado en este tipo de publicaciones [8].

En este contexto, investigaciones previas han identificado patrones lingüísticos<sup>2</sup> recurrentes en personas con depresión, como el uso frecuente de pronombres en primera persona, expresiones negativas y términos con carga emocional [9, 10]. Es por ello que la presente investigación tiene como propósito detectar casos positivos de depresión en textos publicados en plataformas como X<sup>3</sup> (antes *Twitter*) y *Reddit*<sup>4</sup> y comparar el uso del lenguaje asociado a depresión. Para ello, se emplean técnicas de procesamiento de lenguaje natural y aprendizaje automático, utilizando representaciones basadas en bolsa de palabras.

## 1.1. Planteamiento del problema

La depresión es un trastorno del estado de ánimo que puede afectar la vida emocional, social y funcional de quienes la padecen, y puede derivar en consecuencias graves si no se atiende a tiempo, como por ejemplo, el suicidio [3]. Aunque existen métodos clínicos para su diagnóstico, como entrevistas y cuestionarios aplicados por profesionales de la salud, éstos enfrentan limitaciones relacionadas con el estigma [11], las dificultades para verbalizar el malestar [12] y la falta de acceso oportuno a servicios especializados [13]. En este contexto, se ha observado que los individuos expresan voluntariamente sus pensamientos y emociones en redes sociales mediante publicaciones que podrían estar relacionados con su estado emocional [14]. Esto brinda la oportunidad de analizar los textos para crear modelos que detecten patrones de depresión.

---

<sup>1</sup><https://www.facebook.com>

<sup>2</sup>Características recurrentes en el uso del lenguaje, como la elección de palabras, la estructura gramatical y el estilo de escritura

<sup>3</sup><https://www.x.com>

<sup>4</sup><https://www.reddit.com>

Es importante señalar que la forma de comunicación cambia según la red social. Por ejemplo, *Reddit* suele elegirse porque su formato permite a los usuarios publicar textos más largos y detallados [15]. Esta diferencia hace que un modelo diseñado para una plataforma pierda precisión al usarse en otra distinta [16]. Por ello, este trabajo busca analizar estas publicaciones para detectar casos positivos de depresión y, al mismo tiempo, identificar patrones asociados con depresión en cada plataforma.

## 1.2. Pregunta de investigación

En este trabajo de investigación se investiga si ¿Los patrones lingüísticos que revelan depresión en redes sociales están asociados al tipo de red social?

## 1.3. Justificación

En los últimos años, ha aumentado el interés por abordar problemáticas de salud mental, como la depresión, mediante enfoques computacionales [17, 18]. Entre estas herramientas, el procesamiento del lenguaje natural (PLN), rama de la inteligencia artificial, ha demostrado su capacidad para identificar señales relacionadas con el estado mental de las personas a partir de textos publicados en redes sociales [19].

Las redes sociales se han convertido en un espacio donde las personas expresan cómo se sienten, generando textos que pueden ser analizados mediante diversas técnicas para identificar posibles casos de trastornos mentales, como ya se ha demostrado en trabajos previos [17, 18, 20, 21]. Estos textos pueden ser aprovechados desde un enfoque computacional, por ejemplo, usando técnicas de procesamiento del lenguaje natural, para entrenar modelos capaces de detectar casos positivos de depresión. Además, es importante analizar diferencias y similitudes en el uso del lenguaje asociado con depresión.

## 1.4. Motivación

Diversas organizaciones, como la Organización Mundial de la Salud [1] y el Instituto Nacional de Estadística y Geografía [22], han documentado una alta prevalencia de depresión y sus efectos sobre distintos grupos. Como ya se ha expuesto, este trastorno puede manifestarse de distintas formas, muchas de ellas fuera del entorno clínico. En los últimos años, las redes sociales se han convertido en un medio donde es posible observar síntomas asociados con enfermedades mentales [23]. Esto motiva el interés por estudiar cómo se manifiesta la depresión en distintas plataformas y qué diferencias lingüísticas existen entre ellas. La intención es detectar patrones léxicos que indiquen depresión para emitir alertas. Sin embargo, es importante recalcar que los métodos computacionales de detección no intentan reemplazar a los especialistas, más bien, se emiten alertas para canalizar a los posibles casos con un especialista con la finalidad de recibir un diagnóstico y un tratamiento oportuno.

## 1.5. Objetivo

### 1.5.1. Objetivo general

Implementar modelos computacionales para detectar depresión en diferentes plataformas de redes sociales empleando una perspectiva de clasificación de textos.

### 1.5.2. Objetivos específicos

- Desarrollar representaciones textuales de publicaciones en redes sociales a partir de características lingüísticas, con el fin de identificar rasgos depresivos en los usuarios.
- Entrenar diferentes algoritmos de aprendizaje automático a partir de representaciones textuales de publicaciones en redes sociales para la detección de rasgos de depresión.
- Analizar los patrones lingüísticos asociados con la depresión en distintas redes socia-

les, con el fin de compararlos entre plataformas usando representaciones basadas en bolsa de palabras.

## 1.6. Organización de la tesis

El presente documento se encuentra estructurado de la siguiente manera:

- El Capítulo 2 corresponde al marco teórico, donde se describen los términos que permiten familiarizar al lector con el contenido de este trabajo.
- En el Capítulo 3 se revisan trabajos previos relacionados con el tema, identificando qué se ha hecho hasta ahora y qué aspectos aún pueden explorarse.
- El Capítulo 4 describe la metodología seguida y se presentan los resultados obtenidos.
- Finalmente, en el Capítulo 5 se exponen las conclusiones del estudio. También se describe el trabajo futuro de esta investigación.

---

# Capítulo 2

## Marco Teórico

En este capítulo se presentan los conceptos relacionados con los temas abordados en el trabajo de tesis. El enfoque principal se centra en describir el marco referente al aprendizaje automático y procesamiento del lenguaje natural, los cuales constituyen la base de esta investigación.

### 2.1. Aprendizaje automático

El aprendizaje automático es un enfoque que permite aprender patrones complejos a partir de datos y usar ese conocimiento para hacer predicciones sobre nuevos casos [24]. Comúnmente, se describen dos tipos de aprendizaje automático: supervisado y no supervisado [25]; siendo el primero mencionado el que se utilizará en este trabajo.

#### 2.1.1. Aprendizaje Supervisado

El aprendizaje supervisado consiste en entrenar un modelo con ejemplos que ya incluyen una etiqueta, para que aprenda a identificar las relaciones entre los datos y pueda hacer predicciones sobre información nueva [26].

Dentro de este tipo de aprendizaje, destacan dos enfoques: la clasificación, aplicada en tareas como detectar si un correo es spam, y la regresión, que sirve para estimar valores

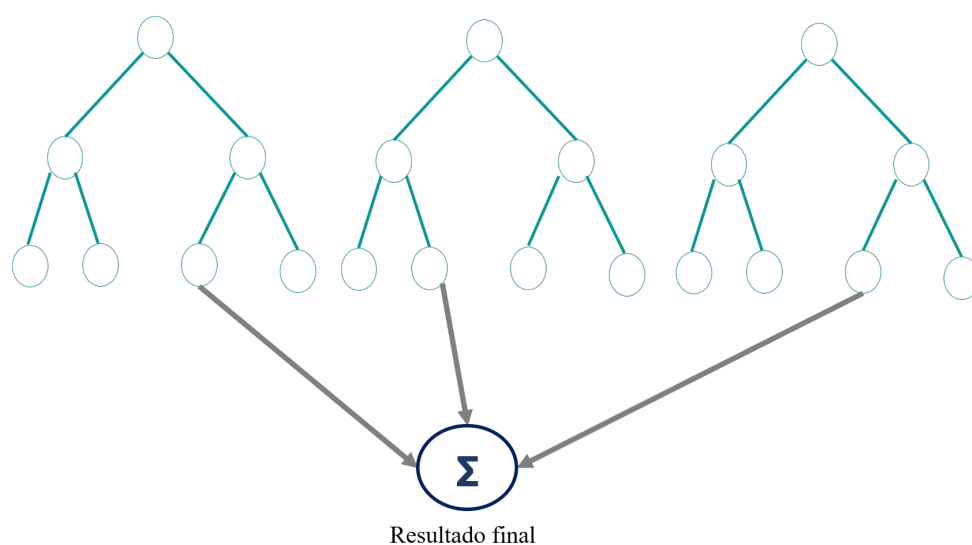
numéricos como el precio de un carro [27] En este trabajo de tesis, se utilizará un enfoque de clasificación para entrenar un modelo que permita detectar casos positivos de depresión.

## 2.1.2. Clasificadores

Un clasificador procesa las características de un dato de entrada para asignarle la clase a la que pertenece ese ejemplo [28]. En este trabajo de tesis se emplearon los cuatro algoritmos de clasificación descritos en las siguientes secciones.

### 2.1.2.1. Bosques aleatorios

Los bosques aleatorios, también conocidos como *Random Forest*, consisten en crear varios árboles de decisión y cada uno de ellos aprende de un subconjunto aleatorio diferente de los datos originales. Al final, se juntan las predicciones de todos los árboles para emitir un único resultado final [29]. En la Figura 2.1 se observa un ejemplo de bosques aleatorios.

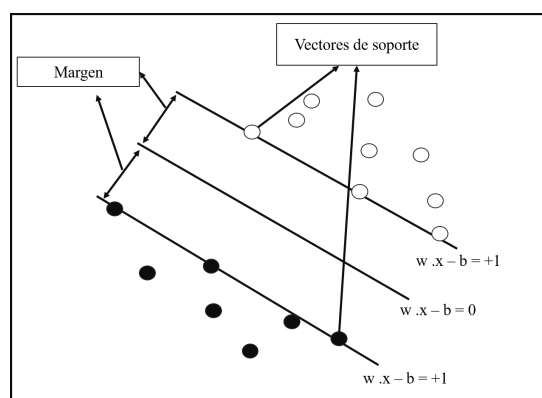


**Figura 2.1:** Representación del algoritmo de Bosques Aleatorios. Se ilustra cómo múltiples árboles de decisión generan una predicción final por mayoría de votos. Imagen basada en [30]

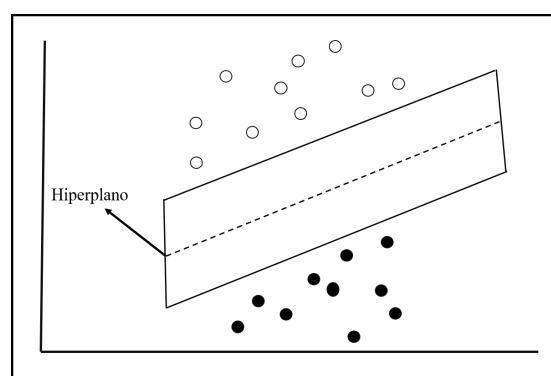
### 2.1.2.2. Máquinas de vectores de soporte

Una máquina de vectores de soporte (SVM, por sus siglas en inglés) representa las muestras como puntos en el espacio y busca una línea o superficie (hiperplano) que separe de forma óptima las distintas clases, centrándose en los ejemplos que se encuentran más próximos a la frontera de separación, denominados vectores soporte [31].

La Figura 2.2a ilustra este principio. En una SVM lineal, la frontera de decisión se define mediante la ecuación  $w \cdot x - b = 0$ , donde  $w$  representa el vector de pesos,  $x$  el vector de características de entrada y  $b$  el término de sesgo. El signo de  $w \cdot x - b$  determina la clase a la que pertenece cada instancia. Asimismo, las rectas  $w \cdot x - b = +1$  y  $w \cdot x - b = -1$  delimitan el margen de separación; los puntos ubicados sobre ellas corresponden a los vectores soporte y el ancho del margen está dado por  $2/\|w\|$ . Por su parte, la Figura 2.2b muestra de forma general cómo el hiperplano separa las dos clases en el espacio de características.



(a) Representación de una SVM lineal.

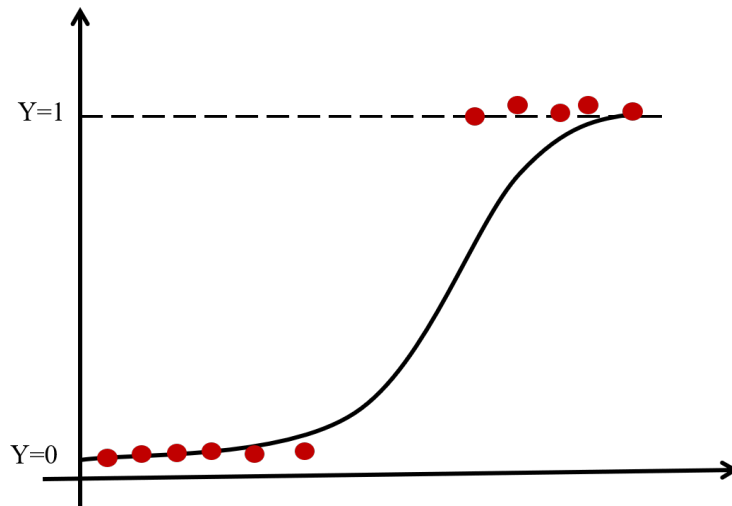


(b) Representación de un hiperplano

**Figura 2.2:** Máquina de vectores de soporte. Se muestra el modelo SVM y la representación de un hiperplano. Imagen basada en [32]

### 2.1.2.3. Regresión Logística

La regresión logística combina las variables de entrada y les aplica una función llamada sigmoide, la cual posee una curva característica en forma de "S"[33], como se muestra en la Figura 2.3.



**Figura 2.3:** Representación de Regresión Logística. Ejemplo de clasificación binaria (puntos rojos) y ajuste probabilístico mediante una curva sigmoide, donde el eje  $y$  representa la probabilidad estimada de la clase  $\mathbf{Y} = \mathbf{1}$  en función de la variable predictora  $\mathbf{x}$ . Imagen basada en [34]

Esta función transforma cualquier valor real en una salida acotada entre 0 y 1 [34], tal como se define en la Ecuación 2.1. De este modo, el modelo puede estimar la probabilidad de que un dato pertenezca a una clase determinada.

$$p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k)}} \quad (2.1)$$

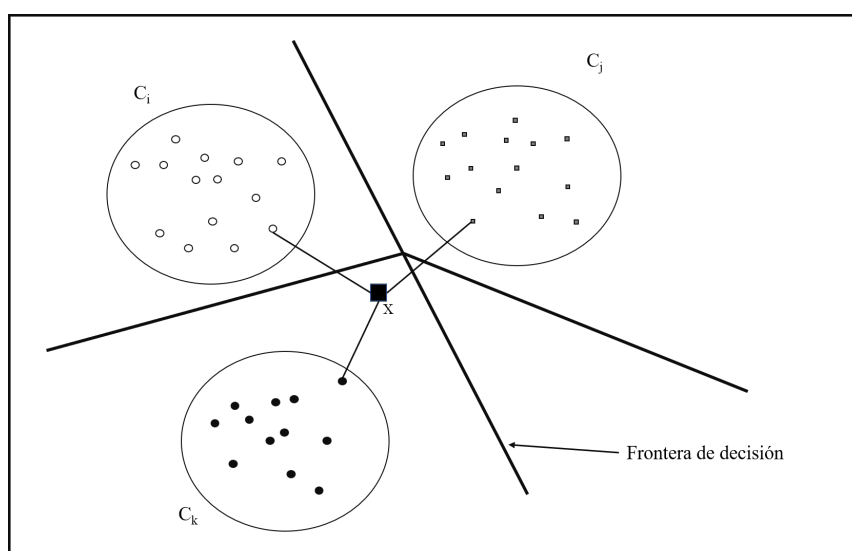
Donde:

- $p$ : probabilidad de que un dato pertenezca a una clase determinada.
- $e$ : número de Euler ( $\approx 2,71828$ ).
- $\beta_0$ : intercepto del modelo.
- $\beta_1, \beta_2, \dots, \beta_k$ : coeficientes asociados a cada variable de entrada.
- $x_1, x_2, \dots, x_k$ : variables de entrada del modelo.

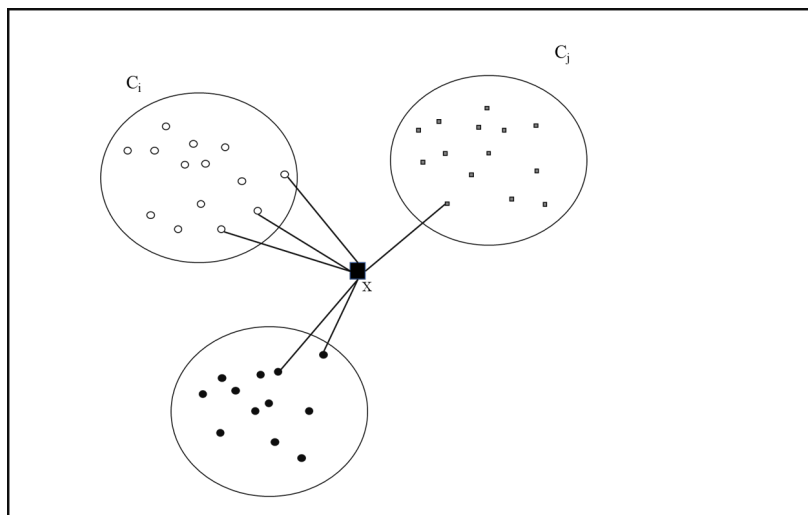
#### 2.1.2.4. Vecinos más cercanos

El algoritmo de vecinos más cercanos (KNN, por sus siglas en inglés) clasifica un nuevo dato calculando la distancia entre este y todos los datos del conjunto de entrenamiento. La clase asignada al nuevo dato será aquella a la que pertenezcan los datos más cercanos. El valor de  $k$  determina cuántos vecinos cercanos se consideran para realizar la clasificación; por ejemplo, con  $k = 1$ , la clase asignada será la del dato más cercano [35].

La Figura 2.4 muestra el caso donde  $K = 1$ : el punto  $x$  se clasifica según su vecino más cercano, lo que puede dar resultados inestables. Al aumentar el valor de  $K$ , la clasificación se vuelve más confiable. La Figura 2.5 muestra el caso con  $K = 6$ , donde el punto  $x$  se clasifica según la mayoría de sus seis vecinos más cercanos.



**Figura 2.4:** Clasificación KNN con  $K = 1$ . El punto  $x$  se asigna a la clase de su vecino más cercano, donde  $C_i$ ,  $C_j$  y  $C_k$  representan las diferentes clases. Imagen basada en [36]



**Figura 2.5:** Clasificación KNN con  $K = 6$ . El punto  $x$  se asigna a la clase mayoritaria entre sus seis vecinos más cercanos, donde  $C_i$ ,  $C_j$  y  $C_k$  representan las diferentes clases.

Imagen basada en [36]

### 2.1.3. Desempeño de modelos de clasificación supervisada

En esta sección, se hablará de diversas métricas utilizadas a lo largo de los experimentos para evaluar el rendimiento de los modelos.

#### 2.1.3.1. Matriz de confusión

En [37] definen a la matriz de confusión como una tabla de datos donde se comparan las predicciones con las clases reales. Destacan 4 conceptos importantes: los Verdaderos Positivos (VP) y Verdaderos Negativos (VN), que representan los aciertos del modelo en la clase positiva y negativa, respectivamente. Por otra parte, los errores se clasifican como Falsos Positivos (FP), cuando una instancia negativa es etiquetada como positiva, y Falsos Negativos (FN), que ocurren cuando el modelo clasifica una instancia positiva como negativa. En la Figura 2.6 se muestra un ejemplo de una matriz de confusión.

		Clase predicha	
		Positivo	Negativo
Clase verdadera	Positivo	VP	FN
	Negativo	FP	VN

**Figura 2.6:** Representación de una matriz de confusión. Se muestran los elementos que la componen [37].

### 2.1.3.2. Exactitud

La exactitud o *accuracy* es una métrica de evaluación que mide la proporción total de predicciones correctas realizadas por el modelo. Sin embargo, resulta engañosa en conjuntos de datos con clases desbalanceadas [38]. Su definición matemática se muestra en la Ecuación 2.2.

$$\text{Exactitud} = \frac{VP + VN}{VP + VN + FP + FN} \quad (2.2)$$

Donde

- VP = Verdaderos positivos
- VN = Verdaderos negativos
- FP = Falsos positivos
- FN = Falsos negativos

### 2.1.3.3. Precisión

Considerando la clase positiva, la precisión mide cuántas de las predicciones positivas de un modelo de clasificación son correctas, calculándose como el número de predicciones positivas correctas dividido entre el total de predicciones positivas [39]. En la Ecuación 2.3 se define matemáticamente. cómo se calcula la precisión.

$$\text{Precisión} = \frac{VP}{VP + FP} \quad (2.3)$$

### 2.1.3.4. Exhaustividad

En [40] se define a la exhaustividad como la capacidad del modelo para detectar todas las instancias que realmente pertenecen a la clase positiva. Matemáticamente se define en la Ecuación 2.4.

$$\text{Exhaustividad} = \frac{VP}{VP + FN} \quad (2.4)$$

### 2.1.3.5. Macro F1

De acuerdo con [27], el F1-score es la media armónica de la precisión y la exhaustividad, favoreciendo clasificadores que tienen valores similares de estas. En la Ecuación 2.5 [41] se describe la fórmula de la puntuación F1 y en la Ecuación 2.6 [42] se describe la fórmula del Macro F1.

$$F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Exhaustividad}}{\text{Precision} + \text{Exhaustividad}} \quad (2.5)$$

$$\text{Macro F1} = \frac{1}{N} \sum_{i=1}^N \frac{2 \cdot P_i \cdot R_i}{P_i + R_i} \quad (2.6)$$

donde  $N$  es el número de clases,  $P_i$  es la precisión para la clase  $i$  y  $R_i$  es la exhaustividad para la clase  $i$ .

### 2.1.3.6. Validación cruzada

La validación cruzada sirve para evaluar el rendimiento de un modelo más estable que dividir los datos en entrenamiento y prueba. En [43] se menciona que la variante más común es la validación cruzada de  $k$ -pliegues ( $k$ -fold). Se dividen los datos en  $k$  partes iguales y se entrenan múltiples modelos, rotando secuencialmente cada pliegue como conjunto de prueba mientras que los restantes se usan para el entrenamiento.

El desempeño final se obtiene promediando las métricas calculadas en cada uno de los  $k$ -pliegues, lo que proporciona una estimación más robusta y estable del modelo.

### 2.1.4. Selección de atributos

La selección de atributos tiene como finalidad identificar el subconjunto de atributos que mejor represente los datos estudiados, eliminando aquellos que resulten irrelevantes o redundantes, lo cual permite reducir la dimensionalidad del conjunto de datos, disminuir el costo computacional y favorecer una mejor generalización del algoritmo de aprendizaje [44]. Hay varias técnicas de selección de atributos. La técnica que se usó en este trabajo de investigación se basa en la prueba de Chi cuadrada, la cual se describe a continuación.

#### 2.1.4.1. Chi cuadrada

La prueba de Chi Cuadrada se utiliza para analizar si existe una relación de dependencia entre dos variables categóricas, sin embargo, no señala la intensidad de dicha relación ni cuál de las dos variables es la que influye sobre la otra [45]. La Ecuación 2.7 define de forma matemática la prueba chi-cuadra.

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} \quad (2.7)$$

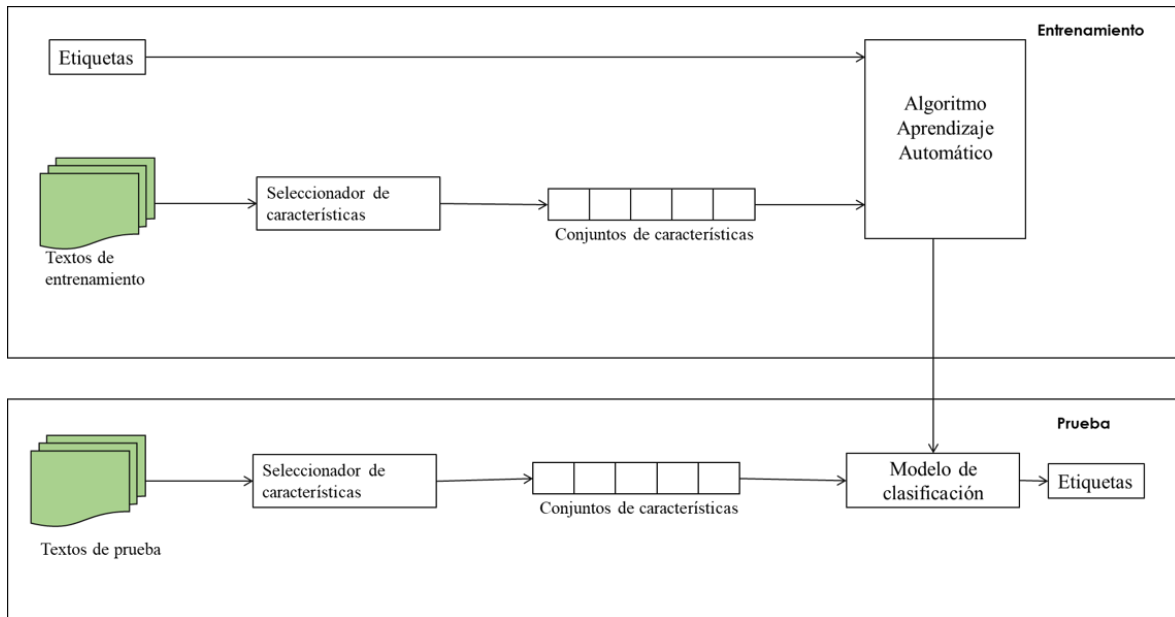
Donde  $O_i$  equivalen a los valores observados y  $E_i$  a los valores esperados.

## 2.2. Procesamiento del Lenguaje Natural

El procesamiento de lenguaje natural (NLP, por sus siglas en inglés) integra los conocimientos sobre cómo los humanos emplean el lenguaje para desarrollar las herramientas y técnicas que permitan a los modelos procesar el lenguaje natural y ejecutar tareas útiles [46]. Entre sus áreas de aplicación, se destaca la traducción automática, la búsqueda de información, y la extracción de información [47]. Otros autores también mencionan que se puede utilizar para el análisis de sentimientos [48] y tareas de clasificación de textos [49], siendo esta última la abordada en este trabajo de tesis.

### 2.2.1. Clasificación automática de texto

La clasificación automática de textos consiste en asignar de manera automática categorías predefinidas a documentos, frases o palabras según su contenido [49]. La Figura 2.7 muestra el flujo de la clasificación de textos dividido en dos etapas: en la etapa de entrenamiento se parte de textos previamente etiquetados, estos pasan por un selector de características que los transforma en conjuntos de características (o vectores), dichos conjuntos, junto con sus etiquetas, se introducen a un algoritmo de aprendizaje automático, el cual aprende patrones y genera un modelo de clasificación. En la etapa de prueba, los textos nuevos se procesan con el mismo selector de características para obtener sus vectores y estos se ingresan al modelo entrenado, que finalmente produce como salida las etiquetas predichas para cada texto.



**Figura 2.7:** Proceso de clasificación de textos. Se muestra lo que ocurre en las etapas de entrenamiento y prueba. Imagen basada en [50]

Para realizar la tarea de clasificación de textos, se requiere pasar el texto en lenguaje natural a representaciones entendibles por las máquinas. Existen diferentes representaciones. En esta tesis se utiliza la representación tradicional conocida como Bolsa de Palabras, la cual se describe a continuación.

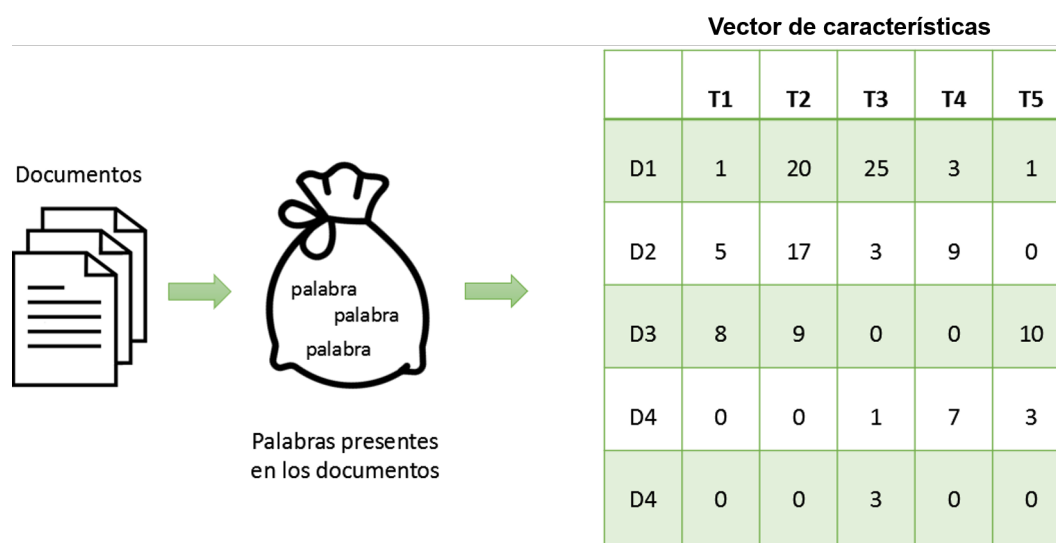
### 2.2.1.1. Bolsa de palabras

El modelo de bolsa de palabras (*bag-of-words*, *BoW*) transforma el texto de entrada en números. Representa un documento como la frecuencia de aparición de cada palabra, sin importar su orden ni la gramática [51].

En las Figuras 2.8 se muestra un ejemplo de representación binaria de bolsa de palabras, donde cada oración se codifica indicando la presencia (1) o ausencia (0) de cada término del vocabulario. La Figura 2.9 extiende este concepto al considerar la frecuencia de aparición de cada término, donde los valores representan cuántas veces aparece cada palabra en el documento correspondiente.

Oración	feliz	hoy	me	siento	triste
me siento triste	0	0	1	1	1
me siento feliz	1	0	1	1	0
hoy me siento triste	0	1	1	1	1

**Figura 2.8:** Ejemplo de representación binaria en bolsa de palabras. Se ilustra la codificación de tres oraciones en un vector de características. Elaboración propia



**Figura 2.9:** Representación de una bolsa de palabras. Se muestra el proceso de conversión de textos en una matriz de frecuencias de palabras, donde cada fila representa un documento y cada columna una palabra específica del vocabulario [52]

En las columnas aparecen los términos del vocabulario, los cuales corresponden al vocabulario compuesto por todas las palabras existentes en la colección de documentos (oraciones en este ejemplo). En cada celda se coloca el número de ocurrencias de cada término en el documento.

### 2.2.1.2. Pesado de términos

El pesado de términos es el proceso de asignar un valor numérico (peso) a cada término de un documento en una representación basada en bolsa de palabras. El esquema más sencillo

es la frecuencia de términos ( $TF$ ), que consiste en asignar como peso el número de veces que un término  $t$  aparece en un documento  $d$ , denotado como  $tf_{t,d}$  [53], de modo que un documento que menciona un término con mayor frecuencia se considera más relevante.

Sin embargo, un problema que surge con este esquema es que no distingue entre términos que aparecen con demasiada frecuencia y aquellos que aparecen en pocos documentos siendo realmente relevantes [40]. Para ello, se busca reducir el peso  $tf$  de un término en función de qué tan común es en la colección, utilizando la frecuencia de documento  $df_t$  que indica en cuántos documentos aparece el término  $t$  [53]. A partir de esta medida se define la frecuencia inversa de documento ( $idf$ ) aplicando logaritmo para obtener:

$$idf_t = \log \frac{N}{df_t} \quad (2.8)$$

donde  $N$  es el número total de documentos, de modo que un término poco común obtiene un valor alto de IDF mientras que uno muy frecuente obtiene un valor bajo.

Finalmente al combinar la frecuencia de términos con la frecuencia inversa de documento se obtiene el esquema de pesado TF-IDF que asigna a cada término  $t$  un peso en el documento  $d$  [40] dado por:

$$tf-idf_{t,d} = tf_{t,d} \times idf_t \quad (2.9)$$

De esta manera el peso asignado es mayor cuando un término aparece muchas veces en pocos documentos y más bajo cuando aparece en prácticamente todos los documentos [53].

---

# Capítulo 3

## Estado del Arte

En este capítulo, se exploran investigaciones sobre la detección de la depresión en redes sociales y foros en línea. En específico, se presentan investigaciones donde se utilizaron técnicas de aprendizaje automático y análisis de texto que abordan esta tarea.

### 3.1. La depresión en entornos digitales

Las redes sociales se han convertido en una actividad cotidiana a nivel global, ya que más de dos de cada tres personas en el mundo las utilizan cada mes [54]. A partir de lo que los usuarios publican en estas redes es posible identificar sus estados de ánimo [55], ya que los usuarios expresan sus emociones tanto con palabras directas como por la forma en la que escriben sus mensajes [56].

Algunas plataformas como *Facebook*, *Instagram*<sup>1</sup>, X (antes *Twitter*) y *Reddit* permiten observar este tipo de expresiones. En particular, plataformas como *Reddit* y X difieren en aspectos como el anonimato y la privacidad; *Reddit* ofrece un mayor anonimato, mientras que en X la norma es mantener las cuentas abiertas [57], lo cual podría influir en la forma en la que los usuarios se expresan.

El hecho de que las publicaciones en redes sociales puedan reflejar emociones, ha llevado a investigar la relación entre cómo escriben los usuarios y posibles síntomas de depresión.

---

<sup>1</sup><https://www.instagram.com/>

- Un ejemplo de ello es el estudio de titulado “*Individuals with depression express more distorted thinking on social media*” [58], en el que los autores analizaron el lenguaje de usuarios de *Twitter* con diagnóstico de depresión, comparándolo con un grupo aleatorio de usuarios de la misma plataforma. Basándose en los principios de la Terapia Cognitivo Conductual (TCC), una psicoterapia que busca identificar formas de pensar negativas o exageradas, como asumir que todo saldrá mal o que uno tiene la culpa de todo, diseñaron un vocabulario clínico compuesto por 241 secuencias de palabras las cuales representan distorsiones cognitivas. Entre ellas, el razonamiento emocional (creer que algo es cierto solo por cómo uno se siente), la personalización (asumir que los demás actúan negativamente por causa de uno mismo) y la sobregeneralización (sacar conclusiones negativas amplias a partir de pocos casos). El análisis de los *tweets* publicados por usuarios que reportaron haber sido diagnosticados de depresión mostró que estas secuencias de palabras aparecían con una mayor frecuencia en comparación con un grupo controlado.

Lo anterior indica que el lenguaje relacionado con la depresión en redes sociales no solo se nota por el tema del que se habla, sino también por la forma en la que son redactadas las publicaciones de los usuarios.

- Otro ejemplo es el trabajo titulado “El impacto de las redes sociales en la salud mental” [59], donde se analiza el uso de las redes sociales y el cómo puede afectar la salud mental en adolescentes y jóvenes en España.

Los autores compararon 31 fuentes bibliográficas, encontrando que el 90% de jóvenes entre los 16 y 24 años de edad usa redes sociales, siendo *Facebook* e *Instagram* las más utilizadas, y que su uso se asocia con un aumento de ansiedad y depresión. Además, señalan que aspectos como los *likes*, comentarios y el número de seguidores pueden afectar la autoestima de los usuarios, generando sentimientos negativos e insatisfacción corporal.

Al final, se menciona que es importante promover un uso más consciente y/o mode-

rado de las redes sociales, junto con la educación digital y atención psicológica para disminuir el impacto negativo que puede tener el uso excesivo en redes sociales.

- En el trabajo titulado “*The Impact of Social Media on the Mental Health of Adolescents and Young Adults: a systematic review*” [60] los autores realizan una revisión sistemática siguiendo el protocolo PRISMA y reportan una búsqueda entre las fechas de enero de 2000 a mayo de 2023, en bases de datos como MEDLINE/PubMed<sup>2</sup>, Scopus<sup>3</sup>, Google Scholar<sup>4</sup> y CINAHL<sup>5</sup> Los 11 estudios incluidos identifican una relación entre el uso de redes sociales como *Facebook, Instagram* y *Twitter* (ahora X), con síntomas depresivos, así como problemas de imagen corporal y trastornos alimenticios, especialmente en mujeres jóvenes.

Los autores señalan que no solo la cantidad de tiempo en redes sociales afecta la salud mental, sino también la calidad de las interacciones, y que un uso adecuado de estas plataformas puede ofrecer oportunidades de conexión y comunicación. Concluyen que existe una conexión entre el uso de redes sociales y la depresión en adolescentes, pero que se necesitan más estudios para poder hacer recomendaciones definitivas.

- De manera similar que el estudio previo, en “*Social Media Use and Depression in Adolescents: A Scoping Review*” [61], los autores realizaron una revisión de alcance (*scoping review*), en la que analizaron 43 artículos publicados entre 2012 y 2022, obtenidos de cinco bases de datos: PubMed, Scopus, ProQuest Psychology<sup>6</sup>, IEEE Xplore<sup>7</sup> y Google Scholar. Aproximadamente el 75 % de los estudios revisados encontraron una asociación entre el uso de redes sociales y la depresión en adolescentes, además de otros efectos negativos como ansiedad, mala calidad de sueño, baja autoestima y ansiedad por la apariencia. Ocho artículos reportaron que las mujeres presentan más síntomas depresivos que los hombres.

---

<sup>2</sup><https://pubmed.ncbi.nlm.nih.gov/>

<sup>3</sup><https://www.scopus.com/sources>

<sup>4</sup><https://scholar.google.com/>

<sup>5</sup><https://www.ebsco.com/products/research-databases/cinahl-database>

<sup>6</sup>[https://about.proquest.com/en/products-services/pq\\_psychology\\_journ/](https://about.proquest.com/en/products-services/pq_psychology_journ/)

<sup>7</sup><https://ieeexplore.ieee.org/Xplore/home.jsp>

En ese trabajo se concluye que, aunque existe una asociación entre el uso de redes sociales y la depresión en adolescentes, se necesitan más estudios para comprender mejor esta relación.

- En el estudio titulado “*A harsher reality for adolescents with depression on social media*” [62], se analizó la experiencia desde el punto de vista de adolescentes que están diagnosticados clínicamente con depresión comparándolos con aquellos que no tienen estos síntomas. En concreto, se llevó a cabo un estudio durante 100 días a 479 adolescentes, de los cuales, sólo el 16.9% presentaba depresión clínica. Los adolescentes deprimidos no participaron más en redes sociales, pero sí manifestaron inseguridad, rechazo y obsesión con las interacciones obtenidas.

Los autores mencionan que la depresión influye principalmente en la vivencia emocional en redes sociales, más que en cuánto tiempo las usan y que no siempre habrá señales o manifestaciones claras de malestar emocional.

- En el trabajo titulado “Detección de depresión a través de análisis textual utilizando aprendizaje automático” [63] se propone un modelo predictivo basado en análisis textual (técnicas automáticas para analizar grandes volúmenes de texto) y aprendizaje automático (capacidad de que una computadora aprenda de los datos sin ser programada explícitamente) para la detección oportuna de la depresión. Se trabajó con una muestra de 596 estudiantes universitarios de la Universidad César Vallejo, a quienes se les aplicaron dos instrumentos estandarizados: el Inventario de Depresión de Beck II (cuestionario de 21 preguntas que mide la severidad de la depresión) y el Test de Frases Incompletas de Sacks (prueba de 60 frases que la persona completa con lo primero que se le ocurra), para detectar la depresión en las personas a analizar.

Para la representación numérica de los textos se utilizó la técnica TF-IDF (medida que expresa qué tan relevante es una palabra dentro de un conjunto de documentos). Posteriormente se compararon siete algoritmos clasificadores y se eligió *Bayes Point Machine* (un método bayesiano de clasificación lineal) como el mejor clasificador para

el modelo final. El modelo obtuvo una exactitud del 93.2% y un AUC (área bajo la curva, que mide la capacidad del modelo para distinguir entre clases) de 91.7%. A partir de este modelo se desarrolló una aplicación web llamada ANTIDEP.

La autora concluye que la herramienta permitió reducir el tiempo de detección de dos semanas a menos de dos horas, y que este tipo de aplicaciones tiene el propósito de apoyar a la disminución de casos de suicidio.

- En el trabajo “Comportamiento de la mirada y análisis mediante aprendizaje automático de la depresión en la juventud: una revisión sistemática” [64], se lleva a cabo una revisión sistemática con el objetivo de comprender las características oculares (como el tamaño de la pupila o los parpadeos) con la depresión en personas jóvenes y explorar su uso en sistemas de aprendizaje automático.

La metodología sigue las recomendaciones PRISMA (protocolo que establece pasos claros para seleccionar y reportar estudios en revisiones sistemáticas). A partir de 1023 títulos identificados, se seleccionaron 79 artículos, de los cuales 36 se analizaron a detalle: 19 enfocados en la correlación entre el comportamiento ocular y la depresión, y 17 orientados a la detección automática de depresión mediante datos oculares.

Entre los estudios de detección revisados, las precisiones reportadas van del 75% al 94.52%, y el algoritmo más empleado fue SVM (máquinas de vectores de soporte, un método de clasificación supervisada). Las precisiones más altas con datos oculares se lograron con regresión logística (93.02%) y con AdaBoost (método que combina varios clasificadores débiles para mejorar la predicción), que alcanzó un 92.95% usando únicamente características de parpadeo.

Los autores concluyen que el seguimiento ocular es una vía prometedora para apoyar la detección de depresión, pero advierten que no existen protocolos estandarizados entre los estudios y que el acceso a los conjuntos de datos es limitado por la sensibilidad de la información clínica.

## 3.2. Enfoques computacionales para la detección de depresión

Anteriormente se mencionó que los usuarios expresan sus pensamientos y emociones a través de las redes sociales [14]. Esto da la oportunidad de analizar la información de forma computacional para detectar información de interés. A continuación, se muestran diversos trabajos que han abordado la detección de depresión mediante técnicas computacionales:

- Por ejemplo, en “*Predicting Depression via Social Media*” [65], se evalúa el uso de publicaciones en *Twitter* para entrenar modelos de aprendizaje automático orientados a la detección de depresión. Para ello, reclutaron 476 participantes mediante *Amazon Mechanical Turk*, quienes respondieron cuestionarios de depresión y compartieron sus perfiles de *Twitter*. De estos, 171 fueron identificados con depresión, y se recopilaron sus publicaciones durante el año previo al inicio reportado del trastorno. Los autores analizaron diferentes aspectos, como la actividad en la plataforma (volumen de publicaciones por día), la interacción con otros usuarios (respuestas, *retweets*), el tono emocional de las publicaciones (positivo o negativo) y uso de léxico relacionado con la depresión. Los resultados mostraron que los usuarios con depresión publicaban menos, usaban más palabras negativas, se referían más a sí mismos y empleaban con mayor frecuencia términos asociados con el trastorno.

Con esta información, entrenaron un modelo que alcanzó una exactitud del 70% y una precisión del 74%, concluyendo que la actividad en redes sociales contiene señales útiles para identificar personas vulnerables a la depresión.

- En el trabajo “*Assessing Mental Health Signals Among Sexual and Gender Minorities Using Twitter Data*” [66], se analiza el estado emocional de personas pertenecientes a minorías sexuales y de género (*Sexual and Gender Minorities*, SGM) a través de datos de *Twitter*. El proceso siguió cinco etapas: recolección desde cuatro fuentes dis-

tintas, preprocesamiento (filtrado de idioma y geolocalización por estado de EE.UU.), clasificación de *tweets*, categorización de usuarios en 11 subgrupos SGM y análisis emocional. Se obtuvieron 27,303,446 *tweets* únicos. Después se filtraron por idioma inglés, quedando 2,296,923 para análisis.

Para clasificar estos *tweets*, se entrenaron tres algoritmos: *Random Forest*, SVM y *CNN* (red neuronal convolucional), utilizando TF-IDF para los dos primeros y *GloVe* (representaciones vectoriales de palabras preentrenadas) para la *CNN*. Los modelos se validaron con 150 *tweets* anotados manualmente, donde *Random Forest* obtuvo el mejor desempeño en ambas tareas: en la primera (distinguir si un tuit realmente trataba sobre temáticas SGM o si la palabra clave aparecía con otro significado) alcanzó una precisión de 95% y un F1 de 95%, y en la segunda (identificar si el usuario se autoidentificaba como SGM) logró una precisión de 89% y un F1 de 86%.

Al aplicar este modelo a los 2,296,923 *tweets*, se identificaron 8439 usuarios potencialmente autoidentificados. Estos fueron revisados manualmente, confirmando 7,033 como personas realmente pertenecientes a minorías sexuales y de género.

El análisis emocional realizado con *LIWC* (*Linguistic Inquiry and Word Count*), una herramienta que mide el porcentaje de palabras asociadas a distintas emociones en un texto, reveló que las personas SGM expresaban tanto más emociones negativas como positivas en comparación con usuarios no SGM. Además, mostraban un estado emocional más positivo en lugares donde existen leyes antidiscriminatorias.

Estos resultados destacan el valor de las redes sociales como herramienta para observar el estado emocional en grupos vulnerables.

- Otros estudios como “*Inferring Social Media Users’ Mental Health Status from Multimodal Information*” [67] tuvo como objetivo desarrollar un enfoque computacional que permitiera deducir el estado mental de los usuarios de redes sociales mediante el análisis de información multimodal. Para ello, se reunieron datos de *Flickr*<sup>8</sup> donde de

---

<sup>8</sup><https://www.flickr.com/>

cada publicación se extrajeron datos visuales (color, brillo, textura, escenas, rostros y objetos detectados en la imagen), lingüísticas (estilo, n-gramas y categorías LIWC del texto que acompaña la imagen) y metadatos de la publicación (hora de publicación, número de vistas y datos del archivo de imagen).

A partir de los datos obtenidos, los autores entrenaron modelos de aprendizaje automático para realizar clasificación binaria en dos puntos de vista: (i) distinguir usuarios saludables de aquellos con alguna condición mental, y (ii) distinguir usuarios sanos y usuarios con posibilidades a desarrollarla.

Los resultados mostraron que los resultados obtenidos a partir de modelos entrenados con diversas fuentes de información (modelo multimodal) tuvieron un mejor rendimiento que aquellos que utilizaron solamente una modalidad (texto, imagen o comportamiento).

Los autores concluyen que existen diferencias de comportamiento entre los grupos: los usuarios con problemas de salud mental tienden a publicar imágenes más oscuras, con alto contraste, en espacios interiores y con menos rostros. Además, señalan que combinar información visual, lingüística y de metadatos mejora la predicción del estado de salud mental en comparación con usar una sola fuente de información.

- Otro trabajo centrado en el análisis de textos es “*Dilated LSTM with Attention for Classification of Suicide Notes*” [68], en el que se emplearon redes neuronales con mecanismos de atención para detectar patrones en notas suicidas y otros escritos vinculados a la salud mental, como la depresión.

Primeramente se realizó un análisis lingüístico con LIWC para identificar diferencias entre los tres tipos de textos, encontrando que las notas de depresión contenían más emociones negativas. Posteriormente, se entrenaron varios modelos de clasificación: el modelo propuesto (*Dilated LSTM* con atención) logró una precisión del 87.34% superando al *BiLSTM* con atención que obtuvo un 82.27%.

Los autores concluyen que el modelo propuesto presta atención a las mismas características lingüísticas identificadas en el análisis con LIWC.

- En “*Depression Detection from Social Media Text Analysis using Natural Language Processing Techniques and Hybrid Deep Learning Model*” [69], proponen un modelo híbrido de aprendizaje profundo llamado FCL (*FastText* + *CNN* + *LSTM*) para detectar depresión a partir de texto publicado en redes sociales.

El modelo utiliza *embeddings FastText* para representar el texto, una red neuronal convolucional (*CNN*) para extraer características globales y una red *LSTM* para capturar dependencias a lo largo del texto. Los experimentos se realizaron sobre dos conjuntos de datos: el primero contiene publicaciones de *Reddit* que fueron balanceadas a 13,000 registros (6,500 de depresión y 6,500 sin depresión); el segundo, obtenido del repositorio de *Kaggle*<sup>9</sup>, contiene *tweets* que originalmente estaban desbalanceados y fueron filtrados a 6,164 registros (3,082 por categoría) para mantener el equilibrio entre clases. Ambos conjuntos pasaron por un preprocesamiento que incluyó tokenización, conversión a minúsculas, eliminación de signos de puntuación y palabras vacías, y lematización.

Se comparó el modelo propuesto contra *CNN* y *LSTM-RNN* utilizando distintos *embeddings* (*Word2Vec*<sup>10</sup> y *GloVe*<sup>11</sup>). Los resultados mostraron que el modelo FCL alcanzó una exactitud del 87% en el conjunto de *Reddit* y del 88% en el de *Twitter*, superando a los modelos anteriores en ambos casos. Los autores concluyen que la combinación de *FastText* con *CNN* y *LSTM* permite detectar depresión con mayor exactitud que los enfoques existentes, tanto en volúmenes pequeños como grandes de datos.

- En “*An hybrid deep learning approach for depression prediction from user tweets*”

---

<sup>9</sup><https://www.kaggle.com/>

<sup>10</sup>Modelo desarrollado por Google basado en redes neuronales, aprende representaciones de palabras a partir del contexto local.

<sup>11</sup>Modelo de representación textual que genera vectores numéricos a partir de las estadísticas globales de co-ocurrencia en un corpus.

using feature-rich CNN and bi-directional LSTM“ [70], proponen un modelo híbrido que combina una red neuronal convolucional (*CNN*) con una *LSTM* bidireccional (*biLSTM*) para clasificar *tweets* como depresivos o no depresivos. El conjunto de datos utilizado está compuesto por tres subconjuntos: D1 con 292,564 *tweets* de 1,402 usuarios depresivos, D2 con más de 10 mil millones de *tweets* de usuarios no depresivos, y D3 con *tweets* que contienen el término "*depress*" de forma general.

En este estudio se utilizaron los subconjuntos D1 y D2. Los datos pasaron por un preprocesamiento que incluyó la eliminación de URLs, hashtags, emojis, menciones de usuario, caracteres especiales y palabras vacías, además de tokenización y *stemming*. Para representar el texto se utilizaron *word embeddings* con una dimensión de 300. El modelo fue comparado contra una *CNN* y una *RNN* por separado, y alcanzó una exactitud del 94.28 %, una precisión del 96.99 % y un F1 del 94.78 %, superando a ambos modelos individuales.

Además, una prueba t confirmó que existe una diferencia significativa en la longitud promedio de los *tweets* entre usuarios depresivos y no depresivos. Los autores concluyen que la combinación de *CNN* con *biLSTM* permite extraer tanto características locales como información contextual, logrando mejores resultados que los enfoques previos.

- En “An evolutionary approach for depression detection from Twitter big data using a novel deep learning model with attention based feature learning mechanism“ [71], proponen un modelo de aprendizaje profundo llamado *RNT-OLSTM* para detectar depresión en *tweets*, combinando *ResNet50* (una red convolucional de 50 capas para extraer patrones del texto) con un *LSTM* optimizado (una red recurrente capaz de capturar el contexto secuencial de las palabras).

Se utilizaron dos datasets públicos de *Twitter*: DS1 de *Kaggle* con más de 4,000 *tweets* y DS2 con 2,000 *tweets*. El preprocesamiento incluyó eliminación de emojis, stopwords, puntuación, conversión a minúsculas, lematización y tokenización. Pa-

ra representar el texto como vectores numéricos, combinaron TF-IDF (que mide la importancia estadística de cada palabra) con *Word2Vec* (que captura relaciones de significado entre palabras), ya que según los autores ambos enfoques se complementan entre sí.

El modelo *RNT-OLSTM* alcanzó una exactitud de 99.55 % en DS1 y 99.23 % en DS2, superando a métodos convencionales de aprendizaje automático y profundo. Los autores concluyen que combinar el balanceo de datos, la extracción de características con ResNet50 y la optimización del *LSTM* permite obtener mejores resultados que los enfoques existentes.

### 3.3. Conjuntos de datos para detección de depresión

En esta sección se describen diversos conjuntos de datos relacionados con la salud mental, los cuales contienen textos de plataformas como *Twitter* y *Reddit* con etiquetas que indican la presencia o ausencia de trastornos relacionados con alteraciones de la salud mental.

#### 3.3.1. *Reddit Self-reported Depression Diagnosis*

Un primer conjunto de datos explorado en este trabajo de investigación fue el llamado “*Reddit Self-reported Depression Diagnosis (RSDD)*” [72], el cual está conformado por publicaciones de 9,210 usuarios de *Reddit* que afirmaron haber sido diagnosticados con depresión (usuarios diagnosticados), junto con 107,274 usuarios sin indicios de depresión (usuarios de control). Además, se aplicó un filtro de actividad, descartando a los usuarios que contaban con menos de 100 publicaciones previas al post en el que reportan el diagnóstico.

Para su construcción, se seleccionaron publicaciones que incluían patrones lingüísticos de alta precisión (por ejemplo, “me diagnosticaron con...”) y menciones cercanas al término “depresión”. Se excluyeron publicaciones ambiguas o referidas a otras personas, así como todas las publicaciones relacionadas con salud mental, con el fin de evitar sesgos.

Los usuarios de control fueron seleccionados buscando similitud con los usuarios diagnosticados en cuanto a su comportamiento general en *Reddit* (distribución de publicaciones por *subreddit* y volumen total de publicaciones). El conjunto de datos está dividido en tres particiones, las cuales se describen en la Tabla 3.1.

**Tabla 3.1:** Distribución de clases en el conjunto de datos *RSDD*. Se muestra el número de usuarios por categoría en los conjuntos de entrenamiento, validación y prueba

Entrenamiento		Validación		Prueba	
Diagnosticados	Control	Diagnosticados	Control	Diagnosticados	Control
3,070	35,753	3,070	35,746	3,070	35,775

### 3.3.2. Dreaddit

*Dreaddit* [73] es un conjunto de datos creado a partir de la necesidad para analizar la presencia de estrés en publicaciones de *Reddit*, y así, facilitar el entrenamiento y la evaluación de modelos supervisados que detecten trastornos de la salud mental.

Se seleccionaron 187,444 publicaciones de 10 *subreddits*, clasificados en cinco dominios que fueron: abuso, ansiedad, problemas financieros, trastorno de estrés postraumático y social. Las publicaciones tienen un aproximado de 420 tokens, para que se puedan capturar de manera más sencilla expresiones de estrés más detalladas que en otras redes sociales.

Para el análisis supervisado, fueron seleccionados 3,553 fragmentos pertenecientes de cinco oraciones, extraídos de 2,929 publicaciones, las cuales fueron etiquetados manualmente por trabajadores de *Amazon Mechanical Turk* como “estrés” o “no estrés”.

Estos fragmentos fueron divididos en 2,838 para entrenamiento, de los cuales 1,465 corresponden a estrés y 1,373 a no estrés; y 715 para prueba, siendo 375 de estrés y 340 de no estrés. También se aplicaron técnicas de validación cruzada y fueron evaluadas diversas arquitecturas de modelos, tales como Regresión Logística, Redes Neuronales Convolucionales

(*CNN*) y *BERT*.

Si bien el enfoque de la presente tesis, se centra en la detección de depresión, *Dreaddit* ofrece una perspectiva más amplia al capturar diversas formas de expresión relacionadas con el malestar emocional, lo que permite considerar su uso en el estudio de otros trastornos psicológicos y situaciones asociadas al estrés. En la Tabla 3.2 se muestra el conjunto de datos etiquetado.

**Tabla 3.2:** Distribución de clases en el conjunto de datos *Dreaddit*. Se muestra el número de publicaciones etiquetadas por categoría en los conjuntos de entrenamiento y prueba

Train		Test	
Estrés	No estrés	Estrés	No estrés
1,465	1,373	375	340

### 3.3.3. *eRisk*

Tras el análisis realizado, se identificaron dos conjuntos de datos particularmente adecuados para el desarrollo del trabajo de tesis propuesto. Estos fueron utilizados en las competencias de las ediciones *eRisk 2017* [74] y *2018* [75], diseñadas para impulsar la investigación en la detección temprana de riesgos psicológicos, especialmente depresión, y en 2018 se incorporó también la anorexia como nueva categoría, utilizando textos publicados por usuarios en *Reddit*.

Cada usuario del conjunto de datos tiene una secuencia de publicaciones ordenadas cronológicamente, segmentada en diez fragmentos (también conocidos como chunks), haciendo posible la simulación de un entorno realista de análisis progresivo.

En *eRisk 2017*, el conjunto usado para entrenamiento incluyó 83 usuarios con depresión y 403 usuarios controlados mientras que el conjunto usado para pruebas estuvo conformado por 52 usuarios con depresión y 349 usuarios controlados, teniendo en total de más de

500,000 publicaciones. Los usuarios controlados son participantes que no presentan indicios de diagnóstico, pero igual mantienen una actividad similar en *Reddit* y en algunos casos, intervienen en conversaciones sobre salud mental, lo que añade complejidad al proceso de clasificación.

En *eRisk 2018*, se incorporaron los datos del año anterior al conjunto de entrenamiento, ampliándose y así alcanzando 135 usuarios con depresión y 752 controlados, además, de un nuevo conjunto de pruebas que se componen por 79 usuarios diagnosticados con depresión y 741 controlados.

Ambos conjuntos presentan un desequilibrio de clases, con predominio de usuarios controlados frente a los usuarios con depresión y/o anorexia (clase positiva), y su evaluación se realizó mediante la métrica *ERDE* (*Early Risk Detection Error*), diseñada específicamente para valorar la capacidad de los sistemas no solo para acertar en sus predicciones, sino también para hacerlo en el menor tiempo posible.

A diferencia de métricas tradicionales como la precisión o el F1, *ERDE* penaliza los aciertos tardíos, asignando un mayor coste cuanto más se tarda en emitir una decisión correcta. Esto permite evaluar de forma más realista el rendimiento de los modelos en escenarios sensibles, como el monitoreo de trastornos mentales, donde la oportunidad de la detección puede ser tan importante como la exactitud.

Las características generales de los usuarios con y sin diagnóstico de depresión que conforman los conjuntos de datos de entrenamiento y prueba pueden consultarse en la Tabla 3.3, donde se detallan aspectos como el número total de publicaciones, el promedio de envíos por usuario y la extensión media de los textos.

**Tabla 3.3:** Distribución de clases en *eRisk 2017* y *eRisk 2018*. Se muestra el número de usuarios etiquetados por categoría en los conjuntos de entrenamiento y prueba

Clase	<i>eRisk 2017</i>		<i>eRisk 2018</i>	
	Train	Test	Train	Test
Depresión	83	52	135	79
No-depresión	403	349	752	741
Total	486	401	887	820

#### 3.3.4. *Binary Labeled Depression Corpus of 60,000 English tweets*

El tercer conjunto de datos usado en el presente trabajo es *Binary Labeled Depression Corpus of 60,000 English tweets* [76] el que está compuesto por 60,172 *tweets*, de los cuales 21,953 están relacionados con depresión ("*depressed*") y 38,219 como no relacionados con depresión ("*non-depressed*"). Estos *tweets* incluyen una extensa variedad de sinónimos en inglés relacionados tanto con la depresión como la felicidad.

Es importante mencionar que este conjunto de datos no incluye particiones predeterminadas de entrenamiento y prueba. Por sus características y disponibilidad, este conjunto de datos representa una opción adecuada para complementar el análisis propuesto en este trabajo de tesis, especialmente al enfocarse en publicaciones breves y espontáneas propias de plataformas como *Twitter* (ahora X).

**Tabla 3.4:** Distribución de clases en el *Binary Labeled Depression Corpus*. Se muestra el número de publicaciones etiquetadas por categoría

<b>Clase</b>	<b>Total de publicaciones</b>
Depresión	21,953
No-depresión	38,219
Total	60,172

---

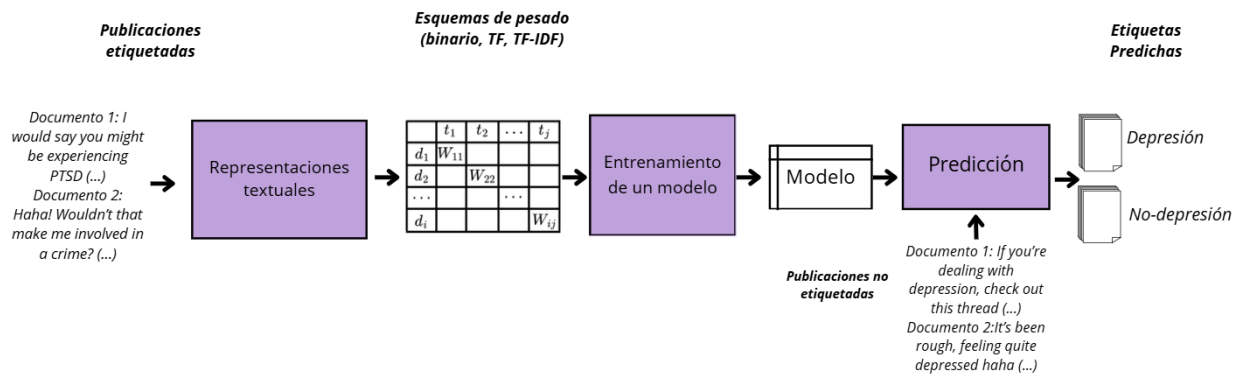
# Capítulo 4

## Experimentos y resultados

En este capítulo se presentan los experimentos realizados y los resultados obtenidos en la detección de depresión en redes sociales. Se describen los métodos utilizados y se analizan los resultados principales.

### 4.1. Metodología

La metodología desarrollada en el presente trabajo se organizó en tres etapas sucesivas enfocadas en la evaluación de diferentes formas de procesar y clasificar textos relacionados con la detección de depresión. A continuación, son descritos los pasos generales del enfoque adoptado, también mostrado en la Figura 4.1.



**Figura 4.1:** Metodología adoptada para la clasificación de textos. Se muestra el flujo desde la representación textual hasta la predicción<sup>1</sup>.

## 1. Construcción de representaciones textuales

Fueron utilizados tres conjuntos de datos compuestos por textos provenientes de diferentes fuentes: *eRisk 2017*, *eRisk 2018* y un conjunto de datos de publicaciones en inglés de la red social *Twitter*. Cada texto está clasificado en una de las dos categorías: *depresión* o *no depresión*.

Se aplicaron distintos esquemas de pesado (binario, TF y TF-IDF) tanto en el texto original, como en versiones donde se aplicó un preprocesamiento. También, se aplicó un proceso de selección de atributos, con el objetivo de conservar un número limitado de palabras representativas y reducir la dimensionalidad del vocabulario.

## 2. Entrenamiento de un modelo

Las representaciones generadas se usaron como entradas para cuatro algoritmos de aprendizaje automático, con el propósito de diferenciar textos asociados a la depresión de los que no están asociados a la depresión.

El entrenamiento tomó en cuenta el preprocesamiento y la reducción de características en el rendimiento de los modelos. La evaluación se centró en la puntuación F1 de la

<sup>1</sup>Los textos de los documentos usados en la imagen fueron parafraseados por motivos de privacidad.

clase positiva, pues esta métrica es la más usada en este dominio.

### 3. Análisis de resultados en diversas redes sociales

Los resultados de los clasificadores se analizaron considerando las diferencias entre los tres conjuntos de datos empleados. Se evaluaron los efectos de las distintas configuraciones de selección de atributos sobre el desempeño de los modelos. Finalmente, se identificaron las palabras más relevantes vinculadas a cada clase y se generaron nubes de palabras que visualizan gráficamente los términos característicos de los textos clasificados.

## 4.2. Experimentos basados en clasificación

La identificación temprana de indicios vinculados a la depresión permite realizar intervenciones oportunas. En este sentido, el análisis textual se presenta como una herramienta favorable para entender cómo las personas comunican sus emociones. Por ello, esta sección expone experimentos en los que se entrenan clasificadores utilizando varias técnicas de preprocesamiento de texto y evaluar su eficacia en la detección de casos positivos de depresión.

### 4.2.1. Configuración experimental

En este apartado se describen las configuraciones empleadas para los cuatro experimentos, organizadas en tres etapas consecutivas: la transformación del texto, la clasificación inicial y la selección de características con una segunda fase de clasificación, implementadas con el marco *Sklearn*<sup>2</sup> para los experimentos.

- **Representación del texto:** para convertir los textos en un formato interpretable por los modelos, se aplicaron tres esquemas distintos: presencia/ausencia de términos (codificación binaria), frecuencia de términos (TF) y TF-IDF. En el caso de TF y

---

<sup>2</sup><https://scikit-learn.org/>

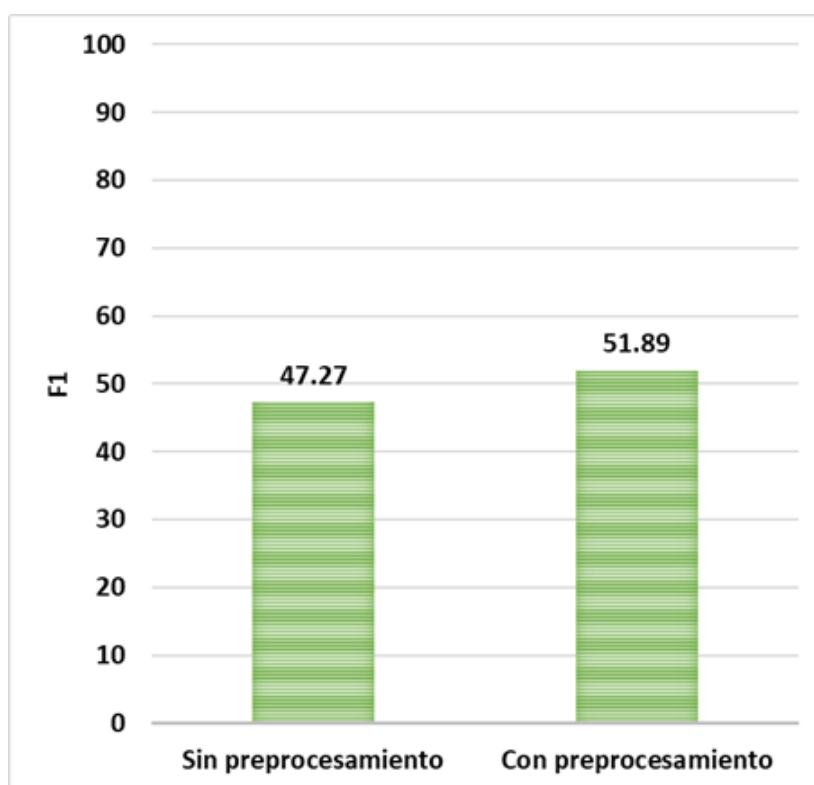
TF-IDF, se utilizó normalización L2. En todos los esquemas, el máximo número de características se estableció a 10,000 términos y se descartaron aquellos que aparecían en menos de tres documentos.

- **Clasificación inicial:** las representaciones generadas se utilizaron como entrada para cuatro modelos de clasificación. Las configuraciones finales fueron las siguientes:
  - **Máquina de Vectores de Soporte (SVM):** núcleo sigmoide,  $C = 10$ ,  $\gamma = 1$ , incorporando balanceo de clases.
  - **Bosques Aleatorios (*Random Forest*):** criterio de entropía, 100 árboles, sin límite de profundidad, balanceo de clases y *random state* de 42.
  - **Regresión Logística:** penalización  $l_2$ ,  $C = 10$ , optimizador *lbfgs*, balanceo de clases, 1000 iteraciones máximas y *random state* de 42.
  - **K-Vecinos más Cercanos (KNN):** 3 vecinos, distancia euclidiana, pesos uniformes.
- **Particiones implementadas:** los conjuntos de datos de *eRisk 2017* y *eRisk 2018* ya incluyen particiones predefinidas de entrenamiento y prueba, en cambio con el conjunto de datos de *Twitter* fue particionado manualmente, se usó el 80% para entrenamiento y 20% para prueba.
- **Selección de características y nueva clasificación:** para las representaciones generadas por TF-IDF, se usó la prueba de Chi-cuadrado para seleccionar solo las características más importantes. Se comparó el rendimiento eligiendo grupos de 500, 1000, 2000, 3000, 4000 y 5000 atributos. Con estas selecciones, se entrenaron nuevamente los modelos utilizando las mismas configuraciones descritas antes.
- **Métricas de clasificación:** se reporta el F1 de la clase positiva.

### 4.2.2. Experimento 1: El impacto del preprocesamiento

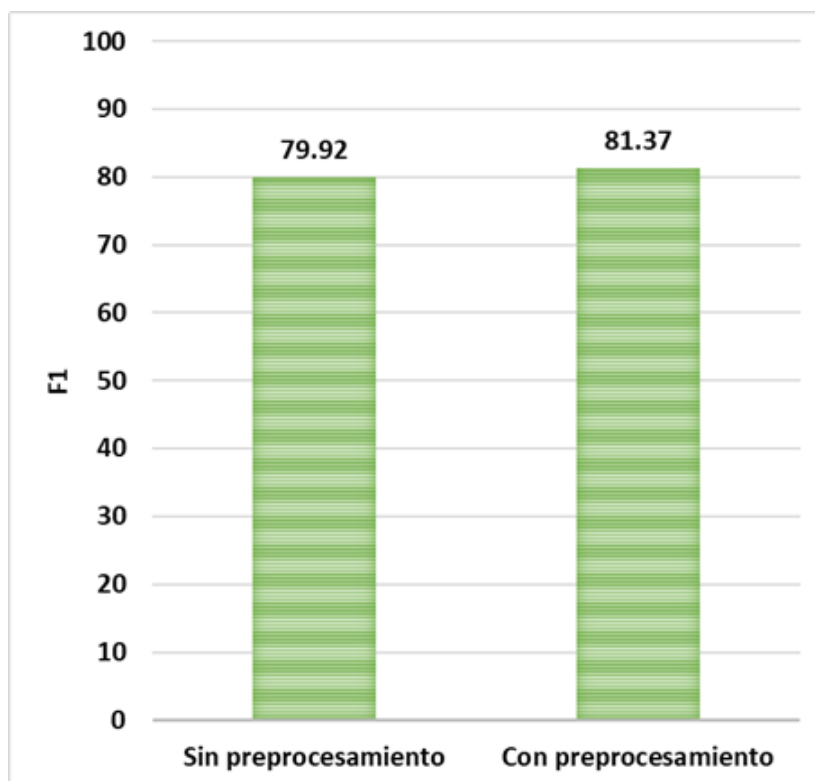
Como primer experimento, se busca evaluar los efectos que tiene el usar o no el preprocesamiento (eliminación de urls, conversión a minúsculas y la eliminación de palabras vacías) aplicando la frecuencia de términos y una SVM al conjunto de datos *eRisk 2018* y al conjunto de *Twitter*.

En la Figura 4.2 se presentan los resultados obtenidos con el conjunto de datos *eRisk 2018*. Se puede observar que el preprocesamiento ayudó a mejorar el rendimiento de SVM, que pasó de 47.27% a 51.89% en F1.



**Figura 4.2:** Análisis del impacto del preprocesamiento en la colección *eRisk 2018*. Se usó frecuencia normalizada como esquema de pesado de términos

En la Figura 4.3 se presentan los resultados obtenidos con el conjunto de datos *Twitter*. En este caso, el preprocesamiento también mejoró el rendimiento del modelo SVM, pasando de 79.92 % a 81.37 %.



**Figura 4.3:** Análisis del impacto del preprocesamiento en la colección de *Twitter*. Se usó frecuencia normalizada como esquema de pesado de términos

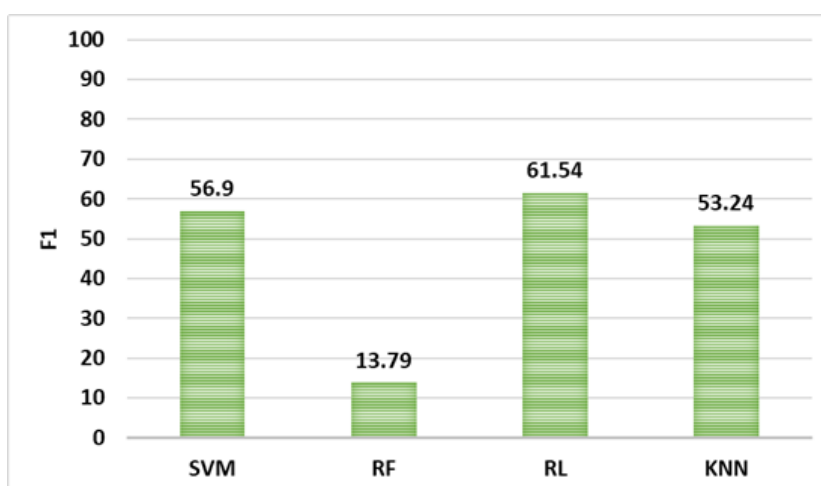
En ambos conjuntos de datos se puede observar que el preprocesamiento contribuyó a mejorar el rendimiento del modelo evaluado. Esto confirma que limpiar y preparar los textos antes de ser clasificados ayuda a obtener mejores resultados.

Es por ello que en los siguientes experimentos se decidió trabajar únicamente con los datos preprocesados.

### 4.2.3. Experimento 2: Evaluación de diferentes algoritmos de aprendizaje

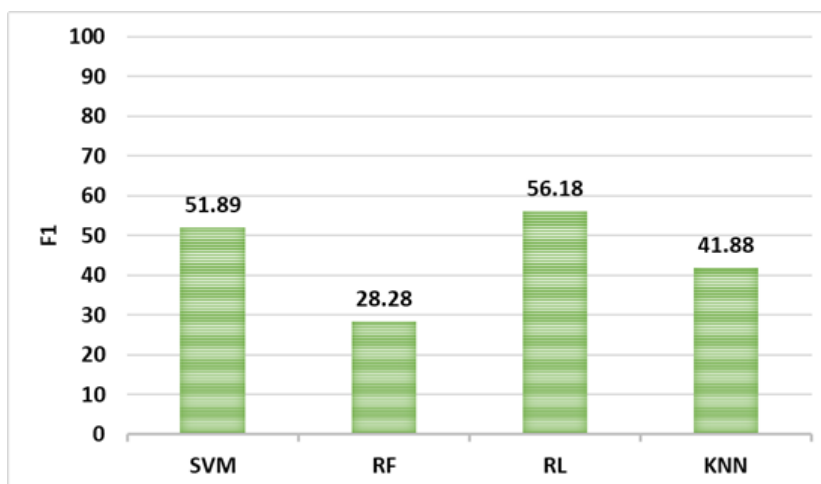
En este experimento se busca comparar el desempeño de cuatro algoritmos de clasificación: máquinas de vectores de soporte (SVM), bosques aleatorios (RF), vecinos más cercanos (KNN) y regresión logística (RL) (véase el apartado 4.2.1 para las configuraciones de los algoritmos).

En la figura 4.4 se presentan los resultados obtenidos para el conjunto de datos *eRisk 2017*. El mejor resultado es obtenido por RL con 61.54 %, seguido de SVM con 56.90 % y KNN 53.24 %. RF obtuvo el valor más bajo con 13.79 %. La diferencia entre el mejor y el peor clasificador es de casi 48 %.



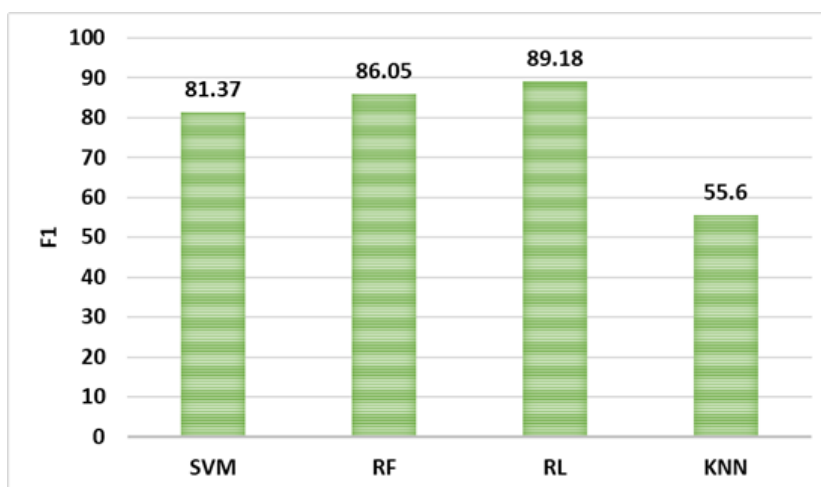
**Figura 4.4:** Comparación del desempeño de diferentes clasificadores en el conjunto de datos *eRisk 2017*. Se utilizó frecuencia normalizada como esquema de pesado de términos

En la figura 4.5 se presentan los resultados obtenidos para el conjunto de datos *eRisk 2018*. Nuevamente RL obtuvo el mejor resultado con 56.18 %, seguido de SVM con 51.89 %. KNN alcanzó 41.88 % y RF quedó en último lugar con 28.28 %.



**Figura 4.5:** Comparación del desempeño de diferentes clasificadores en el conjunto de datos *eRisk 2018*. Se utilizó frecuencia normalizada como esquema de pesado de términos

En la figura 4.6 se presentan los resultados obtenidos para el conjunto de datos de *Twitter*. RL fue otra vez el mejor clasificador con 89.18 %. RF obtuvo 86.05, SVM 81.37 % y KNN 55.60 %. En este conjunto de datos los resultados fueron más altos que en los dos de *eRisk*.



**Figura 4.6:** Comparación del desempeño de diferentes clasificadores en el conjunto de datos de *Twitter*. Se utilizó frecuencia normalizada como esquema de pesado de términos

En general, RL fue el clasificador con mejor desempeño en los tres conjuntos de datos con la representación TF.

#### 4.2.4. Experimento 3: Evaluando diferentes esquemas de pesado

En este tercer experimento se evaluaron tres esquemas de pesado: Binario, Frecuencia normalizada (TF) y TF-IDF. Haciendo uso de los 4 algoritmos de clasificación, el preprocesamiento de datos y aplicado a los 3 conjuntos de datos.

En la Tabla 4.1 se presentan los resultados obtenidos con el esquema de pesado Binario. Regresión Logística fue el clasificador con mejor desempeño en los tres conjuntos de datos, alcanzando 49.41 %, 57.78 % y 87.66 % de F1 en *eRisk 2017*, *eRisk 2018* y *Tweets*, respectivamente. Sin embargo, clasificadores como SVM y RF mostraron resultados bajos en algunos conjuntos, lo que indica que el esquema Binario no resulta igualmente útil para todos los algoritmos.

**Tabla 4.1:** Comparación del desempeño al aplicar el esquema de pesado binario. Se evaluó con F1 en los tres conjuntos de datos con los cuatro clasificadores

Modelo	<i>eRisk 2017</i>	<i>eRisk 2018</i>	<i>Tweets</i>
SVM	22.96	2.80	73.09
RF	14.04	20.00	85.81
RL	<b>49.41</b>	<b>57.78</b>	<b>87.66</b>
KNN	40.71	37.04	79.69

En la Tabla 4.2 se presentan los resultados obtenidos con el esquema de pesado TF. Regresión Logística fue nuevamente el clasificador con mejor desempeño en los tres conjuntos de datos, alcanzando 61.54 %, 56.18 % y 89.18 % de F1 en *eRisk 2017*, *eRisk 2018* y *Tweets*, respectivamente. En comparación con el esquema Binario, los resultados mejoraron en la mayoría de los casos, lo que sugiere que considerar la frecuencia de los términos aporta información útil para la clasificación.

**Tabla 4.2:** Comparación del desempeño al aplicar el esquema de pesado TF. Se evaluó con F1 en los tres conjuntos de datos con los cuatro clasificadores

<b>Modelo</b>	<i>eRisk 2017</i>	<i>eRisk 2018</i>	<i>Tweets</i>
SVM	56.90	51.89	81.37
RF	13.79	28.28	86.05
RL	<b>61.54</b>	<b>56.18</b>	<b>89.18</b>
KNN	53.24	41.88	55.60

En la Tabla 4.3 se presentan los resultados obtenidos con el esquema de pesado TF-IDF. A diferencia de los esquemas anteriores, no fue un solo clasificador el que sobresalió en todos los conjuntos: KNN obtuvo el mejor resultado en *eRisk 2017* con 56.49%, mientras que Regresión Logística fue superior en *eRisk 2018* (58.28%) y *Tweets* (88.81%). Los resultados fueron similares a los obtenidos con TF.

**Tabla 4.3:** Comparación del desempeño al aplicar el esquema de pesado TF-IDF. Se evaluó con F1 en los tres conjuntos de datos con los cuatro clasificadores

<b>Modelo</b>	<i>eRisk 2017</i>	<i>eRisk 2018</i>	<i>Tweets</i>
SVM	50.00	44.58	84.99
RF	20.00	27.37	86.16
RL	54.37	<b>58.28</b>	<b>88.81</b>
KNN	<b>56.49</b>	45.65	35.37

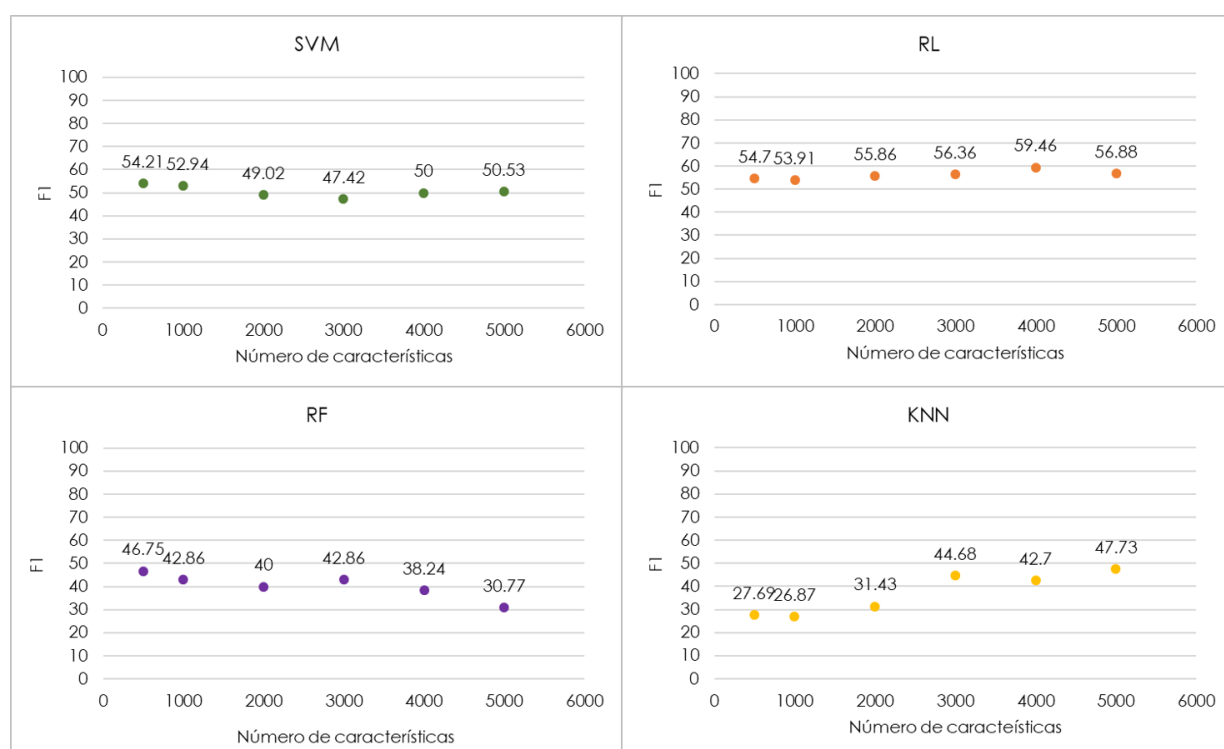
En general, los esquemas TF y TF-IDF mostraron un mejor desempeño que el esquema Binario en la mayoría de los casos, lo que indica que tomar en cuenta la frecuencia de los términos resulta más útil para la tarea de clasificación.

### 4.2.5. Experimento 4: Aplicando selección de características

En este experimento se conservó TF-IDF, y se aplica la selección de atributos mediante Chi-cuadrada para evaluar el desempeño de los modelos con diferentes cantidades de términos.

En la Figura 4.7 se muestran los resultados del conjunto de datos *eRisk 2017*, se puede observar que no todos los clasificadores responden igual al aumento de características. Regresión Logística y KNN mejoran al usar más. RL alcanza su mejor puntuación con 4 mil características (59.46%), siendo el valor más alto entre los cuatro clasificadores, mientras que KNN mejora poco a poco conforme se agregan más variables, aunque se mantiene como el clasificador con los resultados más bajos (47.73% con 5000 atributos).

Por otro lado, SVM y Bosques Aleatorios obtienen sus mejores resultados con pocas características. SVM alcanza 54.21% con 500 y se mantiene estable después, mientras que bosques aleatorios baja de 46.75% con 500 a 36.24% con 5000.

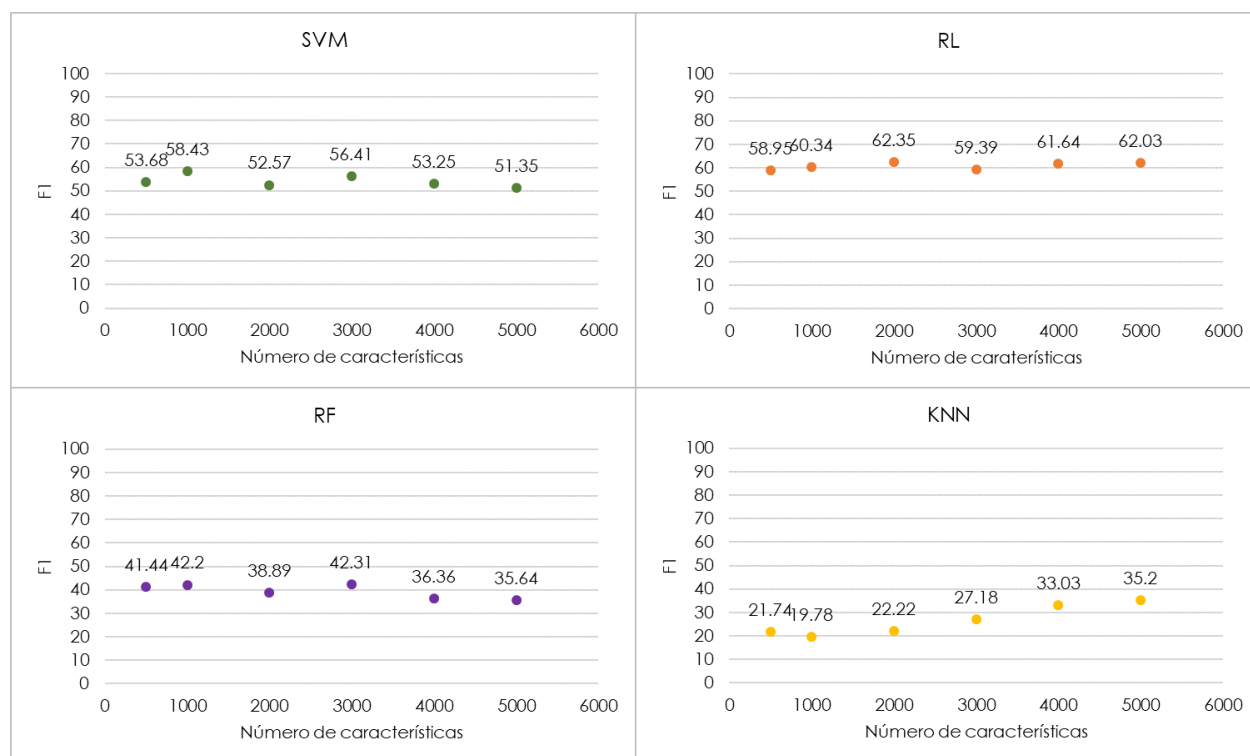


**Figura 4.7:** Selección de atributos en *eRisk 2017*. Se utilizó TF-IDF como esquema de pesado y la prueba de chi-cuadrado como método de selección de términos.

En el conjunto de *eRisk 2018* mostrado en la Figura 4.8 hay un patrón similar al de *eRisk 2017*. Regresión Logística obtiene los mejores resultados entre los cuatro clasificadores, alcanzando 62.35 % con 2 mil características y manteniéndose estable al aumentarlas. SVM alcanza su mejor resultados con mil características (58.43 %), pero no muestra mejoría con más variables.

Por otro lado, Bosques Aleatorios obtiene sus mejores resultados con pocas características, bajando de 42.2 % con mil a 35.64 % con 5 mil, siguiendo la misma tendencia que en *eRisk 2017*. KNN, en cambio, mejora conforme se agregan más características, pasando de 21.74 % con 500 a 35.2 % con 5 mil, aunque sigue siendo el clasificador con los resultados más bajos.

En general, los resultados muestran que Regresión Logística se mantiene como el mejor clasificador, y los patrones de cada algoritmo frente al aumento de características son consistentes entre ambos conjuntos de datos.

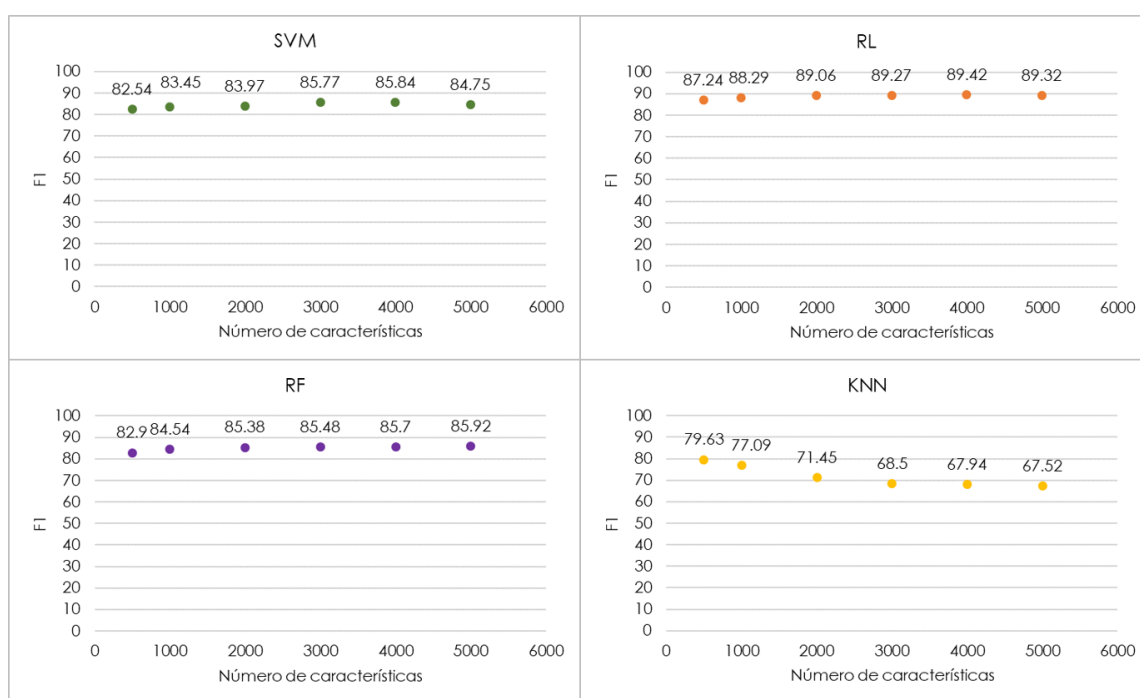


**Figura 4.8:** Selección de atributos en *eRisk 2018*. Se utilizó TF-IDF como esquema de pesado y la prueba de chi-cuadrado como método de selección de términos.

En el caso de *Twitter* que se observa en la Figura 4.9, tres de los cuatro clasificadores mejoran al aumentar el número de características: Regresión Logística alcanza el mejor resultado con 89.42% usando 4 mil, SVM sube de 82.54% con 500 a 85.84% con 4 mil, y Bosques Aleatorios pasa de 82.9% con 500 a 85.92% con 5 mil. En los tres casos la mejora es gradual y los resultados se mantienen estables a partir de las 2 mil características.

KNN es la excepción, ya que su desempeño baja conforme se agregan más características, pasando de 79.63% con 500 a 67.52% con 5000 atributos, lo que indica que para este clasificador usar menos características funciona mejor en este conjunto de datos.

En resumen, SVM, Regresión Logística y Bosques Aleatorios mejoran al usar más características en *Twitter*, mientras que KNN es el único que baja.

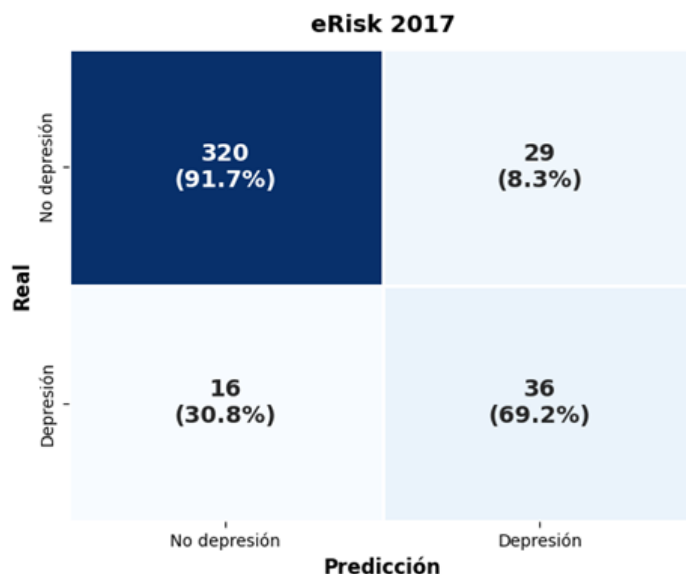


**Figura 4.9:** Selección de atributos en *Twitter*. Se utilizó TF-IDF como esquema de pesado y la prueba de chi-cuadrado como método de selección de términos.

#### 4.2.6. Evaluación de los modelos: matrices de confusión bajo las mejores configuraciones

Después de comparar los diferentes esquemas de pesado, clasificadores y métodos de selección de características, se seleccionó la mejor configuración para cada conjunto de datos. A continuación se presentan las matrices de confusión correspondientes, junto con las métricas de precisión, exhaustividad y F1 de la clase positiva (depresión).

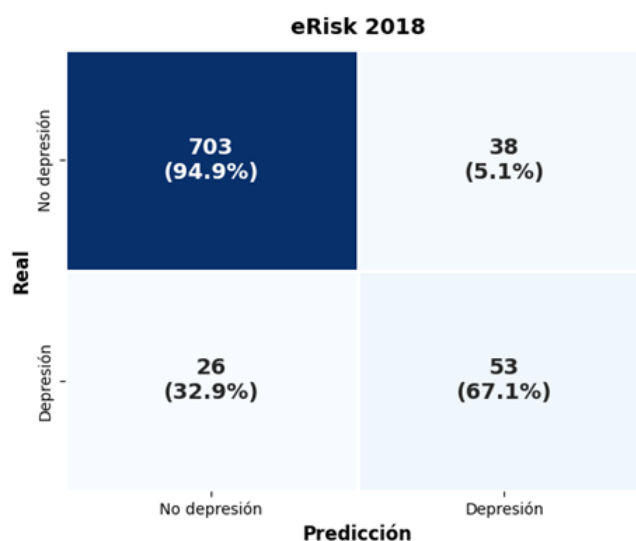
Para el conjunto de datos *eRisk 2017*, el mejor resultado se obtuvo con Regresión Logística y el esquema de pesado TF, alcanzando un F1 de 61.54%, con una precisión de 55.38% y una exhaustividad de 69.23%. En la Figura 4.10 se muestra la matriz de confusión, donde se observa que de los 52 casos de depresión, el modelo logró identificar correctamente 36.



**Figura 4.10:** Matriz de confusión del mejor modelo en *eRisk 2017*. El mejor resultado se obtuvo con frecuencia normalizada como esquema de pesado, sin selección de características y utilizando Regresión Logística

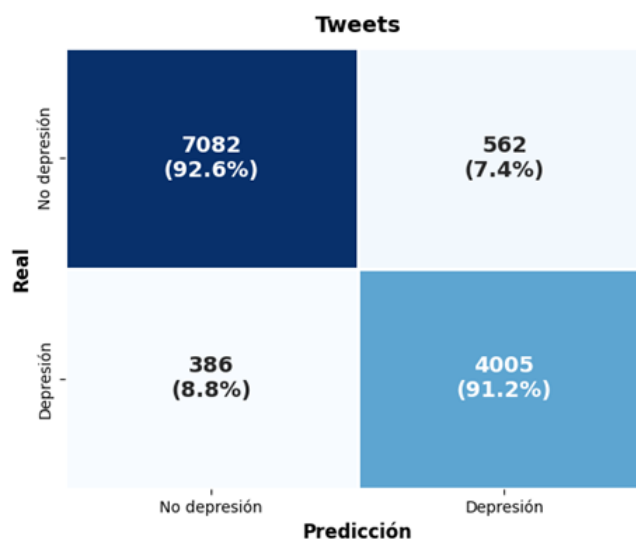
Para el conjunto de datos *eRisk 2018*, el mejor resultado se obtuvo con Regresión Logística y el esquema de pesado TF-IDF con selección de características a 2,000 términos, alcanzando un F1 de 62.35%, con una precisión de 58.24% y una exhaustividad de 67.09%. En la

Figura 4.11 se muestra la matriz de confusión, donde se observa que de los 79 casos de depresión, el modelo logró identificar correctamente 53.



**Figura 4.11:** Matriz de confusión del mejor modelo en *eRisk 2018*. El mejor resultado se obtuvo con TF-IDF como esquema de pesado, selección de características a 2,000 términos y utilizando Regresión Logística

Para el conjunto de datos de *Tweets*, el mejor resultado se obtuvo con Regresión Logística y el esquema de pesado TF-IDF con selección de características a 4,000 términos, alcanzando un F1 de 89.42%, con una precisión de 87.69% y una exhaustividad de 91.21%. En la Figura 4.12 se muestra la matriz de confusión, donde se observa que de los 4,391 casos de depresión, el modelo logró identificar correctamente 4,005.



**Figura 4.12:** Matriz de confusión del mejor modelo en *Tweets*. El mejor resultado se obtuvo con TF-IDF como esquema de pesado, selección de características a 4,000 términos y utilizando Regresión Logística

#### 4.2.7. Análisis comparativo entre plataformas sociales

Como análisis adicional, se incluyó una nube de palabras con los términos más discriminativos, la cual se muestra en la Figura 4.13. En la figura, entre más grande es la fuente de las palabras, mayor es su relevancia para distinguir las clases.

El primer conjunto a analizar fue *eRisk 2017*, donde se observan palabras como *depression*, *anxiety*, *diagnosed*, *medication* y *psychiatrist*, que podrían estar relacionadas con la salud mental. También aparecen términos del ámbito emocional y social como *emotionally*, *feeling*, *relationship*, así como palabras asociadas al tratamiento como *therapy*, *therapist*, *prescribed* y *med*.

Esto sugiere que el vocabulario más útil para distinguir entre usuarios con y sin depresión son aquellas relacionadas con el padecimiento, las emociones y el tratamiento en el caso de *eRisk 2017*.





**Figura 4.14:** Nube de palabras más discriminativas en *eRisk 2018*. Se obtuvieron los términos con mayor valor de Chi-cuadrada

Al igual que en los conjuntos anteriores, se incluyó la nube de palabras del corpus de *Twitter*, mostrada en la Figura 4.15. Las palabras más grandes en la nube están relacionadas con la depresión, como *depression*, *depressed*, *disorder* e *illness*.

A diferencia de las nubes de *eRisk*, aquí aparecen palabras más directas y fuertes como *death*, *dead*, *kill*, *suicide* y *commit*. Esto puede estar relacionado con que los *tweets* son textos cortos, por lo que los usuarios tienden a expresar lo que sienten con menos palabras.

También aparecen palabras positivas como *happy*, *cheerful*, *optimistic*, *hopeful* y *beautiful*, lo que muestra que no solo las palabras negativas ayudan a distinguir entre las clases.



*happy*, *optimistic* y *cheerful*, así como palabras más directas sobre muerte y autolesión como *suicide*, *kill*, *dead*, mientras que en *eRisk* las palabras tienen que ver más con el tratamiento y las relaciones personales como *boyfriend* y *relationship*.

---

# Capítulo 5

## Conclusiones y trabajo futuro

### 5.1. Conclusiones

En este trabajo de tesis se evaluaron diferentes métodos de clasificación de textos para la detección de casos de depresión en publicaciones de redes sociales. Se experimentó con cuatro algoritmos de clasificación, tres esquemas de pesado de términos y selección de características. La idea principal fue comparar la tarea de detección en diferentes plataformas de redes sociales. Para este propósito, los modelos de detección de depresión fueron evaluados en tres conjuntos de datos: *eRisk 2017*, *eRisk 2018* y una colección de datos obtenida de Twitter.

Los resultados muestran que Regresión Logística fue el clasificador más consistente, obteniendo el mejor desempeño en los tres conjuntos de datos. En cuanto a los esquemas de pesado, el esquema Binario presentó el menor rendimiento, mientras que TF y TF-IDF obtuvieron resultados similares entre sí y superiores al Binario, lo que indica que considerar la frecuencia de los términos aporta información útil para la clasificación.

La selección de características mediante Chi-cuadrada permitió reducir la cantidad de términos sin afectar negativamente el rendimiento, e incluso mejoró los resultados en *eRisk 2018* y *Tweets*, donde las mejores configuraciones utilizaron 2,000 y 4,000 términos, respectivamente. En contraste, para *eRisk 2017* el mejor resultado se obtuvo con TF sin selección

de características.

Además, las nubes de palabras generadas a partir de los términos con mayor valor de Chi-cuadrada permitieron visualizar las palabras más discriminativas para cada conjunto de datos, mostrando diferencias en el vocabulario asociado a la depresión según la plataforma de origen.

El conjunto de *Tweets* fue el que presentó los mejores resultados en general, alcanzando un F1 de 89.42 %, mientras que los conjuntos de *eRisk* mostraron un desempeño más bajo, posiblemente debido al mayor desbalance de clases presente en estos conjuntos de datos.

## 5.2. Limitaciones

Aunque la presente investigación obtuvo resultados de interés, es necesario reconocer algunas limitaciones.

1. Los tres conjuntos de datos empleados en la experimentación se encuentran escritos en inglés, por lo que es necesario evaluar el comportamiento de los modelos usando otras colecciones de datos con textos en otros idiomas para generalizar los resultados.
2. Los datos provienen de *Reddit* y *Twitter*, dos plataformas donde los usuarios han mostrado estilos de escritura muy diferentes. Estas diferencias permiten estudiar la relación del tipo de plataforma con patrones lingüísticos asociados a depresión. Sin embargo, en este trabajo no se proponen modelos generales para detección de depresión en diferentes plataformas, en tal caso, se necesitaría entrenar un modelo con textos de entrenamiento de las diferentes plataformas.
3. La técnicas de bolsa de palabras y los clasificadores tradicionales tienen una capacidad limitada para capturar el contexto del lenguaje. Existen otros modelos más recientes y robustos que se explorarán como parte del trabajo futuro para integrar el contexto como los transformadores.

4. Un aspecto importante en esta temática es la interpretación por los especialistas. De ahí que, los términos identificados como discriminativos necesitan ser validados por personal especializado en salud mental. Además, es importante señalar que este tipo de métodos constituyen una herramienta para los especialistas más nunca se propone la sustitución de ellos.
5. Los modelos se evaluaron con datos limpios y etiquetados, pero en la práctica los textos suelen ser más desordenados, conteniendo textos en diferentes idiomas con fragmentos que exponen ironía o sarcasmo, por lo tanto, se requiere más trabajo y mejor desempeño de los modelos para lograr su aplicabilidad en entornos reales.

### 5.3. Trabajo futuro

Con el fin de ampliar esta investigación, se proponen los siguientes puntos.

1. Enriquecer el análisis estudiando la detección de depresión con datos de distintas redes sociales.
2. Explorar modelos de aprendizaje profundo para mejorar la detección de indicadores depresivos en diferentes textos. Por ejemplo, se tiene el interés de evaluar el desempeño de arquitecturas basados en *transformers*.
3. Incluir otras modalidades de información además del texto para enriquecer el análisis, como imágenes o patrones de actividad en redes sociales (por ejemplo, horario de publicaciones e interacción con otros usuarios).
4. Ampliar y construir nuevos conjuntos de datos en español. Actualmente, la mayoría de los conjuntos de datos de entrenamiento disponibles contienen textos en Inglés. De ahí que surge el interés de crear recursos que permitan desarrollar modelos predictivos que procesen textos en español.
5. Diseñar versiones de prueba de aplicaciones que permitan integrar modelos predicti-

vos como apoyo al diagnóstico, con protección de datos y supervisión de especialistas en salud.

6. Adaptar los modelos generados para identificar otras condiciones de salud como ansiedad o ideación suicida.

## Referencias

- [1] World Health Organization. Depressive disorder (depression). Disponible en: <https://www.who.int/news-room/fact-sheets/detail/depression>, Agosto 2025. Consultado el 05 de septiembre de 2025.
- [2] National Institute of Mental Health. Depresión. Disponible en: <https://www.nimh.nih.gov/health/publications/espanol/depression-sp>, 2024. Consultado el 07 de septiembre de 2025.
- [3] Secretaría de Salud. Descifremos. *Género y Salud en Cifras*, 15(2):38–40, 2017.
- [4] Jorge Héctor Alvarado Barranco. Factores asociados a la depresión postpandemia por el virus del SARS-COV-2 Covid-19 en estudiantes de Educación Superior. *Ciencia Latina Revista Científica Multidisciplinar*, 7(3):1553–1565, 2023.
- [5] Sandra Constanza Cañón Buitrago and Jaime Alberto Carmona Parra. Ideación y conductas suicidas en adolescentes y jóvenes. *Pediatría Atención Primaria*, 20(80):387–397, 2018.
- [6] Megan A. Moreno, Lauren A. Jelenchick, Kristin G. Egan, Elissa Cox, Henry Young, Kathryn E. Gannon, and Tara Becker. Feeling bad on Facebook: Depression disclosures by college students on a social networking site. *Depression and Anxiety*, 28(6):447–455, 2011.
- [7] UNICEF España. 4 de cada 10 adolescentes manifiestan haber tenido o creer haber tenido un problema de salud mental. Disponible en: <https://www.unicef.es/noticia/>

- 
- [adolescentes-manifiestan-problema-salud-mental](#), Octubre 2024. Consultado el 14 de septiembre del 2025.
- [8] Sho Tsugawa, Yusuke Kikuchi, Fumio Kishino, Kosuke Nakajima, Yuichi Itoh, and Hiroyuki Ohsaki. Recognizing Depression from Twitter Activity. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI '15*, page 3187–3196, New York, NY, USA, 2015. Association for Computing Machinery.
- [9] James W. Pennebaker, Matthias R. Mehl, and Kate G. Niederhoffer. Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology*, 54:547–577, 2003.
- [10] Molly Ireland, Kaitlin Adams, and Sean Farrell. Tracking Mental Health Risks and Coping Strategies in Healthcare Workers’ Online Conversations Across the COVID-19 Pandemic. In Ayah Zirikly, Dana Atzil-Slonim, Maria Liakata, Steven Bedrick, Bart Desmet, Molly Ireland, Andrew Lee, Sean MacAvaney, Matthew Purver, Rebecca Resnik, and Andrew Yates, editors, *Proceedings of the Eighth Workshop on Computational Linguistics and Clinical Psychology*, pages 76–88, Seattle, USA, July 2022. Association for Computational Linguistics.
- [11] Adalberto Campo-Arias, Heidi Celina Oviedo, and Edwin Herazo. Estigma: barrera de acceso a servicios en salud mental. *Revista Colombiana de Psiquiatría*, 43:162 – 167, 07 2014.
- [12] O. Taycan, A. Özdemir, and S. Erdoğan Taycan. Alexithymia and Somatization in Depressed Patients: The Role of the Type of Somatic Symptom Attribution. *Noro Psikiyatir Arsivi*, 54(2):99–104, jun 2017.
- [13] Robert Kohn, Ali Ahsan Ali, Victor Puac-Polanco, Chantal Figueroa, Victor López-Soto, Kristen Morgan, Sandra Saldivia, and Benjamín Vicente. Mental health in the Americas: an overview of the treatment gap. *Revista Panamericana de Salud Pública*, 42:e165, oct 2018.

- 
- [14] Xiaoli Ye and Haohao Gao. Distress disclosure on social media and depressive symptoms among college students: the roles of social comparison and gender. *Frontiers in Psychology*, 16:1520066, 2025.
- [15] Michael M. Tadesse, Hongfei Lin, Bo Xu, and Liang Yang. Detection of Depression-Related Posts in Reddit Social Media Forum. *IEEE Access*, 7:44883–44893, 2019.
- [16] Keith Harrigian, Carlos Aguirre, and Mark Dredze. Do Models of Mental Health Based on Social Media Data Generalize? In Trevor Cohn, Yulan He, and Yang Liu, editors, *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 3774–3788, Online, November 2020. Association for Computational Linguistics.
- [17] Wusat Ullah, Patrícia Oliveira-Silva, Muhammad Nawaz, Rana Muhammad Zulqarnain, Imran Siddique, and Mohammed Sallah. Identification of depressing tweets using natural language processing and machine learning: Application of grey relational grades. *Journal of Radiation Research and Applied Sciences*, 18(1):101299, 2025.
- [18] Uma Yadav, Ashish K Sharma, and Dipti Patil. Review of automated depression detection: Social posts, audio and video, open challenges and future direction. *Concurrency and Computation: Practice and Experience*, 35(1):e7407, 2023.
- [19] Rafael A. Calvo, David N. Milne, M. Sazzad Hussain, and Helen Christensen. Natural language processing in mental health applications using non-clinical texts. *Natural Language Engineering*, 23(5):649–685, 2017.
- [20] Stevie Chancellor and Munmun De Choudhury. Methods in predictive techniques for mental health status on social media: a critical review. *NPJ Digital Medicine*, 3(1):43, 2020.
- [21] Glen Coppersmith, Mark Dredze, and Craig Harman. Quantifying Mental Health Signals in Twitter. In *Proceedings of the workshop on computational linguistics and clinical psychology: From linguistic signal to clinical reality*, pages 51–60, 2014.

- 
- [22] Instituto Nacional de Estadística y Geografía (INEGI). Encuesta Nacional de Bienestar Autorreportado (ENBIARE) 2021. Comunicado de prensa Núm. 772/21. Disponible en: [https://www.inegi.org.mx/contenidos/saladeprensa/boletines/2021/EstSociodemo/ENBIARE\\_2021.pdf](https://www.inegi.org.mx/contenidos/saladeprensa/boletines/2021/EstSociodemo/ENBIARE_2021.pdf), Diciembre 2021. Consultado el 29 de septiembre de 2025.
- [23] Sharath Chandra Guntuku, David B Yaden, Margaret L Kern, Lyle H Ungar, and Johannes C Eichstaedt. Detecting depression and mental illness on social media: an integrative review. *Current Opinion in Behavioral Sciences*, 18:43–49, 2017.
- [24] Chip Huyen. *Designing Machine Learning Systems: An Iterative Process for Production-Ready Applications*. O’Reilly Media, Sebastopol, CA, 2022.
- [25] Michael W. Berry, Azlinah Mohamed, and Bee Wah Yap. *Supervised and Unsupervised Learning for Data Science*. Springer, 2019.
- [26] Sebastian Raschka, Yuxi (Hayden) Liu, and Vahid Mirjalili. *Machine Learning with PyTorch and Scikit-Learn*. Packt Publishing, Birmingham, UK, 2022.
- [27] Aurélien Géron. *Hands-On Machine Learning with Scikit-Learn, Keras, & TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O’Reilly Media, Sebastopol, CA, 2nd edition, 2019.
- [28] Francisco Pereira, Tom Mitchell, and Matthew Botvinick. Machine learning classifiers and fMRI: a tutorial overview. *NeuroImage*, 45(1, Supplement 1):S199–S209, 2009.
- [29] Hasan Ahmed Salman, Ali Kalakech, and Amani Steiti. Random Forest Algorithm Overview. *Babylonian Journal of Machine Learning*, 2024:69–79, 2024.
- [30] Eda Kavlakoglu. ¿Qué es el bosque aleatorio? IBM. Disponible en: <https://www.ibm.com/mx-es/think/topics/random-forest>. Consultado el 05 de enero de 2026.
- [31] Gustavo A Betancourt. Las máquinas de soporte vectorial (SVMs). *Scientia et Technica*, 1(27), 2005.

- 
- [32] Ashis Pradhan. SUPPORT VECTOR MACHINE-A Survey. *International Journal of Emerging Technology and Advanced Engineering*, 2(8):82–85, 2012.
- [33] Lee, Fangfang. What is logistic regression? IBM. Disponible en: <https://www.ibm.com/think/topics/logistic-regression>. Consultado el 15 de enero de 2026.
- [34] Avinash Navlani. Comprender la regresión logística en el tutorial de Python. DataCamp. Disponible en: <https://www.datacamp.com/es/tutorial/understanding-logistic-regression-python>, Abril 2024. Consultado el 02 de febrero de 2026.
- [35] Gonzalo Berástegui Arbeloa and Mikel Galar Idoate. Implementación del algoritmo de los k vecinos más cercanos (k-NN) y estimación del mejor valor local de k para su cálculo, 2018.
- [36] María Luisa Micó Andrés. *Algoritmos de búsqueda de vecinos más próximos en espacios métricos*. PhD thesis, Universidad Politécnica de Valencia, Valencia, España, March 1996.
- [37] Jordi Gironés Roig, Jordi Casas Roma, Julià Minguillón Alfonso, and Ramon Caihuellas Quiles. *Minería de datos: modelos y algoritmos*. Editorial UOC, Barcelona, 2017.
- [38] Alice Zheng. *Evaluating Machine Learning Models: A Beginner's Guide to Key Concepts and Pitfalls*. O'Reilly Media, Sebastopol, CA, 2015.
- [39] David M. W. Powers. Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation. *Journal of Machine Learning Technologies*, 2(1):37–63, 2011.
- [40] Daniel Jurafsky and James H. Martin. *Speech and Language Processing*. Stanford University, 2023. Draft of 3rd edition.
- [41] Khaled Mahmud Sujon, Rohayanti Hassan, Kwonhue Choi, and Md Abdus Samad. Accuracy, precision, recall, f1-score, or MCC empirical evidence from advanced sta-

- 
- tistics, ML, and XAI for evaluating business predictive models. *Journal of Big Data*, 12(1):268, 2025.
- [42] Juri Opitz and Sebastian Burst. Macro F1 and Macro F1. *arXiv preprint arXiv:1911.03347*, 2019.
- [43] Andreas C. Müller and Sarah Guido. *Introduction to Machine Learning with Python: A Guide for Data Scientists*. O’Reilly Media, Sebastopol, CA, 2016.
- [44] Yanela Rodríguez, Yumilka Fernández, Rafael Bello, and Yailé Caballero. Selección de atributos relevantes aplicando algoritmos que combinan conjuntos aproximados y optimización en colonias de hormigas. *Revista Cubana de Ciencias Informáticas*, 8(1):79–86, 2014.
- [45] David Lastre, Melanie Páez Santana, and Olga López Tumbaco. Análisis estadístico de tablas de contingencia y chi cuadrado para medir el flujo migratorio en el Ecuador en el 2018. *Ecuadorian Science Journal*, 3(1):23–30, 2019.
- [46] K. R. Chowdhary. *Natural Language Processing*, pages 603–649. Springer India, New Delhi, 2020.
- [47] Dorys Moreira, Ivan Cruz, Karolina Gonzalez, Andrea Quirumbay, Christian Magallan, Teresa Guarda, Alicia Andrade, and Carlos Castillo. Análisis del Estado Actual de Procesamiento de Lenguaje Natural. *Revista Ibérica de Sistemas e Tecnologías de Informação*, pages 126–136, 2021.
- [48] Denis Cedeno-Moreno and Miguel Vargas. Aprendizaje automático aplicado al análisis de sentimientos. *I+ D Tecnológico*, 16(2):59–66, 2020.
- [49] Mari Vallez and Rafael Pedraza. El Procesamiento del Lenguaje Natural en la Recuperación de Información Textual y Áreas Afines. *Hipertext. net*, 5, 2007.
- [50] Mingyu Wan, Alex Chengyu Fang, and Chu-Ren Huang. The Discriminateness of Internal Syntactic Representations in Automatic Genre Classification. *Journal of*

- 
- Quantitative Linguistics*, 28(2):138–171, 2021.
- [51] Wisam A. Qader, Musa M. Ameen, and Bilal I. Ahmed. An Overview of Bag of Words; Importance, Implementation, Applications, and Challenges. In *2019 International Engineering Conference (IEC)*, pages 200–204, 2019.
- [52] Vamshi Prakash. An Introduction to Bag of Words (BoW). Medium. Disponible en: <https://medium.com/@vamshiprakash001/an-introduction-to-bag-of-words-bow-c32a65293ccc>. Consultado el 04 de febrero de 2026.
- [53] Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. *Introduction to Information Retrieval*. Cambridge University Press, 2008.
- [54] Kepios (DataReportal). Global Social Media Statistics. DataReportal. Disponible en: <https://datareportal.com/social-media-users>, Enero 2025. Consultado el 04 de febrero de 2026.
- [55] Munmun De Choudhury and Scott Counts. The Nature of Emotional Expression in Social Media: Measurement, Inference and Utility. *Human Computer Interaction Consortium (HCIC)*, 2012.
- [56] Ema Kušen, Giuseppe Cascavilla, Kathrin Figl, Mauro Conti, and Mark Strembeck. Identifying Emotions in Social Media: Comparison of Word-Emotion Lexicons. In *2017 5th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW)*, pages 132–137. IEEE, 2017.
- [57] Shelley Boulianne, Christian P. Hoffmann, and Michael Bossetta. Social media platforms for politics: A comparison of Facebook, Instagram, Twitter, YouTube, Reddit, Snapchat, and WhatsApp. *New Media & Society*, 27(11):6006–6037, 2025.
- [58] Krishna C Bathina, Marijn ten Thij, Lorenzo Lorenzo-Luaces, Lauren A Rutter, and Johan Bollen. Depressed individuals express more distorted thinking on social media. *arXiv preprint arXiv:2002.02800*, 2020.

- 
- [59] Juana Cristina Troya Fernández, Noelia Perriñán Sotelo, and Paula Sánchez-Movellán-Pérez. El Impacto de las Redes Sociales en la Salud Mental. Revisión Bibliográfica. *Sanum: Revista Científico-Sanitaria*, 7(1):18–28, 2022.
- [60] Abderrahman M. Khalaf, Abdullah A. Alubied, Ahmed M. Khalaf, and Abdallah A. Rifaey. The Impact of Social Media on the Mental Health of Adolescents and Young Adults: A Systematic Review. *Cureus*, 15(8), 2023.
- [61] Layan Azem, Rafea Al Alwani, Augusto Lucas, Balqes Alsaadi, Gilbert Njihia, Bushra Bibi, Mahmood Alzubaidi, and Mowafa Househ. Social Media Use and Depression in Adolescents: A Scoping Review. *Behavioral Sciences*, 13(6), 2023.
- [62] Loes HC Janssen, Patti M Valkenburg, Loes Keijsers, and Ine Beyens. A harsher reality for adolescents with depression on social media. *Scientific Reports*, 15(1):10947, 2025.
- [63] Ariana Maybee Orué Medina. Detección de depresión a través de análisis textual utilizando aprendizaje automático. Tesis de licenciatura, Universidad Ricardo Palma, 2018.
- [64] Derick Axel Lagunes-Ramírez, Gabriel González-Serna, Leonor Rivera-Rivera, Nimrod González-Franco, Dante Mújica-Vargas, and María Y Hernández-Pérez. Comportamiento de la mirada y análisis mediante aprendizaje automático de la depresión en la juventud: una revisión sistemática. *XIKUA Boletín Científico de la Escuela Superior de Tlahuelilpan*, 12(23):56–68, 2024.
- [65] Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. Predicting Depression via Social Media. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 7, pages 128–137, 2013.
- [66] Yunpeng Zhao, Yi Guo, Xing He, Yonghui Wu, Xi Yang, Mattia Prospero, Yanghua Jin, and Jiang Bian. Assessing mental health signals among sexual and gender minorities using Twitter data. *Health informatics journal*, 26(2):765–786, 2020.

- 
- [67] Zhentao Xu, Verónica Pérez-Rosas, and Rada Mihalcea. Inferring Social Media Users' Mental Health Status from Multimodal Information. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 6292–6299, 2020.
- [68] Annika M. Schoene, George Lacey, Alexander P. Turner, and Nina Dethlefs. Dilated LSTM with Attention for Classification of Suicide Notes. In *Proceedings of the Tenth International Workshop on Health Text Mining and Information Analysis (LOUHI 2019)*, pages 136–145, 2019.
- [69] Vankayala Tejaswini, Korra Sathya Babu, and Bibhudatta Sahoo. Depression Detection from Social Media Text Analysis using Natural Language Processing Techniques and Hybrid Deep Learning Model. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, 23(1), January 2024.
- [70] Harnain Kour and Manoj K. Gupta. An hybrid deep learning approach for depression prediction from user tweets using feature-rich CNN and bi-directional LSTM. *Multimedia Tools and Applications*, 81:23649–23685, 2022.
- [71] Prabhakar K and Kavitha V. An evolutionary approach for depression detection from Twitter big data using a novel deep learning model with attention based feature learning mechanism. *Automatika*, 65(2):441–453, 2024.
- [72] Andrew Yates, Arman Cohan, and Nazli Goharian. Depression and Self-Harm Risk Assessment in Online Forums. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2958–2968. Association for Computational Linguistics, 2017.
- [73] Elsbeth Turcan and Kathy McKeown. Dreddit: A Reddit Dataset for Stress Analysis in Social Media. In Eben Holderness, Antonio Jimeno Yepes, Alberto Lavelli, Anne-Lyse Minard, James Pustejovsky, and Fabio Rinaldi, editors, *Proceedings of the Tenth International Workshop on Health Text Mining and Information Analysis (LOUHI 2019)*, pages 97–107, Hong Kong, November 2019. Association for Computational

Linguistics.

- [74] David E. Losada, Fabio Crestani, and Javier Parapar. eRISK 2017: CLEF Lab on Early Risk Prediction on the Internet: Experimental Foundations. In *Experimental IR Meets Multilinguality, Multimodality, and Interaction: 8th International Conference of the CLEF Association, CLEF 2017, Dublin, Ireland, September 11–14, 2017*, pages 346–360. Springer, 2017.
- [75] David E. Losada, Fabio Crestani, and Javier Parapar. Overview of eRisk 2018: Early Risk Prediction on the Internet (Extended Lab Overview). In *Proceedings of the 9th International Conference of the CLEF Association, CLEF 2018*, pages 1–20, 2018.
- [76] Radwa Nassar, AbdelMoniem Helmy, and Nagy Ramadan. Binary Labeled Depression Corpus of 60,000 English Tweets, 2022.