



Universidad Autónoma del Estado de México

Unidad Académica Profesional Tianguistenco

Identificación automática de ideación suicida en textos
empleando clasificadores en cascada

T E S I S

Que para obtener el grado de
Maestra en Ciencias de la Computación

Presenta

María del Carmen García Galindo

Tutor académico

Dr. Ángel Hernández Castañeda

Tutores adjuntos

Dr. René Arnulfo García Hernández

M. en C. C. José Rafael Cruz Reyes

Tianguistenco, Estado de México

septiembre 2023

Resumen

A nivel mundial, miles de personas deciden quitarse la vida cada año. Por tal motivo, el suicidio se considera un grave problema social y sanitario que afecta a toda la población. Este fenómeno se relaciona con las emociones debido a que la mayoría de las personas que expresan sus intenciones suicidas describen su motivación como resultado de diversos estados emocionales. En este sentido, algunos autores afirman que las personas que experimentan mayor presencia de emociones negativas tienen más probabilidades de suicidarse.

A lo largo de los años, en diferentes campos de investigación se han desarrollado estrategias para prevenir y reducir las muertes por suicidio. En el área computacional, se han desarrollado modelos que emplean técnicas de aprendizaje automático y procesamiento del lenguaje natural para aprender patrones textuales e identificar en periodos cortos de tiempo millones de publicaciones suicidas en redes sociales.

Estas investigaciones han producido resultados prometedores, sin embargo, la mayoría de ellas aplican una clasificación directa, la cual determina si una nota expresa ideación suicida en función de las palabras utilizadas asociadas directamente al suicidio. Como consecuencia, las técnicas basadas en este tipo de clasificación pueden no ser efectivas para identificar notas que expresan ideas suicidas de forma implícita porque en estas notas generalmente no se emplean términos relacionadas con el suicidio.

Con el objetivo de detectar notas implícitas de ideación suicida, en este estudio se propone el uso de dos clasificadores de aprendizaje automático configurados en una arquitectura en cascada. Esta propuesta se caracteriza porque mediante las palabras utilizadas el primer clasificador obtiene la distribución de emociones latentes en cada texto. Posteriormente, el segundo toma la distribución emocional para determinar si se trata de una nota suicida. Para lograr tal objetivo se emplean diferentes métodos de generación de características con el fin de identificar qué tipo de información (léxica o semántica) es más útil para detectar este tipo de notas. Los clasificadores implementados en este estudio son máquinas de vectores de soporte, Naïve Bayes multinomial y una red neuronal de perceptrón multicapa. Estos clasificadores se seleccionaron debido a que han sido usados en la tarea en cuestión, alcanzado resultados prometedores.

Abstract

Every year, thousands of people around the world take their own lives. Therefore, suicide is considered a serious social and health problem that affects the entire population. This phenomenon is related to emotions because most people who express suicidal intentions describe their motivation as a result of various emotional states. In this sense, some authors affirm that people who experience more negative emotions are more likely to commit suicide.

Over the years, different fields of research have developed strategies to prevent and reduce suicide. In the computational area, models have been developed that use machine learning and natural language processing techniques to learn textual patterns and identify millions of suicidal posts on social networks in short periods of time.

These investigations have produced promising results, however, most of them apply a direct classification that determines if a note expresses some suicidal ideation based on the words used that are directly associated with suicide. Consequently, the techniques based on this type of classification may not be effective for identifying notes that express suicidal ideas implicitly, because in these notes, no terms related to suicide are generally used.

With the aim of detecting implicit notes of suicidal ideation, in this study we propose the use of two machine learning classifiers configured in a cascade architecture. This proposal is characterized by the fact that the first classifier obtains the distribution of latent emotions in each text. Subsequently, the second takes the emotional distribution to determine whether it is a suicidal note. In addition, we use different feature generation methods to identify the type of information (lexical or semantic) that is more useful for detecting this type of notes. The classifiers implemented in this study are support vector machines, multinomial Naïve Bayes and a multilayer perceptron neural network. These classifiers were selected because they have been used in the task at hand and have achieved promising results.

Índice

Capítulo 1 Introducción	12
1.1 Planteamiento del problema	19
1.2 Justificación	20
1.3 Delimitación	21
1.4 Hipótesis.....	21
1.5 Objetivo general	21
1.6 Objetivos específicos	21
1.7 Organización de la tesis	21
Capítulo 2 Marco teórico.....	23
2.1 Análisis de textos.....	23
2.2 Clasificación de textos	24
2.2.1 Minería de opiniones	26
2.2.2 Clasificación directa	27
2.2.3 Clasificación en cascada.....	28
2.2.4 Preprocesamiento	30
2.2.5 Extracción de características.....	31
2.2.5.1 Codificación de un solo paso.....	32
2.2.5.2 Frecuencia de término – frecuencia de documento inversa	33
2.2.5.3 Asignación latente de Dirichlet	34
2.2.5.4 Vector de párrafo	36
2.2.6 Algoritmos de clasificación supervisada	37
2.2.6.1 Máquina de vectores de soporte	38
2.2.6.2 Naïve Bayes multinomial.....	40
2.2.6.3 Perceptrón multicapa	41
2.2.7 Método de validación cruzada.....	43
2.2.8 Medidas de evaluación.....	45
2.2.8.1 Precisión	45
2.2.8.2 Exhaustividad.....	46
2.2.8.3 Medida F	46

Capítulo 3 Estado del arte	47
3.1 Introducción.....	47
3.2 Análisis de depresión y suicidio mediante aprendizaje automático y PLN	50
3.3 Caracterización de la ideación suicida mediante el uso de características de trastornos mentales en microblogs: Una perspectiva de aprendizaje automático	51
3.4 Detección de ideación suicida en foros de redes sociales mediante aprendizaje profundo.....	55
Capítulo 4 Método propuesto.....	60
4.1 Conjuntos de datos.....	61
4.1.1 Notas de suicidio consumado.....	62
4.1.2 Notas de intenciones suicidas	62
4.1.3 Notas no relacionadas al suicidio	63
4.1.4 Notas de sentenciados	63
4.1.5 Conjunto de emociones	64
4.2 Estadísticas básicas de los conjuntos de datos	65
4.3 Generación de características.....	66
4.3.1 Codificación de un solo paso	66
4.3.2 Frecuencia de término-frecuencia de documento inversa	67
4.3.3 Asignación latente de Dirichlet.....	68
4.4 Clasificación en cascada	69
4.4.1 Clasificación de emociones.....	70
4.4.2 Clasificación de notas.....	71
Capítulo 5 Experimentación y resultados.....	72
5.1 Preprocesamiento de los datos.....	72
5.2 Generación de los vectores de características	73
5.3 Clasificación	74
5.3.1 Intención suicida y ejecución: dos caras de la muerte	74
5.3.2 Suicidio consumado: la necesidad de analizar más allá de las palabras	78
5.3.3 Emociones finales: notas de suicidio consumado y personas sentenciadas.....	81
Capítulo 6 Conclusiones	84
6.1 Aportaciones	85
Referencias.....	87

Índice de figuras

Figura 1. Metodología de la clasificación directa	27
Figura 2. Arquitecturas de los sistemas multclasificadores	29
Figura 3. Esquema de representaciones de palabras	36
Figura 4. Máquina de vectores de soporte	38
Figura 5. Modelo básico de una red neuronal artificial	41
Figura 6. Estructura de una red neuronal	42
Figura 7. Clases de redes neuronales	43
Figura 8. Proceso de validación cruzada	45
Figura 9. Redes sociales estudiadas para la detección de ideación suicida	49
Figura 10. Idiomas estudiados para la detección de ideación suicida	49
Figura 11. Metodología de Jain et al. (2022)	50
Figura 12. Temas latentes en tuits	53
Figura 13. Emociones extraídas	54
Figura 14. Direcciones para la detección de ideación suicida	56
Figura 15. Enfoque combinado de aprendizaje profundo	57
Figura 16. Esquema general del método propuesto	60
Figura 17. Ejemplos de construcción de vectores con MTD	67
Figura 18. Distribución de tópicos en una nota de ideación suicida.....	69
Figura 19. Vector de probabilidad emocional.....	70
Figura 20. Probabilidad emocional en una nota de ideación suicida.....	70
Figura 21. Emociones en notas de ideación suicida	75
Figura 22. Emociones en notas de personas sentenciadas	75
Figura 23. Emociones en notas de suicidio consumado	78
Figura 24. Emociones en notas no asociadas con el suicidio	79
Figura 25. Emociones previas al suicidio consumado.....	81
Figura 26. Emociones antes de ser ejecutados	82

Índice de tablas

Tabla 1. Resultados de la clasificación entre notas de depresión e ideación suicida	51
Tabla 2. Resultados de la clasificación realizada por Sarsam et al. (2022)	54
Tabla 3. Ejemplos de publicaciones suicidas y no suicidas	55
Tabla 4. Resultados de la clasificación realizada por Tadesse et al. (2019)	58
Tabla 5. Ejemplos de oraciones del conjunto CEASE.....	64
Tabla 6. Distribución de las notas del conjunto CEASE	65
Tabla 7. Descripción de los conjuntos de datos utilizados	65
Tabla 8. Ejemplo de matriz resultante con FT-FDI	68
Tabla 9. Generación de tópicos en notas de ideación suicida	68
Tabla 10. Desempeño de clasificación con ALD y MVS	73
Tabla 13. Resultados de la clasificación entre notas asociadas con la muerte.....	76
Tabla 14. Vocabulario más frecuente en las notas relacionadas con la muerte	77
Tabla 15. Resultados de la clasificación entre notas suicidas y no suicidas	79
Tabla 16. Términos más frecuentes en las notas suicidas y ajenas al suicidio	80
Tabla 17. Resultados de la clasificación entre notas previas a la muerte	82
Tabla 18. Vocabulario más frecuente antes de morir	83

Índice de ecuaciones

Ecuación 1. Modelo de espacio vectorial.....	32
Ecuación 2. Modelo matemático de la representación vectorial	32
Ecuación 3. Frecuencia de término	33
Ecuación 4. Frecuencia de documento inversa.....	33
Ecuación 5. Expresión de FT-FDI	34
Ecuación 6. Hiperplano óptimo	38
Ecuación 7. Teorema de Bayes.....	40
Ecuación 8. Cálculo de la precisión	46
Ecuación 9. Cálculo de la exhaustividad	46
Ecuación 10. Cálculo de la medida F	46



Capítulo 1

Introducción

El suicidio es un fenómeno universal tan antiguo como la humanidad (Ortega González, 2018). Generalmente, ha sido tachado, estigmatizado y despreciado con el objetivo de ocultarlo debido a la falta de sensibilidad sobre el tema (Organización Mundial de la Salud, 2021; Ortega González, 2018). A pesar de ser un tema controversial, se ha discutido desde la antigua Grecia, donde diversos padres de la filosofía han aportado argumentos a favor y en contra de este suceso (Olivares Pérez, 2019; Ortega González, 2018).

El término suicidio proviene del latín *sui*: de sí mismo y *caedere*: matar (Águila Tejada, 2011; Ortega González, 2018; Real Academia Española, 2023); no obstante, su significado ha variado según los principios filosóficos, religiosos e intelectuales de cada cultura (Ortega González, 2018). Por tal motivo, en este estudio adoptamos la definición de Shneidman (1985): "*Acto consciente de aniquilación autoinducida, que se entiende mejor como un sufrimiento multidimensional en una persona vulnerable que percibe este acto como la mejor solución a sus problemas*".

El comportamiento suicida contempla todos los posibles actos de autolesión con la intención de causar la muerte (Sawhney et al., 2018). Las principales etapas del suicidio son: la ideación suicida, la planeación y finalmente los intentos (Abdulsalam y Alhothali, 2022; Barroso Martínez, 2019; Sawhney et al., 2018). La ideación suicida refiere a las ideas, pensamientos e intenciones del individuo de acabar con su vida (Abdulsalam y Alhothali, 2022; Sarsam et al., 2022; Sawhney et al., 2018); en el plan suicida, se determinan los métodos específicos que una persona puede utilizar para lograr su muerte (Abdulsalam y Alhothali, 2022); el intento suicida describe el comportamiento autohiriente con el propósito de conducir a la propia muerte (Abdulsalam y Alhothali,

2022; Fehling y Selby, 2021; Ortega González, 2018). En el peor de los casos el intento resulta en un acto consumado, mismo que hace referencia a que la persona ha conseguido intencionalmente acabar con su vida (Echeburúa, 2015; Ortega González, 2018).

Las conductas suicidas son un fenómeno complejo en el que influyen múltiples factores de riesgo, tal como: las consideraciones biológicas, clínicas, psicológicas, emocionales, sociales y conductuales (Abdulsalam y Alhothali, 2022; Castillo-Sánchez et al., 2020; Murcia y Vargas, 2020).

La Organización Mundial de la Salud (OMS), en su último informe¹ del 2021, menciona que cada año alrededor de 703,000 personas deciden quitarse la vida a nivel mundial, indicando que por cada acto consumado existen más intentos. Algunos autores afirman que las tasas de suicidio han aumentado globalmente hasta el 60% a nivel mundial en los últimos 45 años (Arroyo Fernández y Bertomeu Ruiz, 2012; Barroso Martínez, 2019).

El suicidio es considerado un problema grave de salud pública que afecta a toda la población (Caicedo, 2020). Este se posiciona entre las primeras cinco causas de muerte no natural más frecuente en adolescentes de entre 15 y 29 años (Abdulsalam y Alhothali, 2022; Castillo-Sánchez et al., 2020; Organización Mundial de la Salud, 2021).

La adolescencia es una etapa crucial en el desarrollo del ser humano por ser el periodo clave de transición entre la niñez y la vida adulta (Cox, 2020; Murcia y Vargas, 2020). En esta etapa los adolescentes son considerados vulnerables a adquirir diferentes conductas, como consecuencia del desequilibrio emocional y psicológico que enfrentan, primordialmente por la búsqueda de su identidad (Cox, 2020; De la Cruz y Zúñiga, 2017; Murcia y Vargas, 2020). Esta búsqueda involucra la propia autoestima y autoimagen de los adolescentes que dependen en gran parte de la evaluación y aceptación de grupos afines (Cox, 2020; Murcia y Vargas, 2020).

Los adolescentes de la actualidad corresponden a la llamada "Generación² Z" (misma que refiere a aquellos nacidos a partir de 1996), que los define como "nativos digitales" (Cox, 2020; Dimock, 2019; The Center For Generational Kinestics, 2018). Los adolescentes pertenecientes a esta generación están acostumbrados al uso de las tecnologías y

¹ <https://www.who.int/es/news-room/fact-sheets/detail/suicide>

² Conjunto de personas que, habiendo nacido en fechas próximas, adoptan actitudes, preferencias y valores similares a lo largo de su vida (Álvarez Ramos et al., 2019).

redes sociales (Cox, 2020); por consiguiente, usan estos medios para relacionarse, generar y compartir contenido digital (Álvarez Ramos et al., 2019).

Actualmente, las redes sociales e internet son el principal medio de comunicación de los adolescentes. Estos medios de comunicación han remplazado a los medios tradicionales, al permitir que usuarios de todo el mundo se comuniquen de manera ilimitada, frecuente e instantánea (Abdulsalam y Alhothali, 2022; Castillo-Sánchez et al., 2020; Martínez Restrepo y Jossa Jaramillo, 2021).

Principalmente, las redes sociales permiten a los usuarios compartir sus experiencias personales, conocer la opinión de los demás y expresar sus opiniones, pensamientos y emociones sobre diversos temas (Abdulsalam y Alhothali, 2022; Sarsam et al., 2022). Entre los temas que se comparten a través de las redes sociales, se encuentran los problemas de salud, tal como: el suicidio (Abdulsalam y Alhothali, 2022; Farzindar e Inkpen, 2015). Diversos autores mencionan que estos medios sociales juegan un doble papel (factor de riesgo y protección) en la prevención del suicidio (Durkee et al., 2011; Parrott et al., 2019).

Por un lado, el uso inadecuado de estas plataformas podría permitir a los usuarios acceder a contenidos que promueven el suicidio. Además, las redes sociales facilitan la difusión de pensamientos, pactos suicidas e incluso el acceso a recursos materiales para realizar un intento de suicidio (Arroyo Fernández y Bertomeu Ruiz, 2012), conduciendo a los adolescentes a tomar malas decisiones y/o reforzar sus pensamientos y conductas suicidas (Durkee et al., 2011; Murcia y Vargas, 2020).

Por otro lado, los contenidos de redes sociales pueden ser utilizados como factor protector y preventivo, debido a que permiten identificar patrones lingüísticos que frecuentemente aparecen en publicaciones textuales relacionadas con pensamientos suicidas (Murcia y Vargas, 2020; Parrott et al., 2019). En este sentido, Shneidman (1985) afirma que la manera de escribir de una persona, junto con su modo de expresión, pueden proporcionar pistas suficientes para determinar su personalidad y estado emocional.

Es importante enfatizar que las emociones juegan un papel significativo en las conductas suicidas debido a que la mayoría de las personas que expresan intenciones suicidas describen su motivación como resultado de diversos estados emocionales. Por lo tanto, emplean frases como: "*vida cansada*", "*tengo miedo*", "*lo siento*" y "*nunca más*" (Ortega González, 2018; Tadesse et al., 2019). En este sentido, algunos autores

afirman que el suicidio es una forma de afrontar experiencias que causan dolor emocional, de modo que las personas que experimentan mayor presencia de emociones negativas tienen mayor riesgo de quitarse la vida (Bonet Mas et al., 2020; De la Cruz y Zúñiga, 2017; Sarsam et al., 2022; Vecina Jiménez, 2006).

Las emociones y sentimientos están estrechamente relacionadas, sin embargo, Pallarés (2010) define que las emociones son estados de ánimos intensos de corta duración; estas son producto de ideas, recuerdos, deseos y sentimientos, que generan acciones impulsivas y precipitadas de primera impresión. Por el contrario, los sentimientos son estados afectivos de baja intensidad y larga duración que expresan una opinión privada como resultado de las emociones (Akhtar et al., 2020; Pallarés, 2010).

En general, las emociones y los sentimientos se distinguen por su intensidad y duración. Pero particularmente, las emociones se clasifican en diversas clases predefinidas, tales como: alegría, miedo, ira y tristeza (Akhtar et al., 2020); en cambio, los sentimientos se categorizan según su polaridad: positiva, negativa y neutral (Awwalu et al., 2020; Pallarés, 2010).

Uno de los principales signos de alarma de la ideación suicida es la expresión explícita o implícita de la planificación de muerte que se puede manifestar en notas de despedida, también denominadas: *notas de suicidio* (Echeburúa, 2015; Monwabisi et al., 2022). Las notas de suicidio son mensajes textuales que el autor deja con la intención de que sean vistas posterior a su muerte. En estas notas, el suicida declara sus últimas palabras y justifica su deseo de morir expresando las circunstancias, emociones, pensamientos y sentimientos que lo llevaron a tomar tal decisión (Fata et al., 2021; Ghosh et al., 2022; Monwabisi et al., 2022).

Las notas explícitas de suicidio se caracterizan por incluir frases o palabras que hacen referencia a la intención y deseo de morir, por ejemplo: "*la vida no merece la pena*", "*lo que quisiera es morirme*", "*para vivir de esta manera lo mejor es estar muerto*" (Fata et al., 2021; Monwabisi et al., 2022; Murcia y Vargas, 2020).

Por el contrario, en las notas implícitas de suicidio no se declara directamente el deseo de morir. En este tipo de notas se incluyen frases como: "*ya no es tolerable*", "*fin de mi viaje*", "*ahora la batalla ha terminado*" (Monwabisi et al., 2022; Murcia y Vargas, 2020). Sin embargo, algunos autores sostienen que el estilo de escritura y el modo de expresión de una persona pueden ofrecer suficientes detalles para identificar los patrones lingüísticos que permitan inferir sobre su personalidad, estado emocional y evaluar la

tendencia que tiene en relación con el suicidio u otros trastornos de salud mental (Sarsam et al., 2022; Shneidman, 1985).

Dentro de las diez características más comunes de las personas suicidas, Shneidman (1985) menciona que el estado cognitivo común en el suicidio es la ambivalencia, es decir, coexisten dos emociones opuestas, tal como el amor y el odio (Chávez-Hernández y Leenaars, 2010). Algunas investigaciones concluyen que en las notas suicidas mayormente se encuentran emociones negativas, tales como: miedo, tristeza, ira y desesperanza (Monwabisi et al., 2022; Sarsam et al., 2022; Weishaar y Beck, 1992); mientras que las emociones positivas se presentan en niveles muy bajos o nulos (Rodríguez-Esparza et al., 2021).

Debido al incremento de uso de las redes sociales y a su influencia actual en los jóvenes (grupo social más familiarizado con los medios digitales y entre los principales afectados por el suicidio), es más común que ellos compartan sus emociones, pensamientos e incluso revelen sus intenciones suicidas en estas redes (Monwabisi et al., 2022; Rajpura et al., 2021).

El anonimato es la razón principal por la cual los adolescentes prefieren confesar abiertamente sus intenciones suicidas en las redes sociales (Durkee et al., 2011; Rajpura et al., 2021), en lugar de hablarlo personalmente con su círculo social cercano. De esta manera evitan verse expuestos a la estigmatización y a la posición negativa de hablar de un problema de salud mental (Durkee et al., 2011; Parrott et al., 2019). Troya-Capa et al. (2020) mencionan que otra razón por la que los jóvenes prefieren compartir sus pensamientos suicidas en las plataformas sociales es a causa de que la consulta profesional no está al alcance de todos.

A lo largo de los años se han desarrollado estrategias en diferentes áreas de investigación con el objetivo de prevenir y reducir el número de muertes por suicidio. Recientemente, estas estrategias se han centrado en el análisis de contenido de las redes sociales e internet debido a la creciente predisposición de los jóvenes por compartir sus pensamientos suicidas a través de estos canales.

Desde el punto de vista computacional se han desarrollado modelos para detectar textos con contenido de ideación suicida en redes sociales, como Twitter³ y Reddit

³ Servicio gratuito de difusión social que permite publicar mensajes cortos llamados tuits, los cuales están limitados a 280 caracteres (Abdulsalam y Alhothali, 2022; Akhtar et al., 2020)

(Abdulsalam y Alhothali, 2022; Castillo-Sánchez et al., 2020). Principalmente, se han empleado técnicas de aprendizaje automático y procesamiento del lenguaje natural. Estas técnicas permiten a los investigadores examinar de manera automática el contenido de millones de publicaciones en periodos cortos de tiempo y comprender las tendencias y patrones textuales de dichos escritos (Parrott et al., 2019).

Las investigaciones actuales han alcanzado resultados prometedores en identificar automáticamente notas de suicidio mediante modelos computacionales. Sin embargo, la mayoría de los estudios aplican una clasificación directa.

La clasificación directa se centra en reconocer si una nota es de ideación suicida con base en las palabras utilizadas. Por ejemplo, es posible verificar si en el escrito existen palabras tales como: *suicidio*, *muerte*, *plan* o *ideación suicida*. La presencia de estos términos permite identificar la probabilidad de que un texto sea una nota suicida.

Por ejemplo, en la investigación realizada por Jain et al. (2022) se aborda la detección de texto depresivo y texto suicida mediante técnicas de procesamiento del lenguaje natural y aprendizaje automático. Para lograr tal objetivo, los autores recolectaron 600,000 textos explícitos de la red social Reddit. Las publicaciones relacionadas con depresión se extrajeron del subreddit *r/depression*; mientras que la otra mitad correspondiente a suicidio, se obtuvieron del subreddit *r/SuicideWatch*. Posterior a la recolección, preprocesaron los datos y a continuación entrenaron diferentes algoritmos de aprendizaje automático, tal como: regresión logística, Naïve Bayes, máquinas de vectores de soporte y bosque aleatorio. De esta clasificación, la regresión logística obtuvo el mejor desempeño (0.79), seguido por bosque aleatorio, *máquinas de vectores de soporte* y Naïve Bayes.

Con el objetivo de mejorar la precisión de la clasificación directa se han desarrollado sistemas multclasificadores (también conocidos como *combinación*, *conjunto* o *ensamble de clasificadores*), los cuales se categorizan por la forma en que se integran (Duval et al., 2012; Lu, 1996). Dentro de esta categorización existe la arquitectura en cascada (Sección 2.2.3), la cual consta de una secuencia de clasificadores que pueden presentar n niveles (normalmente dos), donde el resultado de un clasificador se introduce como entrada al del próximo nivel (Lu, 1996; Ramírez et al., 2015).

El ensamble de clasificadores se ha empleado previamente en la detección de notas suicidas. Por ejemplo, en el estudio de Sarsam et al. (2022) se utiliza el método de *embolsado* que corresponde a la arquitectura paralela (Sección 2.2.3). Este modelo

obtuvo el mejor rendimiento en la clasificación de notas suicidas y no suicidas, en comparación con los clasificadores de k vecinos más cercanos, regresión logística multinomial y árbol de decisión. Sin embargo, es importante tener en cuenta que los datos de este estudio corresponden a notas explícitas, lo que significa que fueron elegidos por tener palabras clave sobre el suicidio, tal como: *suicidio*, *pensamientos suicidas* y *contemplando el suicidio*, por mencionar algunos ejemplos.

Tadesse et al. (2019) propusieron otro modelo de ensamble de clasificadores para la detección de ideación suicida. Su enfoque consta de la concatenación de los siguientes clasificadores: memoria a largo y corto plazo, y la red neuronal convolucional. Este enfoque combinado supera a otros clasificadores de aprendizaje supervisado, por ejemplo: máquinas de vectores de soporte, naïve Bayes, bosque aleatorio y aumento de gradiente extremo.

Además, Tadesse et al. (2019) exponen que concatenar características a diferente nivel lingüístico puede mejorar el desempeño de los clasificadores. En general, el estudio de estos autores aporta resultados interesantes, sin embargo, las notas empleadas en esta investigación también fueron seleccionadas por reflejar contenido explícito relacionado con el suicidio. Otra desventaja es que los autores determinan de manera manual y subjetiva las emociones que caracterizan a este tipo de textos mediante el análisis del vocabulario utilizado en las clases de estudio.

Como se puede observar, los avances computacionales desarrollados para identificar notas de ideación suicida continúan enfocándose en el análisis de las palabras explícitas utilizadas, haciendo que información relevante que también se expresa en estas notas, como el estado emocional, pierda importancia.

Algunos otros autores como Malini y Tan (2016), Gunn y Lester (2012) han estudiado notas suicidas con contenido implícito, sin embargo, las notas de estudio son seleccionadas manualmente y el análisis de estos textos se realiza con el diccionario LIWC de consulta lingüística y recuento de palabras. Con este recurso, los autores obtienen información sobre las palabras, emociones, preposiciones y pronombres más empleados en este tipo de textos. Muestran la información de manera estadística y no la utilizan para otros fines.

Por ejemplo, Gunn y Lester (2012) analizan 145 publicaciones de Twitter que pertenecen a una joven que consumió el acto suicida. Los autores concluyen que en el tiempo más cercano a la muerte las publicaciones eran de mayor longitud, disminuyó el uso de palabras que referenciaban a actos actuales, aumentando las referencias al pasado.

Otro hallazgo fue se emplearon más términos que referencian a emociones positivas, disminuyendo significativamente las emociones negativas.

Considerando lo anterior, en este estudio se propone un ensamble de dos clasificadores con arquitectura en cascada que implementa diferentes etapas. En primer lugar, consideramos la recopilación de diferentes conjuntos de datos en idioma inglés, debido a que no se dispone públicamente de un corpus sólido de este tipo de textos. Posteriormente, se limpian los datos y se utilizan algunos métodos de generación de características para analizar los textos a diferente nivel lingüístico (léxico y semántico).

Las características obtenidas previamente se utilizan para que el primer clasificador determine la distribución emocional de cada texto de estudio. A continuación, la distribución emocional alimenta al segundo clasificador que determina que texto es suicida en función de esa entrada. Finalmente, se evalúa el modelo mediante validación cruzada en términos de la precisión, exhaustividad y medida F.

1.1 Planteamiento del problema

Detectar ideación suicida a través del análisis textual es una tarea desafiante debido a que únicamente se cuenta con las palabras plasmadas en el documento para determinar si el autor tiene tendencia al suicidio. En este tipo de estudio no se tiene a la persona enfrente para observar su comportamiento o expresiones.

Analizar las notas suicidas es una fuente de información importante, debido a que la expresión explícita o implícita de la planificación de muerte es de las principales señales de alerta ante el suicidio. Además, el estudio de estas notas nos permite conocer las circunstancias, emociones, pensamientos y sentimientos; que llevan a la persona a querer concluir con su vida.

Las emociones desempeñan un papel significativo en las conductas suicidas, dado que la mayoría de las personas que expresan pensamientos suicidas describen sus motivos como resultado de diversas emociones, donde prevalecen aquellas con polaridad negativa.

Computacionalmente, se ha desarrollado diferentes investigaciones que han alcanzado resultados prometedores en la detección automática de notas suicidas. La mayoría de estos desarrollos emplean la metodología de clasificación directa; el ensamble de clasificadores se utiliza muy pocas veces. En cualquiera de los casos, la clasificación sigue enfocando al contenido explícito de las notas, restando importancia

a características relevantes que, de igual manera, pueden extraerse de dichos textos, tal como: el estado emocional.

Por otro lado, los estudios que han abordado el análisis de notas suicidas implícitas no realizan la identificación automática de estas notas, además con base en el recuento de las palabras únicamente proveen información estadística sobre el tipo de palabras, pronombres o emociones que predominan en estas notas; ignorando la gramática y el orden en que se emplean.

En consecuencia, el problema planteado fue: *Detectar automáticamente notas de ideación suicida mediante el análisis emocional expresado en textos no explícitos.*

Teniendo en cuenta lo anterior, la pregunta que guio la presente investigación fue:
¿Cuál es la relación que existe entre las emociones y la ideación suicida expresada en texto?

1.2 Justificación

Hoy en día el suicidio representa un grave problema de salud pública, debido a que afecta a toda la población. Este fenómeno se posiciona entre las primeras cinco causas de muerte no natural más frecuente en adolescentes de entre 15 y 29 años, quienes además, por estar acostumbrados al uso de las tecnologías y redes sociales utilizan estos canales como principal medio de comunicación.

En el área de las ciencias de la computación se han desarrollado diferentes modelos para identificar notas suicidas de redes sociales, dando resultados competentes. La clasificación de estos textos se ha realizado tanto con el enfoque directo como de ensamble; sin embargo, en ambos casos se han estudiado notas explícitas sin tener en cuenta que no todas las personas con pensamientos suicidas expresan abiertamente tal intención; por el contrario, describen su intención de morir como consecuencia de diversos actos que afectan sus emociones.

En consecuencia, es necesario desarrollar un modelo computacional que permita detectar automáticamente notas de ideación suicida con contenido no explícito mediante el análisis emocional expresado en dichos textos. De esta manera, se podrán detectar notas suicidas independientemente de que el contenido sea explícito o no; además, se conocerá la relación existente entre las emociones y los pensamientos suicidas, que se expresan textualmente.

1.3 Delimitación

Este estudio está enfocado a la detección automática de notas de ideación suicida mediante el análisis emocional expresado en dichos textos. Para lograr tal objetivo se emplea el ensamble de clasificadores configurados en la arquitectura en cascada; por tal motivo se limita a los siguientes aspectos:

- Se considera el estudio de notas suicidas y notas de personas sentenciadas.
- Se analizan textos exclusivamente en idioma inglés.
- La clasificación se realiza únicamente con el ensamble de dos clasificadores en cascada.

1.4 Hipótesis

Si se utiliza la distribución emocional que se expresa en las notas de suicidio, entonces se podrán detectar automáticamente este tipo de textos.

1.5 Objetivo general

Detectar automáticamente notas de ideación suicida mediante el análisis emocional expresado en textos no explícitos.

1.6 Objetivos específicos

- Recolectar y analizar los conjuntos de datos.
- Aplicar preprocesamiento a los conjuntos de datos.
- Experimentar con diferentes métodos de generación de características.
- Experimentar con diferentes clasificadores para obtener la distribución emocional.
- Experimentar con diferentes clasificadores para identificar notas suicidas.
- Evaluar el rendimiento de los clasificadores.
- Analizar el impacto de las emociones en la clasificación de notas suicidas.

1.7 Organización de la tesis

La estructura de la tesis está organizada de la siguiente manera:

- Capítulo 1. Introducción: En este capítulo se describe la problemática donde está inmersa la investigación.
- Capítulo 2. Marco Teórico: Este capítulo aborda brevemente los elementos teóricos que fundamentan la metodología propuesta.

- Capítulo 3. Estado del Arte: En esta sección se analizan diferentes investigaciones que han empleado técnicas de aprendizaje automático y procesamiento del lenguaje natural para detectar notas de ideación suicida.
- Capítulo 4. Método Propuesto: En este capítulo se describen las etapas del enfoque propuesto.
- Capítulo 5. Experimentación: Se presentan los resultados obtenidos de cada caso de estudio realizado.
- Capítulo 6. Conclusiones: Se indican las conclusiones obtenidas con base en los resultados alcanzados.



Capítulo 2

Marco teórico

En este capítulo se describen las bases teóricas computacionales que fundamentan esta investigación.

2.1 Análisis de textos

Debido a la gran cantidad de información que se comparte diariamente en las plataformas sociales y a la capacidad limitada de las personas para analizar dicho contenido, es necesario utilizar métodos computacionales que permitan extraer conocimiento útil a partir de grandes colecciones de datos (Farzindar e Inkpen, 2015; Han y Kamber, 2006; Žižka et al., 2019).

El análisis de textos, también denominado minería de textos, es la metodología que se ocupa específicamente del estudio de textos en lenguaje natural (Mariñelarena-Dondena et al., 2017; Sarkar, 2016; Viera, 2017). Se entiende por lenguaje natural al medio de comunicación que utiliza el ser humano de manera cotidiana para expresarse con las demás personas (Cortez et al., 2009). Dicha comunicación puede llevarse a cabo de distintas formas, por ejemplo, de manera oral o escrita. Algunas lenguas naturales son: el español, inglés, japonés; entre otros (Chowdhary, 2020; Cortez et al., 2009; Ledeneva y García Hernández, 2017).

La minería de texto es un campo multidisciplinario que se fundamenta en otras disciplinas de las ciencias de la computación, tal como: la minería de datos, el aprendizaje automático, la lingüística computacional, la estadística y el procesamiento del lenguaje natural; por mencionar algunos ejemplos (Chowdhary, 2020; Talib et al., 2016; Troya-Capa et al., 2020; Viera, 2017). La minería de textos se basa principalmente en estas áreas porque le permiten extraer patrones ocultos y representativos de gran

cantidad de documentos textuales expresados en lenguaje natural (Mariñelarena-Dondena et al., 2017; Talib et al., 2016; Troya-Capa et al., 2020; Viera, 2017).

Específicamente, la minería de texto está relacionada con la minería de datos debido a que ambas disciplinas buscan extraer información relevante de grandes conjuntos de datos (Talib et al., 2016; Viera, 2017). La diferencia entre ambas radica en que la minería de textos se centra en datos no estructurados, como el texto; mientras que la minería de datos es aplicable tanto a datos estructurados, como no estructurados (Han y Kamber, 2006; Viera, 2017).

El aprendizaje automático es uno de los principales enfoques de la inteligencia artificial (Rouhiainen, 2018). Este campo de estudio se ocupa de construir programas computacionales que puedan aprender automáticamente patrones de los datos y utilizar la experiencia adquirida para mejorar su desempeño en la toma de decisiones de alguna tarea (Lee, 2019; Rouhiainen, 2018; Viera, 2017).

La lingüística es el estudio científico del lenguaje que incluye la forma, sintaxis, significado y semántica del lenguaje en el contexto de uso (Sarkar, 2016). En resumen, la sintaxis es el área lingüística que estudia el orden y la relación de las palabras utilizadas en una frase o en una oración (Sarkar, 2016). En otro caso, la semántica implica el estudio del significado de las palabras en referencia a objetos del mundo real (Chowdhary, 2020; Sarkar, 2016). En consecuencia a lo anterior, la lingüística computacional se ocupa de desarrollar algoritmos computacionales que permiten el estudio del lenguaje natural considerando sus diferentes niveles lingüísticos (Chowdhary, 2020).

El procesamiento del lenguaje natural se refiere a un conjunto de técnicas y métodos necesarios para que la computadora sea capaz de procesar el lenguaje humano (Chowdhary, 2020; Ledeneva y García Hernández, 2017).

Entre las principales tareas automáticas del análisis de texto se encuentran: el análisis de opiniones, la identificación de temas ocultos, clasificación, agrupación y resúmenes de textos (Sarkar, 2016; Viera, 2017; Žižka et al., 2019).

2.2 Clasificación de textos

La clasificación automática de textos, también conocida como categorización de textos o clasificación de documentos (Russell y Norvig, 2016; Sarkar, 2016), es una tarea

específica del procesamiento del lenguaje natural que consiste en asignar etiquetas de clase a los documentos de acuerdo a su contenido, tema o criterio de interés (Castillo-Sánchez et al., 2020; Chowdhary, 2020; Sarkar, 2016).

Se denomina documento a todas las formas de contenido textual del mundo real, por ejemplo: un artículo científico, un escrito personal, las publicaciones de redes sociales, un correo electrónico, entre otros (Mariñelarena-Dondena et al., 2017; Sarkar, 2016). Por lo tanto, cada documento puede contener una oración e inclusive párrafos de palabras (Sarkar, 2016).

De acuerdo con el tipo de información que se disponga para clasificar los documentos, existen principalmente los siguientes dos enfoques:

- *Clasificación supervisada*: se caracteriza por disponer de un conjunto muestra de documentos previamente etiquetados por humanos (Chowdhary, 2020; Duval et al., 2012; Viera, 2017). A este conjunto muestra se le conoce como conjunto de entrenamiento, el cual, se utiliza para que un algoritmo pueda descubrir y aprender las relaciones existentes entre las características extraídas de los datos proporcionados y sus etiquetas, con el fin de predecir la clase a la que pertenece un nuevo documento (Duval et al., 2012; Troya-Capa et al., 2020).

Un clasificador supervisado puede generar algunas de las siguientes salidas: la etiqueta de clase del nuevo objeto, un conjunto de etiquetas ordenadas por la probabilidad de ser la etiqueta correcta o un vector numérico donde cada dimensión representa el grado de pertenencia a cada clase (Duval et al., 2012).

- *Clasificación no supervisada*: en este tipo de clasificación los algoritmos no disponen de datos previamente etiquetados, ni conocen cuantas clases existen para indicar cómo tendría que ser categorizada la nueva información (Rouhiainen, 2018; Žižka et al., 2019). Por lo tanto, el objetivo de este enfoque es encontrar la manera de agrupar los documento en función de la similitud que presentan (Duval et al., 2012; Žižka et al., 2019).

La clasificación automática de textos tiene diversas aplicaciones, por ejemplo: el filtrado de correo basura, minería de opiniones y la mejora en motores de búsqueda (Dalal y Zaveri, 2011).

2.2.1 Minería de opiniones

El análisis de sentimientos, también conocido como *minería de opiniones*, análisis de afectos, análisis de emociones o minería de reseñas; es el campo de estudio computacional que reconoce las opiniones, sentimientos, valoraciones, actitudes y emociones, que las personas expresan de manera textual hacia diversas entidades como productos, servicios, organizaciones, individuos, problemas, eventos y temas (Awwalu et al., 2020; Lui, 2012; Ramírez et al., 2015).

Lui (2012) declara que los términos de análisis de sentimientos y minería de opiniones se utilizaron por primera vez en 2003; sin embargo, las investigaciones sobre estos temas se realizan desde 2001, convirtiéndose en un área de investigación muy activa debido a las numerosas aplicaciones que se pueden realizar en casi todos los ámbitos. Por ejemplo: gran parte de las empresas se centran en el análisis de reseñas para pronosticar si el servicio o producto brindado tiene o tendrá una aceptación favorable. Otras aplicaciones se centran en eventos sociales, financieros, políticos y sanitarios (Farzindar e Inkpen, 2015; Lui, 2012).

La minería de opiniones en foros médicos ha aumentado con el uso de redes sociales debido a que los pacientes tienden a publicar información sobre su salud, medicación, tratamientos y experiencias emocionales en estos canales (Farzindar e Inkpen, 2015; Lui, 2012). A partir de la información que se comparte públicamente, los investigadores en salud mental y ciencias de la computación tienen la oportunidad de inferir automáticamente el estado de salud de una persona (Farzindar e Inkpen, 2015; Rouhiainen, 2018).

La minería de sentimientos se divide en dos grupos: análisis de grano grueso y análisis de grano fino. El objetivo del análisis de grano grueso es predecir la polaridad del documento, clasificándolo como positivo, negativo o neural (Akhtar et al., 2020; Awwalu, et al., 2020; Lui, 2012); por otro lado, el análisis de grano fino, también denominado análisis de emociones surgió como una tarea más específica del análisis de opiniones (Desmet y Hoste, 2013; Farzindar e Inkpen, 2015).

El análisis de emociones se centra en determinar las emociones plasmadas en el documento considerando diferentes clases como: alegría, ira, miedo y amor; por citar algunos (Akhtar et al., 2020; Awwalu et al., 2020; Desmet y Hoste, 2013).

2.2.2 Clasificación directa

La Figura 1 muestra el proceso genérico de la clasificación supervisada de textos mediante el enfoque directo. Las fases principales de este tipo de clasificación son: 1) recopilación del conjunto de entrenamiento, 2) preprocesamiento de datos, 3) extracción de características, 4) entrenamiento del modelo de clasificación y 5) evaluación del modelo (Chowdhary, 2020; Dalal y Zaveri, 2011; Talib et al., 2016).

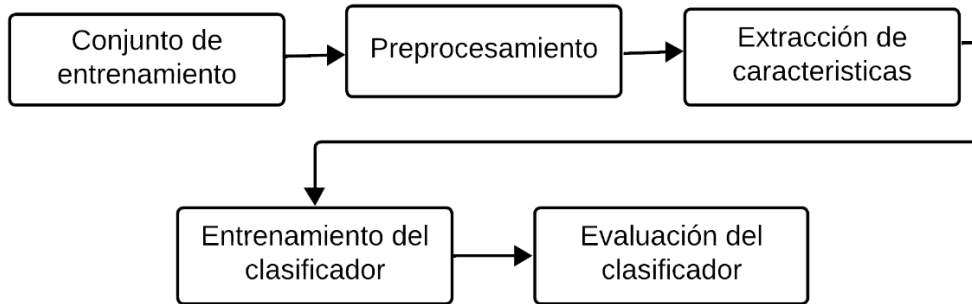


Figura 1. Metodología de la clasificación directa

La recopilación de datos de entrenamiento se refiere al conjunto de documentos no estructurados que han sido etiquetados correctamente de forma manual (Sarkar, 2016). Posteriormente, estos documentos textuales (datos brutos) deben transformarse en un formato que los modelos de aprendizaje puedan comprender (Tadesse et al., 2019); para lograr tal objetivo es necesario eliminar el ruido lingüístico (errores ortográficos, gramaticales, de puntuación, palabras vacías y otros elementos no relevantes para la tarea en cuestión) mediante diferentes métodos de limpieza y normalización de textos (Farzindar e Inkpen, 2015; Han y Kamber, 2006; Tadesse et al., 2019; Talib et al., 2016).

A continuación, el módulo de extracción de características recibe los documentos limpios y utiliza diferentes técnicas para extraer los atributos relevantes que ayudarán en la toma de decisiones (Sarkar, 2016; Talib et al., 2016). Hasta esta etapa, los documentos textuales se transforman automáticamente en matrices o vectores numéricos debido a que los algoritmos de aprendizaje automático así los requieren (Mariñelarena-Dondena et al., 2017; Talib et al., 2016). Tras la obtención de las características se emplea un clasificador de aprendizaje automático, el cual se entrena para determinar la clase de cada documento (Chowdhary, 2020). Finalmente, se evalúa el desempeño del clasificador con un nuevo conjunto de documentos (Dalal y Zaveri, 2011; Talib et al., 2016).

Algunos ejemplos de algoritmos que pueden emplearse en la clasificación supervisada de documentos, son los siguientes: máquinas de vectores de soporte, naïve Bayes, redes neuronales artificiales, Regresión Logística (RL), bosque Aleatorio (BA) y árboles de decisión (Abdulsalam y Alhothali, 2022; Troya-Capa, et al., 2020; Viera, 2017).

La clasificación directa tiene como objetivo determinar la clase a la que pertenece cada documento según las palabras que contiene (Castillo-Sánchez et al., 2020; Žižka et al., 2019). Sin embargo, Duval et al. (2012) afirman que se ha llegado a la etapa donde la precisión proporcionada por un clasificador no es suficientemente precisa para cumplir los requisitos de determinado problema.

Con el propósito de desarrollar sistemas de clasificación más robustos y altamente precisos que mejoren el rendimiento de los clasificadores individuales, se han propuesto los sistemas multclasificadores, también conocidos como: combinación, *ensamble*, conjunto o fusión de clasificadores (Duval et al., 2012; Ramírez et al., 2015).

2.2.3 Clasificación en cascada

El ensamble de clasificadores parte del hecho de que si se tiene un conjunto de clasificadores base donde sus resultados poseen cierto grado de independencia entre sí, estas salidas pueden complementar unas a otras (Akhtar et al., 2020; Duval et al., 2012).

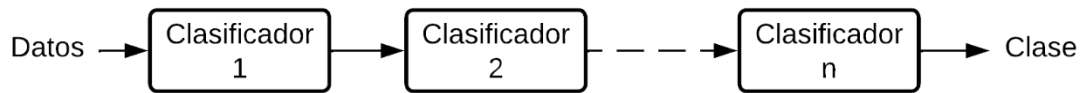
Según Duval et al. (2012) los multclasificadores se categorizan de la siguiente manera:

- Métodos basados en generación de ensamblados: consisten en el entrenamiento independiente de diversos clasificadores, los cuales promedian sus resultados para arrojar un resultado más preciso. El ejemplo más destacado de este método es el embolsado⁴.
- Métodos basados en selección de clasificadores: se usan diversos clasificadores entrenados individualmente y se selecciona el de mejor desempeño.
- Métodos basados en clasificadores: estos métodos combinan los resultados de diversos clasificadores para retornar una respuesta final.

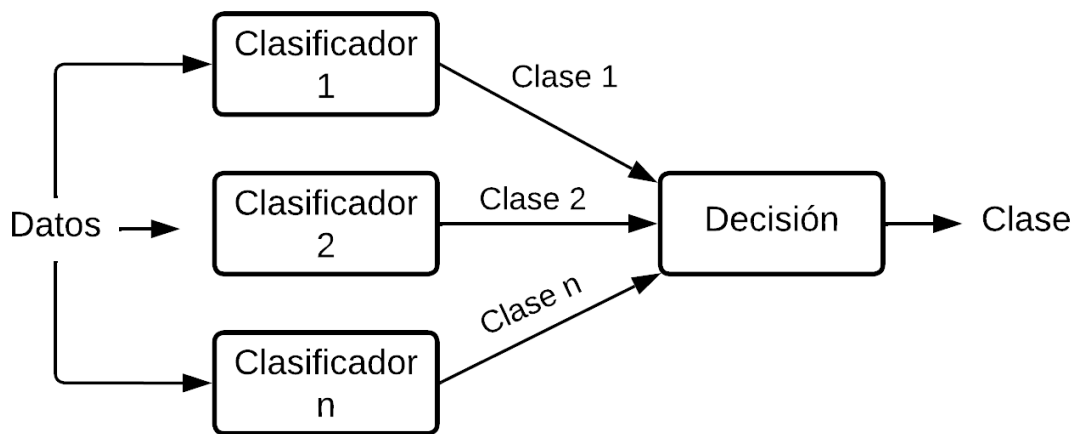
⁴ Método de combinación de clasificadores con arquitectura paralela, donde se selecciona una muestra aleatoria del conjunto de datos de entrenamiento, y cada conjunto es entrenado por un clasificador independiente. De modo que el desempeño del clasificador es el promedio de las predicciones, oh la predicción más repetida (Duval et al., 2012).

Un aspecto notable en los sistemas multclasificadores es la forma en la que se integran. En este sentido, se dividen en tres grupos (Figura 2): cascada(vertical), paralela (horizontal) e híbrida, también conocido como jerárquico (Duval et al., 2012; Lu, 1996).

Arquitectura en cascada



Arquitectura paralela



Arquitectura híbrida

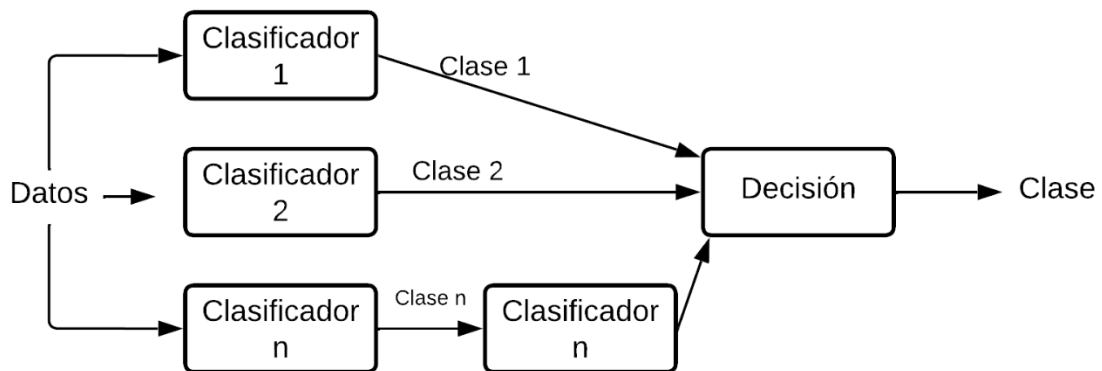


Figura 2. Arquitecturas de los sistemas multclasificadores

La arquitectura en cascada consiste en una secuencia de clasificadores, que pueden presentar n niveles (normalmente dos), donde se permite que el resultado de un clasificador se pueda volver a introducir como entrada del próximo nivel (Duval et al., 2012; Ramírez et al., 2015).

2.2.4 Preprocesamiento

Los documentos en lenguaje natural frecuentemente contienen ruido lingüístico (Han y Kamber, 2006). Se define como ruido lingüístico a los errores ortográficos, gramaticales, de puntuación, palabras vacías y demás elementos no relevantes para la tarea en cuestión (Farzindar e Inkpen, 2015; Han y Kamber, 2006). Estos atributos surgen a partir de la naturaleza humana que comete errores al redactar un texto (Han y Kamber, 2006; Žižka et al., 2019).

Para eliminar o reducir el ruido lingüístico e incrementar la calidad de los documentos que mejoren el entrenamiento y precisión de un sistema de clasificación automático, es necesario preprocesar los documentos (Ledeneva y García Hernández, 2017; Rajpura et al., 2021; Tadesse et al., 2019).

El preprocesamiento de los documentos es una etapa importante en cualquier sistema de procesamiento del lenguaje natural (Hernández-Castañeda, 2017; Sarkar, 2016), debido a que los elementos textuales (palabras, frases, oraciones o cualquier otro token) que resultan de esta etapa son fundamentales para las fases posteriores de cualquier modelo de clasificación de textos, lo que indica que si los documentos no se procesan adecuadamente se pueden producir resultados irrelevantes (Sarkar, 2016).

El preprocesamiento tiene como objetivo transformar los datos no estructurados (datos brutos) en un formato comprensible para los modelos de aprendizaje automático (Tadesse et al., 2019; Viera, 2017). Por tal motivo, se aplican diferentes técnicas de preparación de datos, tal como: la normalización y tokenización (Castillo-Sánchez et al., 2020; Farzindar e Inkpen, 2015).

La etapa de normalización de texto permite limpiar y estandarizar los documentos para reducir el ruido lingüístico (Farzindar e Inkpen, 2015; Sarkar, 2016). Para lograr tal objetivo, generalmente las palabras se transforman a minúsculas, se eliminan las direcciones electrónicas, las contracciones, signos de puntuación, caracteres especiales, números y palabras vacías (Castillo-Sánchez et al., 2020; Sarkar, 2016; Tadesse et al., 2019). Las palabras vacías, también llamadas palabras de parada, son palabras que tienen poco o ningún significado para determinada tarea, por lo cual son comunes en todos los dominios (Ledeneva y García Hernández, 2017; Sarkar, 2016). Ejemplo de este tipo de palabras son los artículos, preposiciones, conjunciones o pronombres (Ledeneva y García Hernández, 2017).

La tokenización es el proceso de dividir o segmentar un documento en caracteres, frases o palabras mediante la localización del punto en que inician y terminan (Farzindar e Inkpen, 2015; Hernández-Castañeda, 2017). La herramienta que realiza esta segmentación se llama tokenizador y a los pequeños segmentos textuales resultantes se les da el nombre de tokens (Farzindar e Inkpen, 2015; Hernández-Castañeda., 2017; Sarkar, 2016). Existen principalmente tres tipos de tokens: palabra, carácter y n-gramas (Jain et al., 2022).

La segmentación de palabras y oraciones son las dos técnicas más utilizadas en el proceso de tokenización (Sarkar, 2016). De acuerdo con Sarkar (2016) la tokenización de palabras es el proceso de segmentar las oraciones de un documento en las palabras que lo componen.

Una palabra es la unidad léxica que tiene una sintaxis, semántica y gramática definida por un lenguaje determinado (Žižka et al., 2019). En este sentido, el sistema de escritura del idioma es el elemento más importante en el proceso de la tokenización porque permite identificar los tokens que deben procesarse, los cuales se indican por los límites entre las unidades lingüísticas del documento. Sin embargo, no todos los lenguajes establecen un límite entre las palabras, por ejemplo, en muchos lenguajes europeos los límites de las palabras están determinados por un espacio en blanco (Hernández-Castañeda, 2017). En cambio, en los textos en lenguajes como el chino o el árabe es complicado identificar los límites entre las palabras debido a que estas se escriben consecutivamente, sin espacios entre ellas (Hernández-Castañeda, 2017; Žižka et al., 2019).

2.2.5 Extracción de características

En el área de aprendizaje automático, las características son los atributos más importantes o frecuentes que permiten indicar la categoría de un documento (Dalal y Zaveri, 2011; Seger, 2018; Sarkar, 2016).

La extracción de características es el proceso de transformar los tokens en una representación computacional adecuada para los clasificadores (Mariñelarena-Dondena et al., 2017). En este sentido, se pueden obtener dos tipos de características: numéricas o categóricas (Seger, 2018; Žižka et al., 2019).

Debido a que un algoritmo es en esencia una operación matemática para optimizar y minimizar los errores de aprendizaje de patrones a partir de un conjunto de datos (Sarkar,

2016), estos algoritmos generalmente reciben las características de los documentos textuales como vectores numéricos (Seger, 2018; Sarkar, 2016).

El modelo de espacio vectorial propuesto por Salton y McGill es el más utilizado para transformar los documentos a vectores (Sarkar, 2016; Žižka et al., 2019). Mediante este modelo, cada documento se transforma a un vector numérico donde la longitud representa el número total de características (términos) distintas para todos los documentos y los valores son los pesos (importancia) de las características (Sarkar, 2016; Žižka et al., 2019).

El vector que representa el número total de n términos distintos en todos los documentos se representa matemáticamente de la siguiente manera:

Ecuación 1. Modelo de espacio vectorial

$$\text{Espacio Vectorial} = \{w_1, w_2, w_3, \dots, w_{Dn}\}$$

Por consiguiente, la representación de un documento al modelo de espacio vectorial se define matemáticamente de la siguiente forma (Ecuación2), donde w_{Dn} denota el peso de la palabra n en el documento D .

Ecuación 2. Modelo matemático de la representación vectorial

$$D = \{w_{D1}, w_{D2}, w_{D3}, \dots, w_{Dn}\}$$

El objetivo del pesado o ponderación de términos, es asignar a cada término un valor numérico que refleje su importancia en relación con los demás. Esto se realiza con la finalidad de seleccionar las características más importantes que representen al texto original (Ledeneva y García Hernández, 2017).

2.2.5.1 Codificación de un solo paso

La codificación de un solo paso, también conocida como Matriz Término-Documento (MTD), es el método de extracción de características más popular en el aprendizaje automático (Hernández-Castañeda, 2017; Seger, 2018).

Esta codificación permite transformar las características a un vector numérico de representación binaria (Hernández-Castañeda, 2017; Seger, 2018). Para lograr tal objetivo, primero se elabora una lista de todas las palabras únicas (w_1, w_2, \dots, w_n) existentes en el conjunto de documentos (Hernández-Castañeda et al., 2020).

Posteriormente, se analiza cada documento para comprobar si el término W_n existe en el documento actual. Si la característica W_n está presente en dicho documento, se establece el valor uno, en caso contrario se asigna el valor cero (Hernández-Castañeda, 2017; Hernández-Castañeda et al., 2020).

2.2.5.2 Frecuencia de término – frecuencia de documento inversa

Se denomina FT-IDF al esquema de ponderación de términos que implica multiplicar la Frecuencia del Término (FT), por la Frecuencia Inversa del Documento, abreviado como FID (Ledeneva y García Hernández, 2017).

De acuerdo con Ledeneva y García Hernández (2017) la FT fue propuesta por Luhn en el año de 1957, mientras que Salton y Buckley propusieron la FDI en 1988.

La FT se ajusta al modelo de bolsa de palabras donde se evalúa el número de veces que la palabra aparece en determinado documento, de manera que se asigna un peso mayor al término que se repite frecuentemente en un documento porque se considera más significativo que aquel que aparece pocas veces (Ledeneva y García Hernández, 2017; Ramírez et al., 2015). Matemáticamente, la FT se representa de la siguiente forma:

Ecuación 3. Frecuencia de término

$$D_i(t_j) = f_{ij}$$

Donde f_{ij} representa la frecuencia del término j en el documento i .

Por otro lado, FDI es la medida que evalúa la distribución de los términos en el documento, por lo tanto, considera que un término cuya frecuencia es menor y que se presenta en pocos documentos discrimina mejor que aquel que aparece con mayor frecuencia en muchos documentos (Ledeneva y García Hernández, 2017; Ramírez et al., 2015).

La fórmula matemática de FDI es la siguiente:

Ecuación 4. Frecuencia de documento inversa

$$D_i(t_j) = \log\left(\frac{N}{n_j}\right)$$

Donde f_{ij} es la frecuencia del término j en el documento i ; N representa el número total de documentos en la colección, y n_j es el número de documentos en los que está presente el término j .

Como se mencionó previamente, la ponderación FT-FDI es el producto de FT por FDI expresándose matemáticamente de la siguiente manera:

Ecuación 5. Expresión de FT-FDI

$$D_i(t_j) = f_{ij} \times \log\left(\frac{N}{n_j}\right)$$

La ponderación FT-FDI refleja tanto la importancia de un término en un documento, como su relevancia en la colección de documentos (Hernández-Castañeda et al., 2020; Ledeneva y García Hernández, 2017).

Dado lo anterior, Ramírez et al. (2015) declaran que el vector de características que se obtiene de FT-FDI para representar un documento, se expresa de la siguiente forma:

$$D_i = (D_{j1}, D_{j2}, D_{j3}, \dots, D_{jn})$$

Donde:

$$D_{ij} = fd_{ij} \times fdi_{ij}$$

Por consiguiente, Blei et al. (2003) exponen que el resultado final de FT-FDI es una matriz X término-documento cuyas columnas contienen los valores ponderados FT-FID de cada uno de los documentos de la colección y las filas corresponden a los documentos. De esta manera, el esquema FT-FID reduce la longitud de los documentos a vectores de longitud fija.

Este esquema de ponderación de términos resulta útil para conocer la relevancia de las palabras o frases en un documento (Hernández-Castañeda et al., 2020). Sin embargo, su limitante es que representa cada documento como una colección de palabras sin ningún orden concreto (Ramírez et al., 2015).

2.2.5.3 Asignación latente de Dirichlet

Una de las técnicas de minería de texto más eficaces en el procesamiento del lenguaje natural para identificar y agrupar las palabras de un conjunto de documentos y conocer los temas latentes (ocultos) que se tratan en ellos, es el modelado de temas (Jelodar et al., 2019). Se define como tema o tópico al grupo de términos semánticamente relacionados (Hernández-Castañeda., 2017; Mariñelarena-Dondena et al., 2017).

La Asignación Latente de Dirichlet (ALD) propuesta por Blei et al. (2003) es uno de los métodos de modelado de temas más populares (Jelodar et al., 2019).

ALD es un modelo probabilístico generativo para colecciones de datos discretos, tal como las colecciones de texto (Blei et al., 2003; Hernández-Castañeda Á., 2017; Jelodar et al., 2019).

El principio fundamental de ALD es representar cada documento del conjunto de datos como una mezcla de k diferentes temas latentes, donde cada tópico se define por una distribución de palabras que están vinculadas entre sí (Blei et al., 2003; Hernández-Castañeda., 2017; Jelodar et al., 2019). Como resultado, cada documento contiene diferentes tópicos en diferentes proporciones (Hernández-Castañeda., 2017; Srivastava y Sahami, 2009).

Los tópicos no tienen una etiqueta, sin embargo, las palabras con mayores probabilidades suelen dar una idea del tema tratado (Farzindar e Inkpen, 2015; Jelodar et al., 2019). El tema de cada tópico es diferente con base en el conjunto de datos analizado (Hernández-Castañeda., 2017).

Blei et al. (2003) indican que el procedimiento que sigue ALD por cada documento w de la colección D , es el siguiente.

- 1) La determinación del número N de palabras que un documento contendrá según la distribución de Poisson.
- 2) La elección de una mezcla de tópicos para el documento de un conjunto fijo de k temas de acuerdo con la distribución de Dirichlet.
- 3) La generación de cada palabra del documento de la siguiente manera:
 - a. Elegir un tema.
 - b. Elegir una palabra de ese tema.

De acuerdo con Hernández-Castañeda. (2017) para encontrar el conjunto más probable de tópicos probablemente tratados en un documento, se debe utilizar el modelo ALD de forma inversa.

ALD resulta ser útil para la selección de características en la tarea de clasificación de textos, reduciendo el tamaño de los vectores de características debido a que la longitud de los vectores que se crean con este método depende del número de tópicos establecido por el usuario (Hernández-Castañeda et al., 2020).

2.2.5.4 Vector de párrafo

El *vector de párrafo*, también conocido como *Doc2Vec* es un algoritmo de aprendizaje no supervisado que permite representar los documentos como vectores numéricos de longitud fija (Hernández-Castañeda et al., 2020; Le y Mikolov, 2014). Este algoritmo fue desarrollado por Le y Mikolov (2014), quienes se inspiraron en la representación de palabras conocida como *Word2Vec*, propuesta por Mikolov et al. (2013).

Le y Mikolov (2014) declaran que la diferencia entre *Word2Vec* (W2V) y *Doc2Vec* (D2V) radica en que el primero predice la siguiente palabra de una oración dadas otras palabras; mientras que D2V supone el siguiente término considerando los diferentes contextos y palabras del párrafo.

Por lo tanto, W2V concatena los vectores que representan el significado semántico de cada palabra en una oración para predecir la siguiente palabra (Figura 3 a). En contraste, D2V vectoriza cada párrafo completo, y además cada palabra para posteriormente concatenar ambos vectores y predecir la siguiente palabra en un contexto (Figura 3 b). De modo que el vector del párrafo actúa como una memoria que recuerda la palabra faltante en el contexto o tópico actual (Le y Mikolov, 2014).

Es importante destacar que Le y Mikolov (2014) afirman que el vector de párrafos se comparte solo en los contextos generados a partir del mismo párrafo, pero no entre párrafos; sin embargo, la matriz vectorial W de palabras se comparte entre los párrafos.

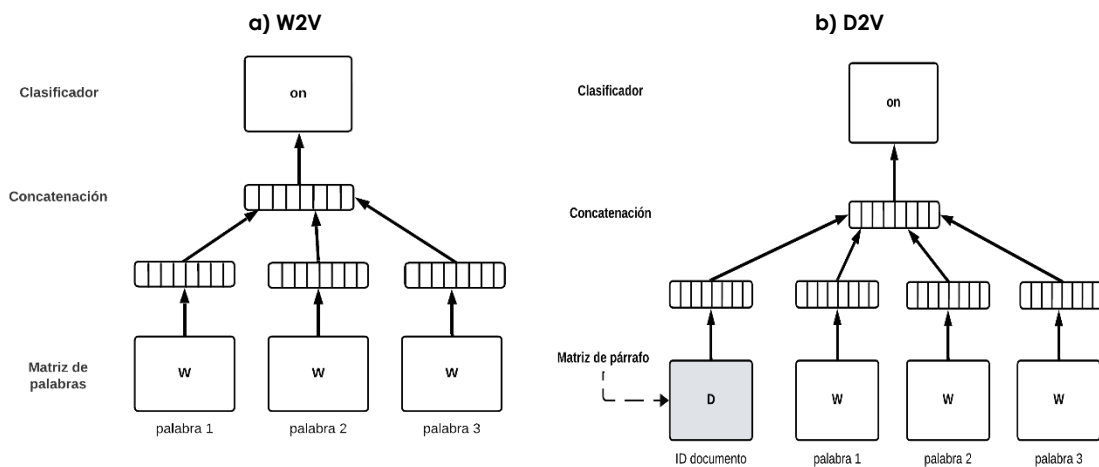


Figura 3. Esquema de representaciones de palabras

La formalización matemática de W2V es descrita de la siguiente manera por Hernández-Castañeda et al. (2020); Le y Mikolov, (2014). Dada una secuencia de palabras de entrenamiento $w_1, w_2, w_3, \dots, w_T$, el objetivo es maximizar la probabilidad logarítmica promedio dada por:

$$\sum_{t=k}^{T-k} \log p(w_t | w_{t-k}, \dots, w_{t+k}).$$

La predicción se realiza mediante un clasificador multiclase (softmax), siguiendo la fórmula:

$$p(w_t | w_{t-k}, \dots, w_{t+k}) = \frac{e^{ywt}}{\sum e^{yi}}$$

Cada y_i de la fórmula anterior es una probabilidad logarítmica no normalizada para cada palabra i de salida, que se calcula como:

$$y = b + Uh(w_{t-k}, \dots, w_{t+k}; W)$$

Donde h se construye mediante la concatenación de los vectores de palabras extraídos de la matriz de palabras w ; U, b son los parámetros softmax. Se utiliza un softmax jerárquico para un entrenamiento rápido. Después, los vectores de palabras basados en redes neuronales son entrenados mediante el descenso de gradiente estocástico, donde el gradiente se obtiene mediante propagación hacia atrás. Cuando el algoritmo converge, las palabras con significados similares deben estar lo más cerca posibles en el espacio vectorial

En comparación con el modelo de palabras, D2V tiene la ventaja de considerar el orden y la semántica de las palabras en el contexto del documento (Hernández-Castañeda et al., 2020; Le y Mikolov, 2014).

2.2.6 Algoritmos de clasificación supervisada

Los algoritmos de clasificación supervisada son técnicas del aprendizaje automático que se utilizan para predecir la etiqueta de la clase a la cual pertenece un nuevo objeto, basándose en un conjunto de datos de entrenamiento (Chowdhary, 2020; Duval et al., 2012; Viera, 2017).

Algunos ejemplos de los algoritmos utilizados en la clasificación de texto son: naïve Bayes multinomial, máquinas de vectores de soporte, redes neuronales, árboles de decisión, entre otros (Duval et al., 2012; Sarkar, 2016; Viera, 2017).

2.2.6.1 Máquina de vectores de soporte

La Máquina de Vectores de Soporte (MVS) es un algoritmo de aprendizaje supervisado que se utiliza para resolver problemas de clasificación (Chowdhary, 2020; Sarkar, 2016; Žižka et al., 2019). De acuerdo con Žižka et al. (2019) este algoritmo fue desarrollado por el matemático Vladimir Vapnik y otros investigadores, en el año de 1963.

Ramírez et al. (2015) afirman que la MVS es un clasificador binario que examina el espacio de características representado por x , con el objetivo de encontrar el hiperplano óptimo (Ecuación 6) que separe los dos posibles valores (clases) de la variable etiquetada y $\in \{+1, -1\}$.

Ecuación 6. Hiperplano óptimo

$$W * x + b = 0$$

Han y Kamber (2006) mencionan que: W es un vector de peso $W=(w_1, w_2, \dots, w_n)$; n es el número de atributos. X es un punto cualquiera en el espacio; b es un valor escalar y el signo indica a qué lado del hiperplano se encuentra el punto X .

Se conoce como hiperplano óptimo a la zona más ancha o margen máximo de separación entre los puntos más próximos de los atributos positivos y negativos del conjunto de datos de entrenamiento formado por N observaciones (Chowdhary, 2020; Ramírez et al., 2015). Los puntos más próximos de cada clase que determinan la zona límite, se denominan vectores de soporte (Chowdhary, 2020; Viera, 2017); estos vectores determinan la dirección y posición del hiperplano (Žižka et al., 2019). Lo mencionado anteriormente se observa gráficamente en la Figura 4.

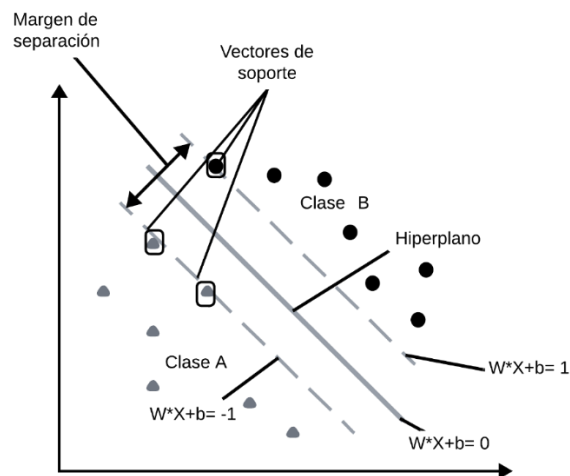


Figura 4. Máquina de vectores de soporte

Una MSV representa los ejemplos de entrenamiento como puntos en el espacio que se distribuyen comúnmente en dos clases (Chowdhary, 2020; Lee, 2019).

Chowdhary (2020) expone que durante la etapa de entrenamiento el algoritmo de MSV recibe como entrada un conjunto S de k elementos x_1, x_2, \dots, x_k etiquetados correctamente y_1, y_2, \dots, y_k .

$$(x_1, y_1), (x_2, y_2), \dots, (x_k, y_k)$$

Donde:

$$\begin{cases} y_i = +1 & \text{si } x_i \in \text{clase A} \\ y_i = -1 & \text{si } x_i \in \text{clase B} \end{cases}$$

Posteriormente el algoritmo utiliza estos ejemplos para establecer las funciones de decisión $D(x)$ que le permitan categorizar un patrón desconocido x .

$$\begin{aligned} x &\in A, && \text{si } D(x) > 0 \\ x &\in B, && \text{en caso contrario} \end{aligned}$$

Para buscar el hiperplano adecuado es necesario mapear el espacio de entrada en un espacio de características de una dimensión mayor y buscar el hiperplano ahí.

La tarea de aprendizaje de una MSV se formaliza con el siguiente problema de optimización:

$$\min_{w,b} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i$$

Sujeto a

$$y_i(w * x_i + b) \geq 1 - \xi_i \forall i \in \{1, \dots, N\} \xi_i \geq 0, \forall i \in \{1, \dots, N\}$$

Una vez que se ha producido el aprendizaje, los ejemplos nuevos se mapean en el espacio de características y se predice su pertenencia a una de las dos clases en función del lado de la brecha en que se encuentren (Chowdhary, 2020; Desmet y Hoste, 2013).

Farzindar e Inkpen (2015) mencionan que la MVS son probablemente los clasificadores más utilizados, debido a su alto desempeño en muchas tareas, por ejemplo, en la clasificación de emociones logran los mejores resultados de clasificación. Además, Tadesse et al. (2019) indican que las MVS tienen buen desempeño en tareas de salud mental.

2.2.6.2 Naïve Bayes multinomial

Naïve Bayes (NB) o también llamado Bayes ingenuo, es un algoritmo probabilístico del aprendizaje automático que pone en práctica el conocido teorema de Bayes (Sarkar, 2016; Viera, 2017). Según Žižka et al. (2019), Han y Kamber (2006) Thomas Bayes propuso este teorema en 1702, de ahí el nombre del algoritmo.

Debido a su simplicidad, aprendizaje rápido y desempeño con gran volumen de datos, NB es uno de los algoritmos más utilizados en la clasificación de textos (Han y Kamber, 2006; Viera, 2017).

Este algoritmo aplica el teorema de Bayes para estimar la suposición ingenua, también conocida como independencia condicional de clase, la cual establece que cada característica de un documento es independiente de los demás atributos (Han y Kamber, 2006; Sarkar, 2016; Viera, 2017).

De acuerdo con Farzindar e Inkpen (2015), Han y Kamber (2006) en la clasificación de textos NB asigna la categoría o clase más probable a un documento d a partir de un conjunto de N clases predefinidas como c_1, c_2, \dots, c_N . La función de clasificación f asigna un documento a una clase maximizando la probabilidad de la siguiente ecuación:

Ecuación 7. Teorema de Bayes

$$P(c|d) = \frac{P(c) \times P(d|c)}{P(d)}$$

Es importante recordar que, en la clasificación de textos, un documento puede representarse mediante un vector de T características, quedando de la siguiente manera: $X = (t_1, t_2, \dots, t_T)$. Asumiendo que todos los atributos $\frac{x}{t_i}$ son independientes dada la categoría c , se puede calcular la probabilidad posterior $P(X|H)$ de X condicionada a H con la siguiente ecuación:

$$\begin{aligned} \underset{c \in \mathcal{C}}{\operatorname{argmax}} P(c|d) &= \underset{c \in \mathcal{C}}{\operatorname{argmax}} P(c) \times \prod_{i=1}^T P(t_i|c) \\ &= P(t_1|c) \times P(t_2|c) \times \dots \times P(t_T|c) \end{aligned}$$

El atributo t_i puede ser un término del vocabulario, un n -grama, la longitud media de una palabra, una característica sintáctica o semántica (Farzindar e Inkpen, 2015)

La meta de la clasificación es encontrar la clase óptima para un determinado documento, calculando la clase que da la probabilidad posterior máxima (Viera, 2017)

El algoritmo Naïve Bayes Multinomial (NBM) es una extensión del popular algoritmo naive Bayes. El NBM principalmente se utiliza en tareas de clasificación donde se tiene más de dos clases (Sarkar, 2016).

Sarkar (2016) menciona que los vectores de características se pueden representar como $P_y = \{P_{y1}, P_{y2}, \dots, P_{yn}\}$ para cada etiqueta de clase y , y el número total de características es n , que puede representarse como el vocabulario total de análisis.

2.2.6.3 Perceptrón multicapa

Las redes neuronales biológicas del cerebro humano son la fuente de inspiración de las Redes Neuronales Artificiales (RNA), por tal motivo, una RNA esta compuesta por elementos que tratan de simular las funciones básicas de la neurona biológica, para realizar tareas complejas de clasificación (Chowdhary, 2020; Viera, 2017).

Viera (2017) menciona que las RNA se representan de manera simplificada mediante un grafo. Los nodos son las unidades de preprocesamiento, los vínculos representan las conexiones sinápticas y la fuerza de conexión entre los nodos se determina por los pesos asociados. Dependiendo de la activación del nodo se envía una señal al nodo vecino, por lo cual, la fuerza de esa señal dependerá del peso asociado a esa conexión. Cada uno de los componentes de las RNA mencionados anteriormente están representados matemáticamente por números reales, de ahí que cada conexión entre neuronas tiene un peso, llamado fuerza sináptica (Chowdhary, 2020). En la Figura 5 se muestra el modelo básico de una neurona propuesto por McCulloch Pitts.

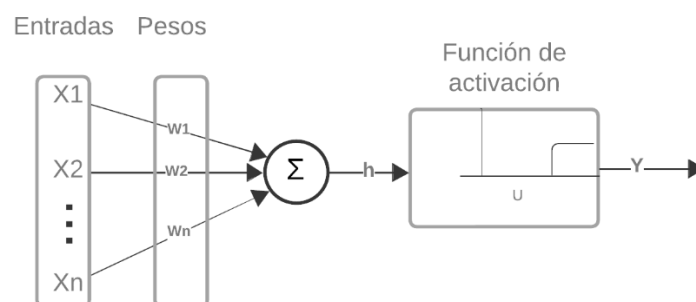


Figura 5. Modelo básico de una red neuronal artificial

La neurona calcula una suma ponderada de sus n señales de entrada $x_j, j = 1, 2, \dots, n$ y genera una salida de uno si la suma está por encima de cierto umbral u , y una salida de cero en caso contrario. Matemáticamente:

$$y = \theta \left(\sum_{j=1}^n w_j x_j - \mu \right)$$

Donde θ es una función de paso, w_j es el peso de la sinapsis asociado con la entrada j^{th} . Para simplificar la notación a menudo se considera el umbral μ como otro peso $w_0 = -\mu$ que está unido a la entrada con una constante $w_0 = 1$ (Chowdhary, 2020).

Una RNA está compuesta principalmente por tres partes: Una capa de entrada (datos de entrada); una o varias capas ocultas; y una capa de salida (Figura 6).

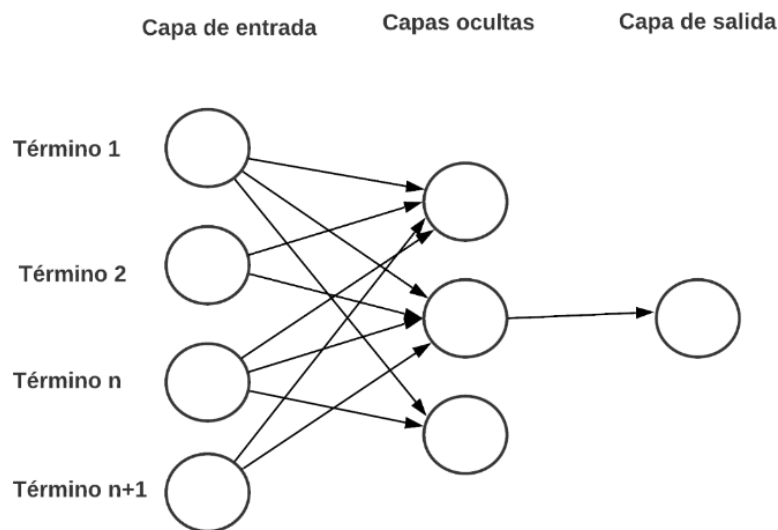


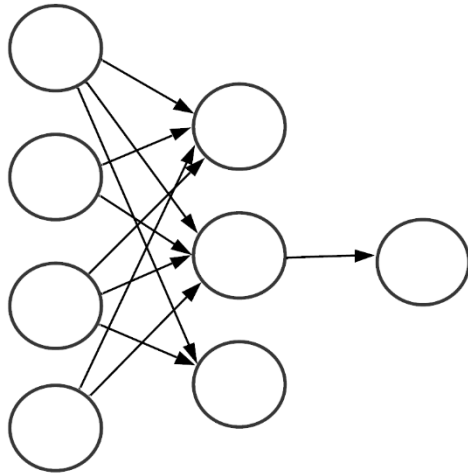
Figura 6. Estructura de una red neuronal

De acuerdo con Viera (2017) los clasificadores de texto basados en RNA están formados por unidades de entrada que representan los términos existentes en los documentos, la capa oculta y las unidades de salida que representan las clases de interés. La cantidad de neuronas en la unidad de salida depende de la cantidad de categorías en que deben ser clasificados los patrones de entrada

Las RNA son redes que utilizan algoritmos de aprendizaje automático (Troya-Capa, et al., 2020; Viera, 2017). De modo que aprenden los pesos de conexión a partir de los patrones de entrenamiento disponibles, cuyo rendimiento se mejora con el tiempo y con la actualización interactiva de sus pesos (Chowdhary, 2020).

Chowdhary, (2020) asegura que las RNA se clasifican de las siguientes formas, según el tipo de conexión que presentan: redes prealimentadas o redes de retroalimentación recurrente. La distinción entre ambas es que el último tipo de redes mencionadas presenta ciclos debido al proceso de retroalimentación que realiza, mientras que en el otro tipo de redes se carece de tal característica (Figura 7). La RNA prealimentada más común es el perceptrón multicapa (PMC).

Redes prealimentadas



Redes de retroalimentación

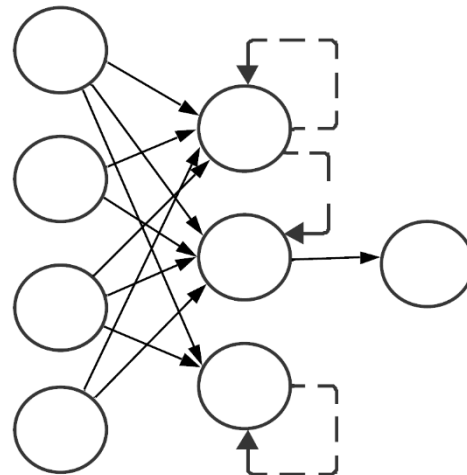


Figura 7. Clases de redes neuronales

De acuerdo con Haykin(2009) el perceptrón fue propuesto por el psicólogo Rosenblatt. El perceptrón se desarrolla a partir del modelo de neurona propuesto por McCulloch-Pitts (Figura 5).

Un perceptrón es la forma más simple de la RN utilizada para la clasificación de patrones que son linealmente separables. Este consta de una sola neurona con sinapsis ajustable, pesos y sesgos (Haykin, 2009).

2.2.7 Método de validación cruzada

Para evaluar el rendimiento y la calidad de un modelo de aprendizaje automático, es necesario utilizar un nuevo conjunto de datos, también conocido como conjunto de validación el cual incluye datos que no se utilizaron en la fase de entrenamiento (Sarkar, 2016; Žižka et al., 2019).

Cuando solo se dispone del conjunto de entrenamiento es necesario buscar un método que permita realizar la evaluación del modelo debido a que evaluar al clasificador con estos mismos datos podría provocar un sobreajuste (Duval et al., 2012; Mariñelarena-Dondena et al., 2017). El sobreajuste se refiere a que el modelo se adapta a los datos de

entrenamiento y pierde la capacidad de clasificar con precisión nuevos documentos (Mariñelarena-Dondena et al., 2017).

Existen diferentes métodos que permiten efectuar la evaluación del modelo en caso de que se disponga de un conjunto de datos, una de estas alternativas es la validación cruzada (Mariñelarena-Dondena et al., 2017; Žižka et al., 2019).

Hay varios tipos de validación cruzada, sin embargo, uno de los más comunes es la *validación cruzada de k interacciones* (Anguita et al., 2012; Berrar, 2019). Este método de validación consiste en dividir aleatoriamente un conjunto de datos en k subconjuntos del mismo tamaño (Anguita et al., 2012; Berrar, 2019); luego, un subconjunto se utiliza para evaluar al clasificador que se entrenó con los $k-1$ subconjuntos restantes. Este procedimiento se repite de forma iterativa hasta que cada uno de los k subconjuntos haya servido como conjunto de validación. Finalmente, se promedian los k valores resultantes de cada evaluación (Anguita et al., 2012; Berrar, 2019; Duval et al., 2012), y mediante distintas medidas de evaluación se determina el rendimiento del clasificador (Duval et al., 2012).

Algunas de las medidas de evaluación más utilizadas en las tareas de clasificación y minería de opiniones, son la precisión, exhaustividad y la medida F (Farzindar e Inkpen, 2015). Estas medidas se describen en la siguiente sección.

La Figura 8 ilustra este proceso para $k = 10$ iteraciones. En la primera iteración, el primer subconjunto sirve como conjunto de validación y los nueve restantes como entrenamiento. En la segunda iteración el segundo subconjunto funje como conjunto de validación y el resto como entrenamiento y así sucesivamente el proceso se repite k veces (Berrar, 2019).

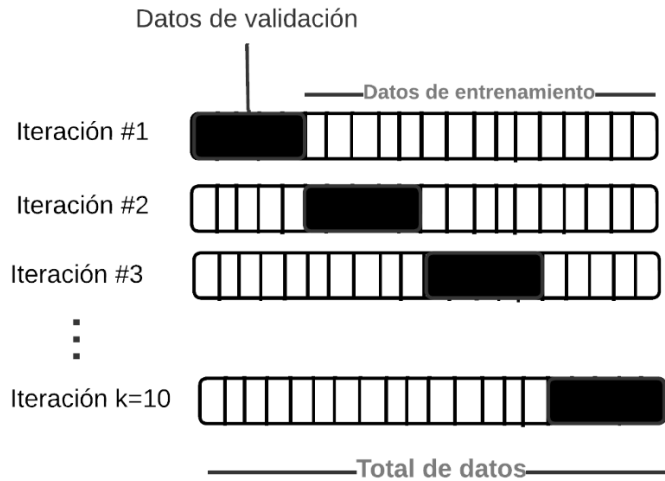


Figura 8. Proceso de validación cruzada

2.2.8 Medidas de evaluación

El rendimiento de los métodos de clasificación se comprueba usualmente por un conjunto de medidas de evaluación, tal como: la precisión, exhaustividad y medida F (Duval et al., 2012; Ramírez et al., 2015; Seger, 2018). A su vez, estas medidas se calculan respecto a los siguientes cuatro términos (Han y Kamber, 2006; Žižka et al., 2019):

- Verdadero positivo (VP): número de elementos positivos que el clasificador etiquetó correctamente.
- Verdadero negativo (VN): elementos negativos que se etiquetaron correctamente.
- Falso positivo (FP): cantidad de elementos negativos que se etiquetaron incorrectamente.
- Falso negativo (FN): elementos a los que se les asignó la etiqueta negativa, pero realmente son positivos.

2.2.8.1 Precisión

La precisión (P) mide la *calidad del modelo* que se determina como el porcentaje de muestras correctamente clasificadas, es decir, el número de elementos etiquetados correctamente dividido por el total de las predicciones clasificadas como positivas (Duval et al., 2012; Ramírez et al., 2015). La ecuación que determina el porcentaje de precisión es la siguiente:

$$\textit{Precisi3n} = \frac{VP}{VP + FP}$$

Ecuaci3n 8. C3lculo de la precisi3n

2.2.8.2 Exhaustividad

La exhaustividad (E) nos informa la *cantidad* de muestras positivas correctamente identificadas. Por lo tanto, este valor corresponde a los elementos correctamente clasificados como positivos, sobre todas las observaciones positivas (Duval et al., 2012; Ram3rez et al., 2015). La exhaustividad se expresa por la siguiente f3rmula:

$$\textit{Exhaustividad} = \frac{VP}{VP + FN}$$

Ecuaci3n 9. C3lculo de la exhaustividad

2.2.8.3 Medida F

La medida F es la medida arm3nica ponderada de la precisi3n y la exhaustividad (Ram3rez et al., 2015; ŹiŹka et al., 2019). Esta medida se calcula por la siguiente f3rmula:

$$\textit{Medida F} = 2 * \frac{\textit{Precisi3n} * \textit{Exhaustividad}}{\textit{Precisi3n} + \textit{Exhaustividad}}$$

Ecuaci3n 10. C3lculo de la medida F



Capítulo 3

Estado del arte

En este capítulo se describe de forma breve la manera tradicional en que se ha estudiado la detección de comportamientos suicidas. Asimismo, se analizan algunos trabajos desarrollados computacionalmente que han empleado técnicas de procesamiento del lenguaje natural y aprendizaje automático para identificar la ideación suicida.

3.1 Introducción

Identificar los comportamientos suicidas ha atraído la atención de investigadores en diferentes campos de estudio. Dominantemente, el suicidio se ha estudiado desde el ámbito de la salud, específicamente en el área psicológica. En este campo de estudio se han implementado diferentes instrumentos, por ejemplo: cuestionarios, pruebas psicométricas, escalas y entrevistas en terapia (Abdulsalam y Alhothali, 2022; Ortega González, 2018). Estos instrumentos principalmente evalúan y cualifican la gravedad de riesgo de comportamiento suicida (Ortega González, 2018).

De acuerdo con Ortega González (2018) y Rangel-Garzón et al. (2015) las principales escalas psicométricas para determinar el perfil psicológico de los suicidas y cualificar el riesgo suicida son: *la escala de ideación suicida de Beck, escala de desesperanza de Beck, cuestionario de intentos de suicidio, escala tentativa suicida*; por mencionar algunos ejemplos.

Las escalas e inventarios previamente citados han sido ampliamente utilizados en el campo clínico, sin embargo, estas escalas e inventarios se construyen a partir de variables clínicas, sociales y demográficas que se reportan de las personas suicidas (Ortega González, 2018; Rangel-Garzón et al., 2015), por lo tanto, no son pruebas ad hoc

a cada persona (Ortega González, 2018). Además, estos instrumentos podrían permitir que los participantes oculten o nieguen información sobre su estado de salud mental (Abdulsalam y Alhothali, 2022).

Otro recurso utilizado para estudiar el suicidio es el análisis de las notas suicidas (Ortega González, 2018). En los últimos años, debido al auge e influencia de las redes sociales e internet en los jóvenes, se han desarrollado diversos estudios para identificar automáticamente comportamientos suicidas expresados en publicaciones de redes sociales (Abdulsalam y Alhothali, 2022).

Recientemente, ha llamado la atención la extracción y el análisis de datos de redes sociales para obtener información relacionada con la salud (Jain et al., 2022). En este sentido, Troya-Capa et al. (2020) mencionan que la inteligencia artificial presenta un potencial elevado para diagnosticar y comprender enfermedades mentales como el suicidio.

Desde el punto de vista computacional se han desarrollado modelos para detectar textos con contenido de ideación suicida en redes sociales, como Twitter⁵ y Reddit (Abdulsalam y Alhothali, 2022; Castillo-Sánchez et al., 2020).

Debido a la creciente predisposición de los jóvenes por compartir pensamientos suicidas a través de las redes sociales, desde el punto de vista computacional se han desarrollado modelos para identificar textos con contenido suicida (Abdulsalam y Alhothali, 2022; Castillo-Sánchez et al., 2020). Principalmente, se han empleado técnicas de aprendizaje automático y procesamiento del lenguaje natural. Estas técnicas permiten a los investigadores examinar de forma rápida y automática el contenido de millones de publicaciones e identificar las tendencias y patrones textuales de dichos escritos (Parrott et al., 2019).

Abdulsalam y Alhothali (2022) en su estudio analizan 24 investigaciones que emplean algoritmos de aprendizaje automático para detectar ideación suicida. De este análisis, los autores concluyen que las redes sociales más utilizadas para identificar textos con contenido suicida son: *Twitter*, *Reddit*, *Tumblr*, *Weibo* y *Vkontakte*. *Twitter* es la plataforma más popular (Figura 9), seguida por *Reddit*. Sin embargo, Yao (2020) menciona que el límite de caracteres de los tuits podría restringir la expresión de las

⁵ Servicio gratuito de difusión social que permite publicar mensajes cortos llamados tuits, los cuales están limitados a 280 caracteres (Abdulsalam y Alhothali, 2022; Akhtar et al., 2020)

personas. Estos autores también indican que mayormente se han analizado publicaciones con contenido suicida en idioma inglés, seguido del chino (Figura 10).

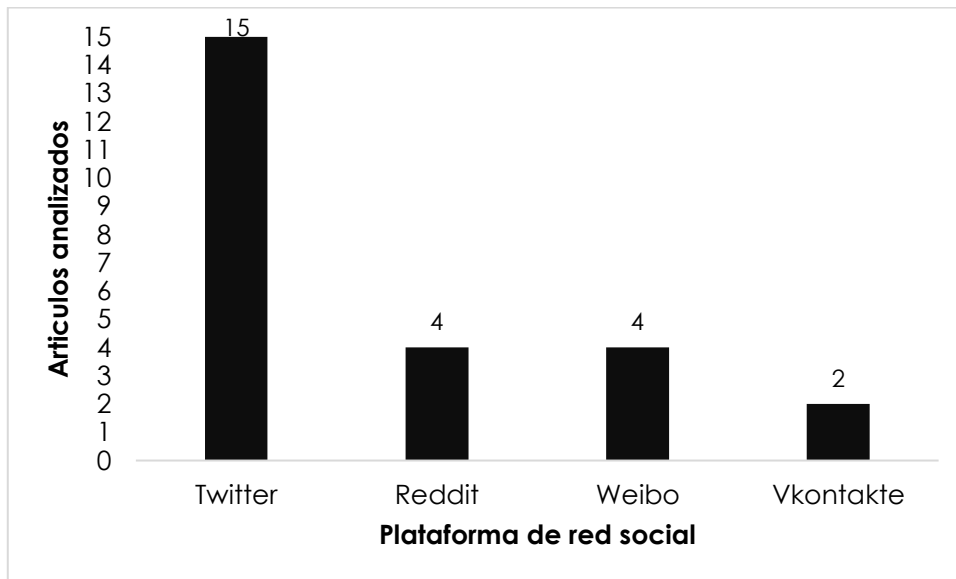


Figura 9. Redes sociales estudiadas para la detección de ideación suicida

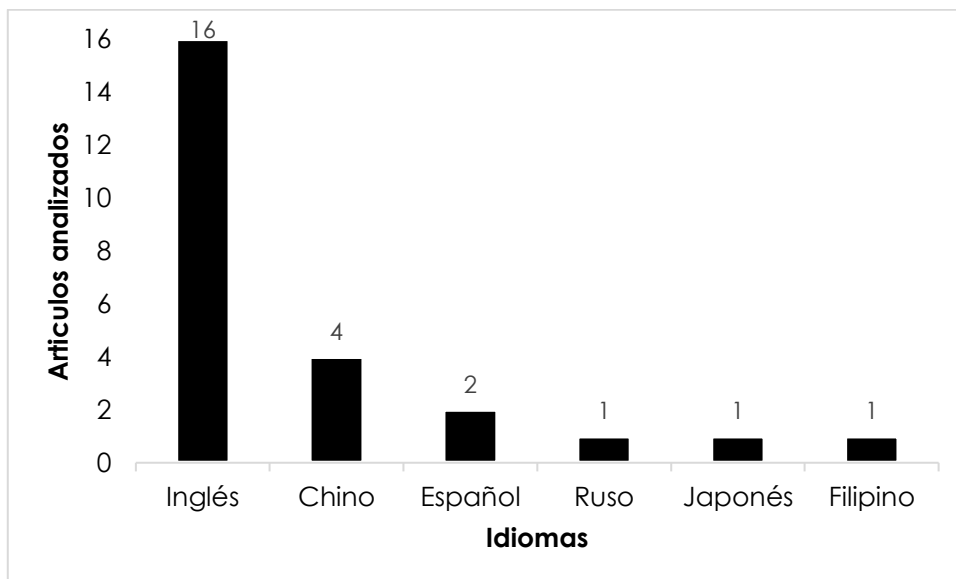


Figura 10. Idiomas estudiados para la detección de ideación suicida

Los conjuntos de datos que se han utilizado hasta el momento para identificar contenido suicida varían en tamaño; estos contienen entre 102 y 1,100,000 documentos (Castillo-Sánchez et al., 2020).

3.2 Análisis de depresión y suicidio mediante aprendizaje automático y PLN

En la investigación de Jain et al. (2022) se aborda la detección de texto depresivo y texto suicida mediante técnicas de procesamiento del lenguaje natural y aprendizaje automático.

El objetivo de la investigación fue clasificar textos referentes a depresión clínica, de aquellos relacionados con el suicidio. Se estableció este objetivo porque la depresión es una enfermedad mental que puede conducir a la ideación e intentos de suicidio.

Los autores extrajeron 600,000 textos de la red social Reddit para crear su conjunto de datos. De esta manera, cada clase contiene 300,000 textos. Los textos referentes a depresión se obtuvieron del subreddit *r/depresión*; mientras que los textos enfocados al suicidio se extrajeron del subreddit *r/SuicideWatch*. Eligieron esta red social porque es una red activa que cuenta con poco spam y son muy pocas las publicaciones que muestran imágenes. Además, los subreddits seleccionados parecen ser lugares confiables en línea donde las personas dan señales significativas y honestas de su estado mental.

La metodología establecida en este trabajo se basa en el modelo general de clasificación (clasificación directa). Los autores definen los siguientes pasos en su propuesta (Figura 11).

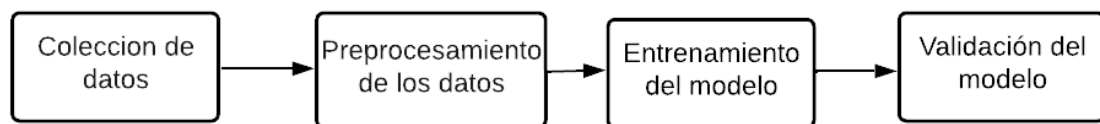


Figura 11. Metodología de Jain et al. (2022)

En la etapa de preprocesamiento, los autores convirtieron los textos en minúsculas, tokenizaron y removieron las palabras vacías.

En la etapa de entrenamiento aplicaron diferentes algoritmos de aprendizaje automático, tal como: RL, NB, MVS y BA.

La Tabla 1 muestra los resultados obtenidos por cada clasificador. La RL obtuvo el mejor desempeño (0.79), seguido por BA, MVS y NB.

Tabla 1. Resultados de la clasificación entre notas de depresión e ideación suicida

Clasificador		P (%)	R (%)	F(%)
RL	1	0.80	0.79	0.79
	2	0.79	0.80	0.79
NB	1	0.78	0.72	0.75
	2	0.74	0.80	0.77
MVS	1	0.77	0.78	0.77
	2	0.77	0.76	0.77
BA	1	0.78	0.77	0.77
	2	0.77	0.77	0.77

Los autores concluyen que Reddit podría utilizarse como complemento del sistema tradicional de salud pública, debido a que este medio permite el intercambio de ideas, emociones y uso de términos médicos.

Como trabajo futuro, los autores plantean el uso de redes neuronales artificiales o redes neuronales recurrentes para mejorar la detección de notas suicidas. Además, proponen analizar los comentarios, las reacciones y el número de veces que se comparte una nota.

3.3 Caracterización de la ideación suicida mediante el uso de características de trastornos mentales en microblogs: Una perspectiva de aprendizaje automático

El objetivo de la investigación de Sarsam et al. (2022) fue identificar los principales temas y emociones que caracterizan a los textos de ideación suicida compartidos a través de la red social *Twitter*.

Para lograr tal objetivo, los autores definieron distintas etapas: primero crearon su conjunto de datos y los preprocesaron. Posteriormente, agruparon los datos para descubrir los temas latentes en cada documento. A continuación extrajeron los sentimientos y las emociones latentes. Finalmente, clasificaron documentos no suicidas y suicidas.

En esta investigación, los autores generaron un conjunto de datos de 54,385 tuits en idioma inglés. Los textos fueron recopilados mediante la interfaz de programación de aplicaciones (API) de transmisión gratuita de Twitter. Los textos seleccionados debían

poseer alguna palabra clave relacionada con el suicidio, tal como: *suicidio*, *pensamientos suicidas* y *contemplando el suicidio*.

Además, los autores consideraron la participación de dos humanos expertos en salud mental para evaluar la relevancia de las publicaciones recopiladas.

En la etapa de preprocesamiento los autores implementaron el modelo de bolsa de palabras. Para lograr este objetivo, los autores tokenizaron, convirtieron a minúsculas, eliminaron las palabras vacías y estandarizaron la longitud de los tuits. Con las palabras resultantes construyeron un diccionario. Posteriormente, con un algoritmo de agrupación crearon tres grupos.

Con el propósito de descubrir los temas latentes de cada grupo emplearon el algoritmo de ALD. Posteriormente, se les pidió a los expertos en salud mental que evaluaran el contenido de cada tema y proporcionaran la etiqueta adecuada a cada grupo. Como resultado, los expertos indicaron que el grupo uno y el grupo dos eran de contenido *no suicida*, mientras que el grupo tres se categorizó como *suicida*.

En la Figura 12 se muestran los grupos generados. Cada círculo representa un tema específico, el tamaño demuestra la frecuencia del tema y la distancia entre círculos refleja la similitud de los temas.

El primer grupo refiere a documentos sobre consejos para la prevención del suicidio. Entre los temas que abarca se encontró la propuesta de intervenciones psicosociales, la terapia conductual y opciones de tratamiento.

El segundo grupo corresponde a los documentos centrados en presentar información estadística sobre el suicidio. En este grupo se abarcan temas como: *el número de personas que mueren anualmente*, *los métodos de suicidio más utilizados* o *la tasa de suicidios por género*.

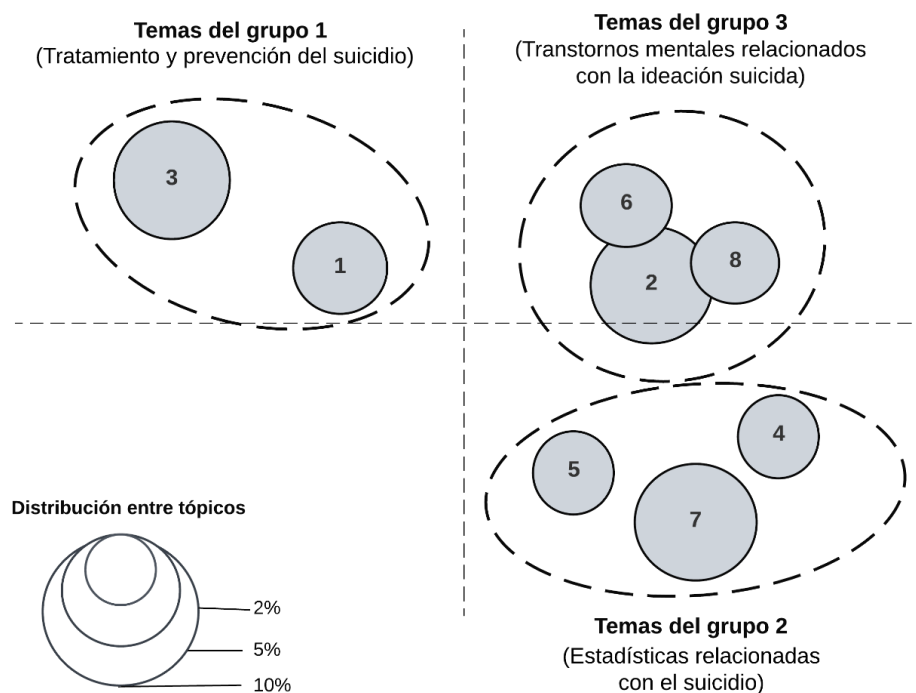


Figura 12. Temas latentes en tuits

Los documentos altamente asociados con trastornos mentales, por ejemplo: *depresión*, *alcoholismo* y *esquizofrenia*; corresponden al tercer grupo. En cuanto a los temas relacionados, se encontró que algunas personas con depresión que cometieron suicidio tenían problemas personales, tal como la pérdida de trabajo. Por otro lado, se indica que a menudo las personas con pensamientos suicidas recurren al alcohol para sobrellevar algún trauma. Finalmente, en relación con la esquizofrenia, se mostró que las personas con este padecimiento tienen más posibilidades de suicidarse.

Con el propósito de identificar la polaridad de sentimientos asociados a cada grupo, los autores implementaron *SentiStrength*. Este recurso asigna a cada grupo una de las siguientes categorías: *positiva*, *negativa* o *neutral* con base en el vocabulario que contenga. Los autores descubrieron que en el grupo uno predominaba la polaridad positiva. Por el contrario, en el grupo tres prevalecía la polaridad negativa y en el grupo dos la polaridad neutral.

Posteriormente, para comprender las emociones suicidas y no suicidas, utilizaron el lexicón de intensidad afectiva NRC. Este lexicón realiza un análisis afectivo categorizando los documentos en emociones como: *ira*, *miedo*, *tristeza* y *confianza*. De esta manera encontraron un alto nivel de confianza y emociones positivas en los documentos del grupo uno. Por el contrario, en el grupo tres se indica mayor presencia

de emociones negativas como el miedo y la tristeza. La gráfica de este análisis se observa en la Figura 13.

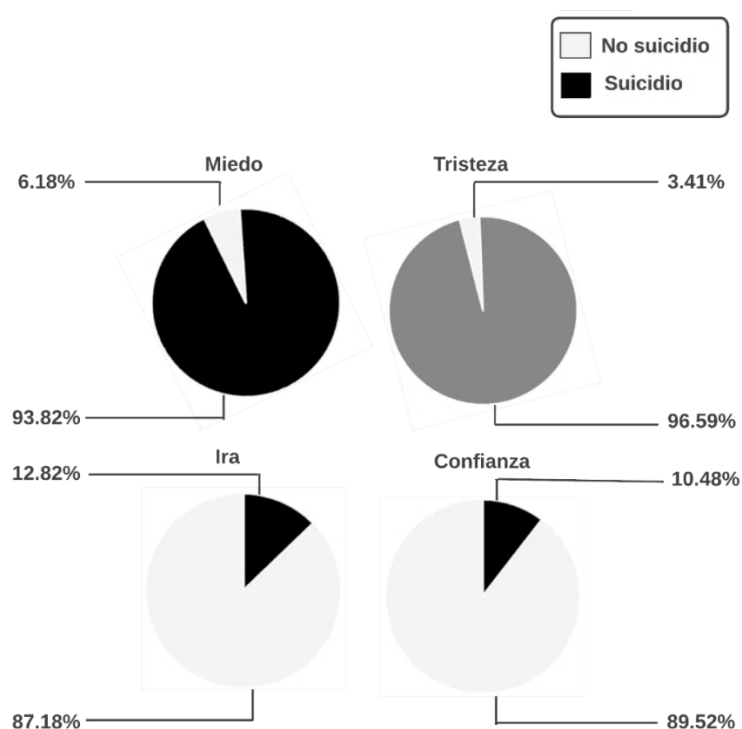


Figura 13. Emociones extraídas

Finalmente, clasificaron entre documentos no suicidas y suicidas (correspondientes a los del grupo uno y tres respectivamente). Esta clasificación la realizaron con los siguientes algoritmos: k-vecinos más cercanos, regresión logística multinomial, árbol de decisión y el método de embolsado. Además, evaluaron las predicciones con validación cruzada de diez interacciones.

En la Tabla 2 se compara la exactitud, estadística Kappa y el error cuadrático medio, obtenidos por cada clasificador. Estos resultados revelan que la mejor exactitud y estadística Kappa se alcanzó con el método de embolsado, seguido de regresión logística multinomial.

Tabla 2. Resultados de la clasificación realizada por Sarsam et al. (2022)

Algoritmo	Exactitud (%)	Estadística Kappa (%)	Error cuadrático medio (%)
K-vecinos más cercanos	64.18	68	35
Regresión logística multinomial	49.43	55	91
Embolsado	97.64	96	3

Los autores concluyen que las emociones tienen un gran impacto en la clasificación del suicidio. Además, mencionan que, en comparación con otros algoritmos de aprendizaje automático, el método de embolsado (*modelo de ensamble*) tiene gran potencial para la detección de declaraciones suicidas. Otro hallazgo es que la ALD permite encontrar los temas latentes relacionados con el suicidio, siendo útil para comprender dicho tema.

3.4 Detección de ideación suicida en foros de redes sociales mediante aprendizaje profundo

En la investigación de Tadesse et al. (2019) se aborda la detección de ideación suicida a través de clasificadores de aprendizaje profundo y aprendizaje automático. Específicamente se centraron en el análisis de publicaciones en idioma inglés, extraídos de la red social *Reddit*.

En este estudio se analizó el conjunto de datos desarrollado por Ji et al. (2018). Este conjunto consta de 3,549 publicaciones indicativas de suicidio y 3,652 publicaciones no suicidas, dedicadas a apoyar a personas potencialmente en riesgo. La Tabla 3 muestra algunos documentos de cada categoría.

Tabla 3. Ejemplos de publicaciones suicidas y no suicidas

Publicaciones suicidas	Publicaciones no suicidas
Want to die, end it now, I wanna die with a blunt. I'm going to kill myself this weekend. I want to slit my wrists tonight, I tried to commit suicide.	There is no one "correct" way to talk to someone struggling with suicidal thoughts, Children of suicide parents. I think you should tell people how you feel, I think suicide is a permanent option that most of the time results out of a temporary issue.
I wish guns for suicide, nobody cares if I die. I want to die Where can i go to commit Suicide??	Will you ever get over the news that one of your parents committed suicide?
Where can i go to commit suicide?? I don't know what else to do.	Friend has given up, seriously considering suicide.

Los autores argumentan que los métodos de aprendizaje profundo han logrado avances impresionantes en el campo de la visión por computadora y el reconocimiento

de patrones. Además, mencionan que la combinación de clasificadores puede mejorar el rendimiento de la clasificación de texto. Por lo anterior, proponen un nuevo enfoque para mejorar la clasificación de ideas suicidas. Su propuesta se basa en combinar los clasificadores de aprendizaje profundo: Memoria a Largo y Corto Plazo [MLCP], y Red Neuronal Convolutiva (RNC).

Con el objetivo de comparar el enfoque propuesto con clasificadores de aprendizaje automático, los autores establecieron dos direcciones (Figura 14). La primera consiste en métodos tradicionales de aprendizaje automático y técnicas de extracción de características base del procesamiento de lenguaje natural.

La segunda dirección consiste en una técnica de incrustación de palabras y clasificadores de aprendizaje profundo.

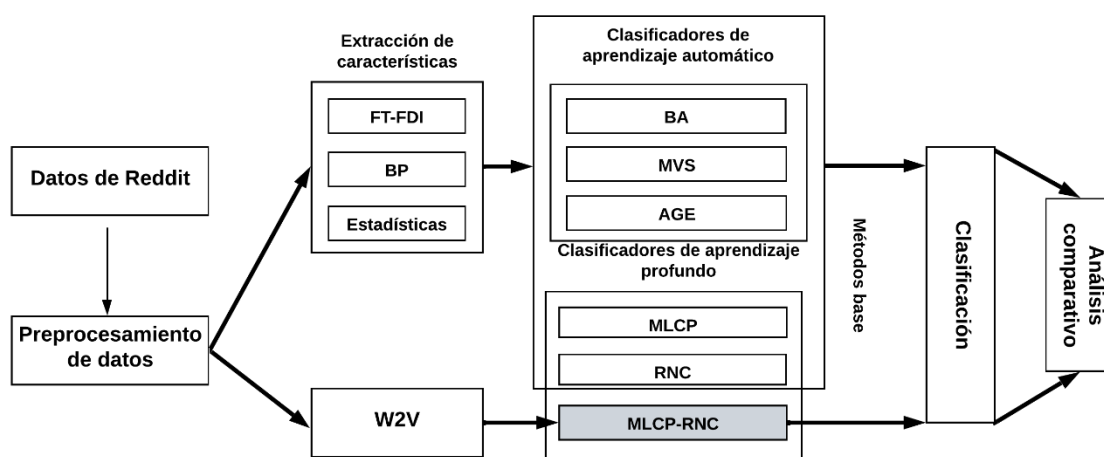


Figura 14. Direcciones para la detección de ideación suicida

En la etapa de preprocesamiento, los autores concatenaron los títulos con los cuerpos de las publicaciones, eliminaron las oraciones duplicadas, tokenizaron, eliminaron caracteres especiales, direcciones de sitios webs y contracciones. Por último, convirtieron todos los textos a minúsculas.

En la primera dirección enfocada al uso de técnicas de procesamiento de lenguaje natural con clasificadores tradicionales de aprendizaje automático, los autores extrajeron las características textuales mediante técnicas como: FT- FDI y Bolsa de Palabras (BP). Asimismo, extrajeron las características estadísticas, por ejemplo: el número de tokens, oraciones y su longitud.

Posteriormente, las características obtenidas previamente ingresaron de manera individual y concatenadas a los siguientes clasificadores: MVS, NB, BA y Aumento de

Gradiente Extremo (AGE). El objetivo de concatenar las características fue examinar qué combinación favorece la detección de ideación suicida.

En la segunda dirección se obtuvieron las características textuales mediante el modelo de W2V con longitud de 300 dimensiones. A continuación, las características se ocuparon como entrada para los clasificadores individuales de aprendizaje profundo: MLCP, RNC. Estos atributos también se utilizaron como entrada al enfoque que combina ambos clasificadores.

Los clasificadores individuales de MLCP y la RNC utilizaron la función de activación ReLU.

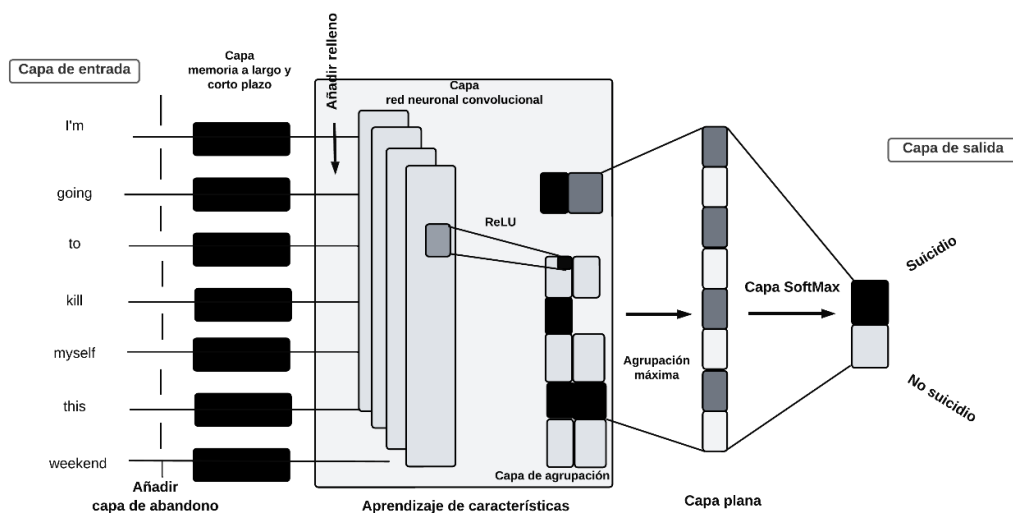


Figura 15. Enfoque combinado de aprendizaje profundo

La Figura 15 muestra la arquitectura del enfoque combinado propuesto por los autores de este estudio. Este enfoque consta de capas diferentes. La primera capa es de incrustación de palabras en la que se utilizó el modelo de W2V para transformar las palabras a una representación vectorial de longitud fija. A continuación, se aplicó una capa de abandono para evitar el sobreajuste, como parámetro definieron una tasa de 0.5. Posteriormente, se empleó la MLCP para capturar la información durante largo tiempo. Seguidamente, se tomó la salida de la MLCP, como vector de entrada de la RNC para realizar la extracción de características. Después, en la capa de agrupación se retuvo la información más importante reduciendo la dimensión del mapa de características. Subsecuentemente, el mapa de características se transformó a un vector mediante una capa plana. Finalmente, la función SoftMax se aplicó para clasificar los documentos.

En esta Tabla 4 se compara la exactitud (A), medida F (F), exhaustividad (E) y precisión (P) en la dirección base y el enfoque combinado.

Tabla 4. Resultados de la clasificación realizada por Tadesse et al. (2019)

Método	Extracción de características	A (%)	F (%)	E (%)	P (%)
RF	Estadísticas	77.2	75.1	73.9	76.3
	FT- FDI	81.8	80.9	83.4	80.5
	BP	81.1	78.6	77.9	81.
	Estadísticas + FT- FDI + BP	85.6	84.1	84	85
MVS	Estadísticas	79.6	79	70	60
	FT- FDI	81.2	82.7	87.2	78.7
	BP	80.6	81.1	81.8	80.4
	Estadísticas + FT- FDI + BP	83.5	83.8	85.5	82.1
NB	Estadísticas	68.2	71.3	76.3	67.6
	FT- FDI	78.6	76.1	75.6	80.5
	BP	79.8	78.4	78.9	79.7
	Estadísticas + FT- FDI + BP	82.5	81.5	83.4	80.8
AGE	Estadísticas	76.3	76.1	75.6	80.5
	FT- FDI	85.6	84.1	84.0	85.8
	BP	83.1	82.6	84.4	81.6
	Estadísticas + FT- FDI + BP	88.3	83.1	84.3	88.4
MLCP		91.7	92.6	90.5	94.8
RNC	W2V	90.6	92.8	93.8	91.8
MLCP + RNC		93.8	93.4	94.1	93.2

Estos resultados revelan que en la línea base el clasificador de MLCP tiene el mejor desempeño con una precisión de 91.7% y una medida F de 92.6%. Seguido del clasificador de AGE que considera las características individuales y combinadas.

Al comparar la línea base completa con el enfoque propuesto, los autores concluyen que el mejor desempeño se logra con el enfoque combinado de MLCP y la RNC. El enfoque combinado alcanza una precisión del 93.8% y una medida F del 92.8%. Estos resultados revelan que el ensamble de clasificadores supera a los demás clasificadores.

Adicionalmente, los autores de este trabajo compararon el vocabulario de ambas clases. Para ello, calcularon las frecuencias de los 200 unigramas y bigramas principales

de cada categoría. A partir de este análisis, los autores afirman que las personas suicidas manifiestan diversas emociones. Por ejemplo, expresan desesperanza, frustración, ansiedad, culpa, arrepentimiento y soledad mediante frases como: "*jodida vida*", "*vida cansada*", "*odio cansado*", "*estoy cansado*", "*tengo miedo*", "*lo siento*" y "*nunca más*".

Por el contrario, en las publicaciones ajenas al suicidio predominan palabras que refieren a momentos felices, actitudes y sentimientos positivos, tal como: "*quiero bromear*", "*quiero divertirme*", "*reír*" y "*quiero ser feliz*".

Otra observación a partir del análisis del vocabulario es que los usuarios con tendencia al suicidio principalmente utilizan frases con connotación de muerte, por ejemplo: "*suicidio*", "*quiero morir*", "*morir jodidamente*", "*deseo de suicidio*", "*deseo morir*", "*quiero matar*", "*quiero terminar*" y "*quiero ir*".



Capítulo 4

Método propuesto

En este capítulo se describe el método propuesto en esta investigación para detectar notas de ideación suicida mediante un análisis emocional.

El enfoque propuesto consiste en un ensamble de clasificadores con una arquitectura en cascada (Figura 16).

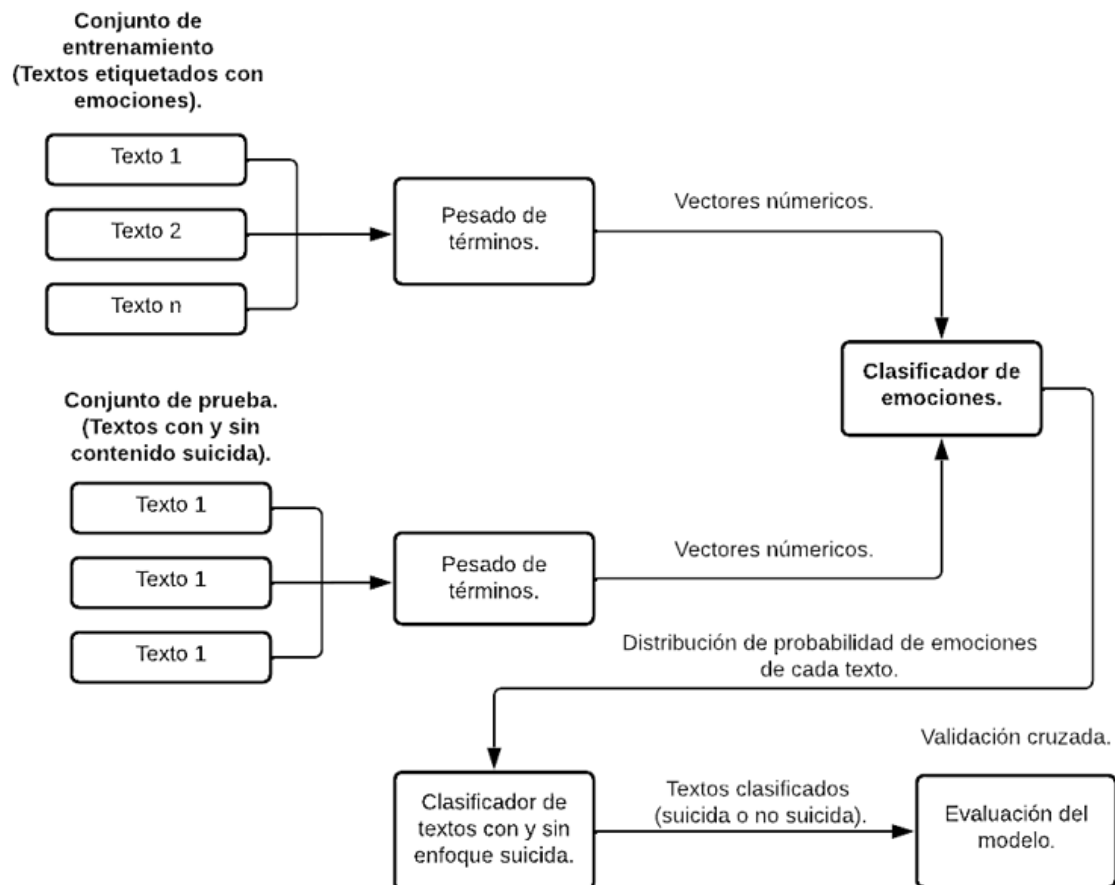


Figura 16. Esquema general del método propuesto

De acuerdo con el esquema propuesto en la Figura 16, el primer paso consiste en la colección de datos, los cuales son preprocesados para eliminar el mayor ruido lingüístico. Posteriormente, se aplican diferentes métodos de generación de características para convertir cada texto en un vector numérico de longitud fija. Luego, la distribución emocional de cada texto se determina con el primer clasificador. A continuación, la distribución emocional alimenta al segundo clasificador que determina qué texto es suicida en función de esa entrada. Para validar los resultados de la clasificación se realizó un procedimiento de validación cruzada igual a 5 iteraciones.

Este capítulo está organizado de la siguiente manera: en la sección 4.1 se describen los conjuntos de datos utilizados. A continuación, en la sección 4.3 se indican los métodos de generación de características empleados en esta investigación, y finalmente en la sección 4.4 se explica el proceso de clasificación desarrollado.

4.1 Conjuntos de datos

Por lo general, una nota de suicidio se escribe momentos antes de que se realice un intento. La nota de suicidio proporciona información sobre el estado emocional, las motivaciones, los pensamientos y deseos; de la persona suicida vinculados al momento en que lleva a cabo su muerte (Ghosh et al., 2020).

En gran medida se han analizado notas suicidas en idioma inglés; no obstante, cabe resaltar que de acuerdo con el conocimiento de los autores no hay disponible públicamente un corpus sólido de este tipo de textos. Por tal motivo, en este estudio se crearon tres conjuntos de datos diferentes en idioma inglés que cubren los siguientes tópicos: notas de suicidio consumado, intenciones suicidas y no suicidas. Estos conjuntos de datos se analizaron para comparar el desempeño del método de clasificación en cascada con respecto a otros enfoques.

El método propuesto está planteado para analizar notas en idioma inglés; sin embargo, es posible adecuarlo para procesar textos en otros idiomas donde cada palabra se delimite por un espacio en blanco. Esta característica es importante porque permite identificar los elementos que componen a los textos y estudiar sus relaciones léxicas y semánticas. El análisis de idiomas no segmentados es un desafío para el procesamiento del lenguaje natural, por lo que se requieren enfoques más robustos que permitan tokenizar dichos textos (Hernández-Castañeda., 2017).

Para obtener la distribución emocional del enfoque propuesto, se utilizaron los siguientes recursos. Por un lado, un conjunto de textos etiquetado con emociones. Por el otro lado, un conjunto de notas suicidas y notas con dominio diferente al suicidio.

Los conjuntos de datos considerados en esta investigación se detallan a continuación.

4.1.1 Notas de suicidio consumado

Esta colección consta de 155 notas escritas por personas que consumaron el intento y que expresaron textualmente su ideación suicida. Es importante destacar que este tipo de notas se caracterizan por expresar intenciones suicidas de forma implícita, tal como se observa en el siguiente ejemplo.

Ejemplo de nota de suicidio consumado

The pain you have caused me everyday has destroyed every bit of me, destroyed my soul. I can't eat or sleep or think or function. I am running away from everything. The career is not even worth it anymore. When I first met you, I was driven, ambitious and disciplined. Then I fell for you, a love I thought would bring out the best in me. I don't know why destiny brought us together. After all the pain, the rape, the abuse, the torture...

El ejemplo anterior muestra que no en todos los textos de este conjunto se plasman palabras asociadas al suicidio que indiquen explícitamente el deseo de morir; sin embargo, suelen expresar emociones tal como: dolor, ira, miedo; entre otras.

Las notas de este conjunto se recolectaron de diversos sitios web y periódicos en línea.

4.1.2 Notas de intenciones suicidas

Esta colección consta de 453 notas de ideación suicida recolectadas manualmente de la red social Reddit, específicamente de los subreddits “r/SuicideWatch” y “r/Depression”. En estos subreddits las personas pueden compartir públicamente sus pensamientos suicidas o depresivos.

Las notas de este conjunto se recopilaron con base en la presencia de algunos términos y/o frases relacionadas con el suicidio, tal como: *suicidio, ideación suicida, intento de suicidio, muerte e intento fallido*. Estos términos se tomaron de la lista propuesta por Parrott et al. (2019).

A continuación, se muestra el ejemplo de una nota de ideación suicida extraída del subreddit SuicideWatch.

Ejemplo de nota de intención suicida

I want to suicide. I am going to try again tonight but I have a feeling it's going to fail again but I hope it goes well.

4.1.3 Notas no relacionadas al suicidio

Esta colección consta de 255 notas que abarcan los siguientes seis dominios: cultura, economía, política, música, comida y deportes. Las notas se recolectaron de la red social Reddit.

El objetivo de crear este conjunto fue tener contraejemplos de lo que no es una ideación suicida y darle al clasificador el conocimiento de lo que no debe considerar como una nota de suicidio.

A continuación, se muestra el ejemplo de una nota referente al tema de comida.

Ejemplo de nota no suicida

I like drinking milk and cereal separately; I know most people would disagree and have it mixed. The reason why is because I feel like cereal is more like cereal when it's eaten by itself and it's more crunchy when you don't pour milk in. I know you have to eat the cereal quickly to prevent it from getting too soggy...

4.1.4 Notas de sentenciados

Este conjunto consta de 453 notas escritas por presos sentenciados a muerte. Las notas se recolectaron del sitio oficial del Departamento de Justicia de Texas⁶.

El siguiente fragmento es un ejemplo de una nota escrita por un sentenciado.

First of all, I want to apologize to the family. I am sorry for what I did to her twelve years ago. I wish they could forgive me for what I did. I am sorry. I am sorry for hurting my family, for hurting my friends...

⁶ https://www.tdcj.texas.gov/death_row/dr_executed_offenders.html

Esta colección se consideró con el objetivo de comparar y analizar las emociones entre las personas que se quitan la vida por voluntad propia y aquellas que no tienen el deseo de morir, pero están obligadas a hacerlo.

4.1.5 Conjunto de emociones

Con el objetivo de obtener la distribución emocional que presenta cada nota de estudio, se utilizó el conjunto de datos CEASE desarrollado por Ghosh et al. (2020). Esta colección consta de 2,393 oraciones extraídas de aproximadamente 205 notas suicidas.

El CEASE consta de las siguientes quince emociones: abuso, ira, culpa*, miedo, perdón, culpa**, felicidad_tranquilidad, esperanza, desesperanza, amor, orgullo, pena, agradecimiento, información e instrucciones; cada oración está asociada a una emoción específica, tal como se muestra en la Tabla 5.

Tabla 5. Ejemplos de oraciones del conjunto CEASE

Notas	Etiqueta
after all, i was stupid to be used because i worked hard to adjust myself to the atmosphere in the presidents room under the direction of the director.	abuso
i feel disgusted when i see guys who make a fuss about the entrance examination all the time.	ira
there is no one that admires jesus christ more than i do although the conventional contemporary method of worshipping him rives me a pain in the neck.	culpa*
i am afraid i am a coward.	miedo
dear god please have mercy on my soul please forgive me i cannot stand the pain anymore.	perdón
cannot go on spoiling your life any longer.	culpa**
i am really happy now.	felicidad_tranquilidad
i hope you overcome the unsecurity and i hope you realise that you are ever so beautiful.	esperanza
my mood is one of profound discouragement and my personal future appears bleak.	desesperanza
i love you so completely whole heartedly without restraint.	amor
i am proud of how i lived.	orgullo
i have been so down so god damn down i cannot get up.	pena
thank you for being a positive influence on my life and making it a little better.	agradecimiento
every one who knows me and hears of it will have a different hypothesis to offer to explain why i did it.	Información
you have to lie on your job applications.	instrucciones

En el presente estudio se excluyeron las sentencias etiquetadas como *información e instrucciones*, por no considerarse como emociones. Por tal motivo, el conjunto de estudio consta de 1,001 oraciones.

En la Tabla 6 se muestra la distribución de las notas según su polaridad de sentimiento y la emoción que presentan.

Tabla 6. Distribución de las notas del conjunto CEASE

Sentimiento positivo		Sentimiento negativo	
Emoción	Número de notas	Emoción	Número de notas
perdón	24	abuso	12
orgullo	16	miedo	29
felicidad_tranquilidad	38	pena	305
esperanza	169	ira	79
agradecimiento	38	culpa*	47
amor	85	culpa**	74
		desesperanza	85
	<i>Total</i>	<i>Total</i>	
	370	631	

*sustantivo.

**verbo.

4.2 Estadísticas básicas de los conjuntos de datos

En la Tabla 7 se muestran el número de tokens y tipos presentes en cada colección. De acuerdo con Hernández-Castañeda (2017) los *tokens* se refieren a la cantidad total de palabras contenidas en los documentos, mientras que los *tipos* son el número total de palabras no repetidas en los documentos.

Tabla 7. Descripción de los conjuntos de datos utilizados

Conjunto de datos	Número de textos	Tokens	Tipos	Promedio de tokens por doc.
Suicidio consumado	155	1,684	852	11
Intensión suicida	453	28,093	4,322	62
No relacionado al suicidio	255	13,961	4,606	55
Sentenciados	453	21,433	3,292	47
Emociones	1,001	5,426	1,839	5

4.3 Generación de características

Mariñelarena-Dondena et al. (2017), Sarkar (2016) y Žižka et al. (2019) mencionan que la extracción de características es el proceso de convertir los tokens obtenidos del preprocesamiento a una representación numérica vectorial, donde la longitud del vector representa el número total de características únicas para todos los documentos y los valores son la importancia de las características en el documento de estudio (Sección 2.2.5).

En este estudio se usaron cuatro fuentes diferentes de generación de características (MTD, FT- FDI, ALD, D2V), con el objetivo de comparar qué tipo de información lingüística es más útil para identificar notas suicidas.

Los métodos de generación de características se eligieron porque han mostrado resultados competentes en la investigación de interés, por otro lado, permiten cubrir diferentes niveles del lenguaje (léxico y semántico); por ejemplo, MTD evalúa si la palabra aparece en el texto de estudio. FT- FDI proporciona características que reflejan la importancia de las palabras en un conjunto de textos. ALD provee características semánticas y D2V considera el contexto de las palabras para proporcionar información semántica.

4.3.1 Codificación de un solo paso

La codificación MTD consiste en convertir los datos categóricos (palabras) a valores binarios, donde si la palabra existe en la nota de estudio, se establece su presencia con el valor uno, en caso contrario, con el valor cero (Hernández-Castañeda et al., 2020).

En la Figura 17, se ejemplifica la representación de los vectores de características obtenidos con MTD.

Vocabulario del conjunto de datos				
palabra 1	palabra 2	palabra 3	palabra n	palabra n +1
want	suicide	going	feeling	commit

Vector de características del texto 1

Texto 1: "I want to suicide. I am going to try again tonight but I have a feeling it's going to fail again but I hope it goes well"

característica 1	característica 2	característica 3	característica n	característica n+1
1	1	1	1	0

Vector de características del texto 2

Texto 2: "I know you can overdose on drugs and that you can get alcohol poisoning and stuff, but can you purposely commit suicide by drinking too much alcohol."

característica 1	característica 2	característica 3	característica n	característica n+1
0	1	0	0	1

Figura 17. Ejemplos de construcción de vectores con MTD

4.3.2 Frecuencia de término-frecuencia de documento inversa

FT-FDI es un esquema de ponderación de términos que refleja la importancia de cada palabra en un documento y en un conjunto de documentos (Ledeneva y García Hernández, 2017).

La frecuencia resultante entre ambas variables (frecuencia del término y frecuencia inversa del documento) se presenta como una matriz. La representación matricial consta de filas que representan el número total de documentos en la colección de datos, mientras que las etiquetas de las columnas representan el número de términos distintos en todos los documentos; cada columna contiene los valores ponderados de FT-FID respecto a cada término (Blei et al., 2003; Sarkar, 2016).

Como ejemplo se muestran los siguientes documentos con contenido de ideación suicida.

- I want to suicide. I am going to try again tonight but I have a feeling it's going to fail again but I hope it goes well.
- I know you can overdose on drugs and that you can get alcohol poisoning and stuff, but can you purposely commit suicide by drinking too much alcohol.
- I need a way to commit suicide. Please just help me out by doing this, i don't want any other things to be said, just a way to suicide myself.

Al calcular las puntuaciones FT- FDI (Sección 2.2.5.2) y eliminar las palabras vacías, se genera la matriz de la Tabla 8.

Tabla 8. Ejemplo de matriz resultante con FT-FDI

	alcohol	going	overdose	please	poisoning	stuff	suicide	tonight	try	want
0	0.0	0.726	0.0	0.0	0.0	0.0	0.276	0.363	0.363	0.363
1	0.5	0.0	0.5	0.0	0.5	0.5	0.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.795	0.0	0.0	0.605	0.0	0.0	0.0

4.3.3 Asignación latente de Dirichlet

ALD (Sección 2.2.5.3) es un modelo probabilístico que representa los documentos como una mezcla aleatoria de k diversos tópicos, donde k es un número definido por el usuario (Hernández-Castañeda, 2017; Sarsam et al., 2022). Cada tópico consta de un conjunto de palabras relacionadas semánticamente (Hernández-Castañeda, 2017; Mariñelarena-Dondena, et al., 2017). A su vez, las palabras son seleccionadas de acuerdo con determinada probabilidad (Hernández-Castañeda, 2017; Sawhney et al., 2018).

La Tabla 9 muestra las palabras que conforman algunos tópicos encontrados en las notas de ideación suicida.

Tabla 9. Generación de tópicos en notas de ideación suicida

Tópico 1	Tópico 2	Tópico 3	Tópico 4	Tópico 5
love	sorry	forgive	father	suicide
life	please	ready	goodbye	kill
like	hope	wish	happy	feel
people	pain	peace	family	pain
thank	yall	live	reason	hate
suicide	home	tired	world	dead
family	stupid	hate	accident	hurt

En la Figura 18 se observa gráficamente el grado de pertenencia de cada tópico en una nota de ideación suicida. El eje *Tópico* representa los diez tópicos más frecuentes de la nota y el eje *Probabilidad* indica la probabilidad de pertenencia de cada tópico en el documento analizado.

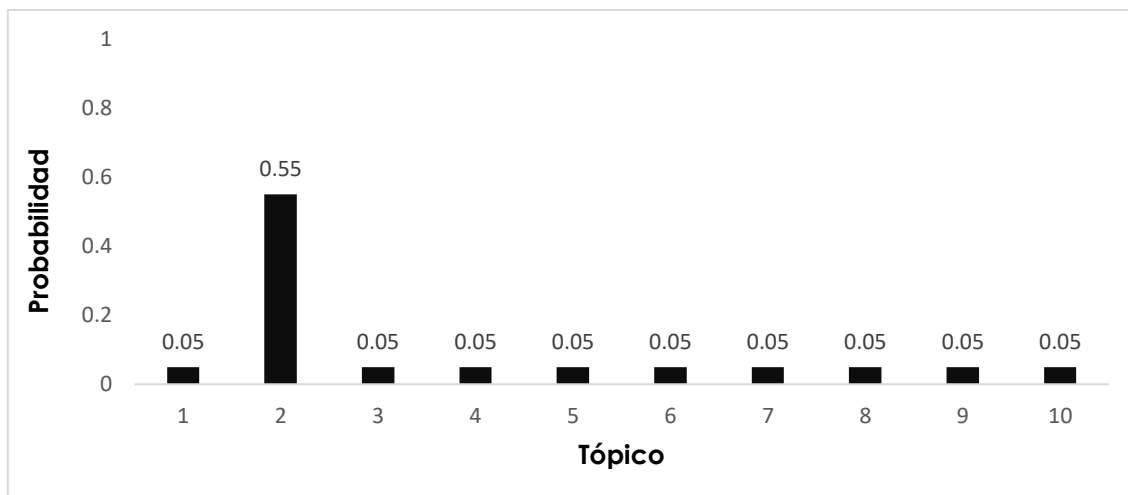


Figura 18. Distribución de tópicos en una nota de ideación suicida

4.4 Clasificación en cascada

El método propuesto en esta investigación para la detección de notas suicidas consiste en dos clasificadores de aprendizaje supervisado, establecidos en una arquitectura en cascada. El primer clasificador obtiene la distribución de emociones latentes en cada texto (Sección 4.4.1) y el segundo, toma la distribución emocional para identificar si el texto es una nota suicida (Sección 4.4.2).

Los clasificadores de aprendizaje automático implementados en este estudio son MVS, NBM y el PMC. Estos clasificadores se seleccionaron porque previamente han sido utilizados en la tarea en cuestión y han alcanzado resultados prometedores.

Los parámetros con los que se utilizaron estos clasificadores fueron los establecidos por defecto, por ejemplo: la MVS se utilizó con un kernel lineal y para controlar el error de la clasificación el parámetro C se estableció con valor igual a uno; el PMC que es un modelo de red neuronal artificial prealimentado que procesa la información de manera unidireccional; se configuró con dos capas ocultas, la función de activación se definió en tangente y el número máximo de interacciones fue igual a 1000 épocas.

Además, se utilizó el modelo de validación cruzada con un parámetro $k = 5$, debido a que este número de interacciones es de los más comunes en la clasificación de textos.

4.4.1 Clasificación de emociones

El objetivo de esta etapa es identificar la distribución emocional latente que se expresa principalmente en notas de estudio. Para lograr tal objetivo, primero se entrena al primer clasificador con el CEASE (Sección 4.1.5). Debido a que este conjunto este etiquetado con trece emociones, se obtienen vectores de 13 dimensiones. Cada dimensión representa el grado en el que determinada emoción está presente en la nota analizada.

Por ejemplo, el documento: *"I always knew it would come to this I was chronically ill then homeless now I'm going to kill myself cause there's no way out and earth is populated by evil devils. I'm going to jump and have no one to say goodbye too I have no friend or love. I'm going to die tonight goodbye"*, genera la distribución emocional que se muestra en la Figura 19 donde se aprecia el grado en que cada emoción está presente en dicho texto.

abuso	ira	culpa*	miedo	perdón	culpa**	felicidad_ tranquilidad	esperanza	desesperanza	amor	orgullo	pena	agradecimiento
0.02	0.30	0.06	0.05	0.01	0.02	0.04	0.08	0.14	0.01	0.02	0.27	0.00

Figura 19. Vector de probabilidad emocional

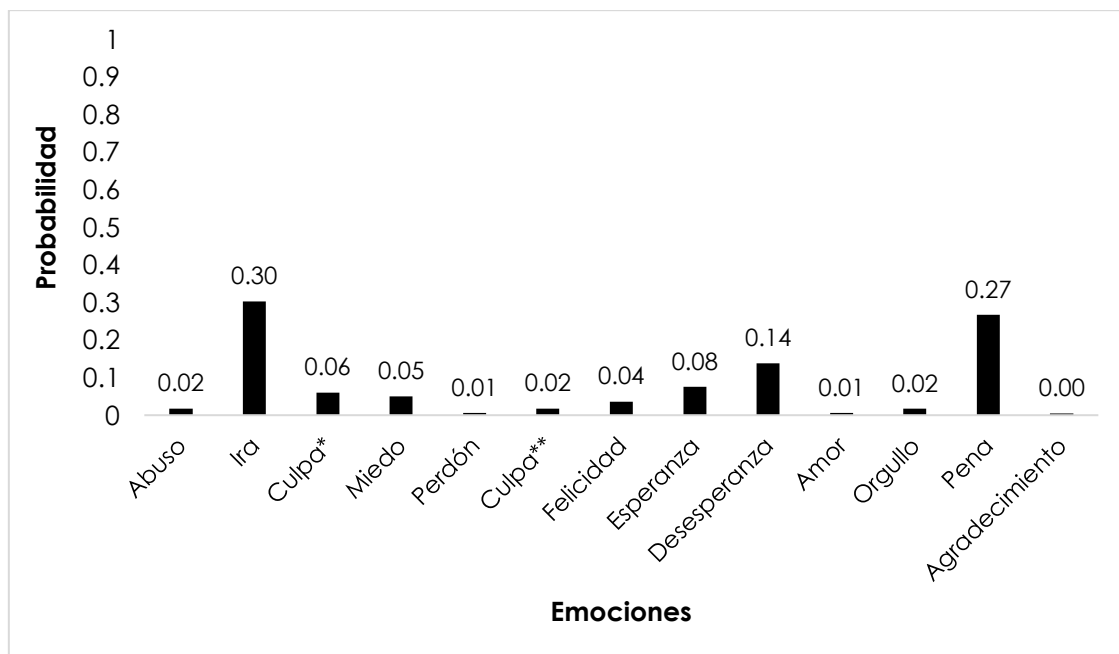


Figura 20. Probabilidad emocional en una nota de ideación suicida

La Figura 20 presenta gráficamente el nivel de cada emoción identificada en la nota analizada. Esta distribución de emociones es la entrada al clasificador de notas (Sección 4.4.2).

En esta etapa se obtuvo las distribuciones emocionales de cada nota analizada. En concreto, se estudiaron notas de dos conjuntos de datos: un conjunto correspondiente a notas suicidas y el otro, a la clase de contraejemplo, que podría ser cualquiera de los siguientes: notas de intenciones suicidas, no relacionadas al suicidio, o las de individuos sentenciados.

4.4.2 Clasificación de notas

El clasificador de notas utiliza como entrada los vectores de distribución emocional de cada nota de los dos conjuntos de datos examinados previamente y considera dos clases (suicida y no suicida). De esta manera, el clasificador determina automáticamente a qué clase pertenece la distribución de entrada que corresponde a determinada nota.

El desempeño de los clasificadores en la detección de notas suicidas se evaluó mediante el método de validación cruzada con $k = 5$ y las medidas de precisión, exhaustividad y medida F.



Capítulo 5

Experimentación y resultados

En este capítulo se describen los experimentos realizados y los resultados alcanzados con el método propuesto en la detección automática de notas de ideación suicida mediante el análisis emocional de los textos.

5.1 Preprocesamiento de los datos

El preprocesamiento es una etapa importante en el análisis de datos textuales debido a que permite eliminar el ruido lingüístico que pudiera afectar la clasificación final (Farzindar e Inkpen, 2015).

Se aplicó el siguiente preprocesamiento a los conjuntos de datos:

- **Conversión a minúsculas:** transformamos el contenido textual a minúsculas, debido a que palabras como *Suicidio* y *suicidio* serían consideradas como términos diferentes.
- **Eliminar caracteres y símbolos especiales:** en esta etapa eliminamos los símbolos, pictogramas, corchetes, caracteres del código ASCII y direcciones URL. Este procedimiento se realizó porque este tipo de caracteres no brindan información útil a la tarea en cuestión.
- **Reemplazar contracciones:** se reemplazaron las contracciones por su expansión equivalente, por ejemplo: en el idioma inglés suelen utilizarse contracciones como *we're*. Dejar de esta forma la contracción podría conllevar la pérdida de información debido a que al eliminar el apóstrofe y separar ambas palabras, tendría significado únicamente un token (*we*).

Para evitar estos casos se realizó la expansión de contracción, por ejemplo, la contracción *we're* se expandió a *we are*, obteniendo dos tokens.

- **Tokenización:** en este estudio se generaron tokens a nivel de palabra.
- **Eliminación de signos de puntuación y números:** se removieron los signos de puntuación y números porque no brindan información que ayuden a la detección de ideación suicida.
- **Eliminación de palabras vacías:** se eliminaron las palabras vacías porque se consideran palabras no significantes. Además, Jain et al. (2022) afirman que eliminar este tipo de palabras no modifica el significado de la sentencia.

5.2 Generación de los vectores de características

En este estudio se realizaron diversos experimentos donde se aplicaron diferentes métodos de generación de características (Sección 4.3). El objetivo de probar estos métodos, por un lado, fue identificar qué tipo de información (léxica o semántica) es más útil en el estudio de notas suicidas.

Por otro lado, nos permiten encontrar el número de características que mejoran el desempeño del clasificador. En la Tabla 10 se muestra un ejemplo de los resultados obtenidos de la clasificación en cascada entre notas de ideación suicida y notas de sentenciados, mediante el método de mapeo ALD y el clasificador de MVS. En esta tabla se puede observar el desempeño del clasificador al considerar diferente longitud del vector de características.

Tabla 10. Desempeño de clasificación con ALD y MVS

Tamaño del vector de características	P	E	F
3	59.24	34.3	47.97
4	56.63	89.01	51.47
5	75.5	7.06	39.88
6	40.78	61.76	40.68
7	54.94	85.71	37.75
8	50.23	85.49	41.92
9	55.49	77.8	44.39
10	62.18	53.63	43.65
20	69.18	42.42	35.45
30	50.47	97.58	36.87
40	50.48	97.36	37.02
50	50.23	99.12	34.85
60	49.95	97.57	34.95
70	49.66	97.14	34.1
80	50.56	47.03	37.97
90	51.52	97.57	40.86

100	51.02	95.6	39.61
200	50.81	95.6	39.12
300	59.24	34.3	47.97

5.3 Clasificación

En este estudio se implementaron y compararon dos enfoques:

- **Clasificación directa:** consiste en ingresar los vectores de características directamente a algún clasificador para determinar la clase la que pertenece el texto de estudio en relación con la presencia o ausencia de ciertas palabras (Sección 2.2.2).

En este estudio, el enfoque de clasificación directa sirve como punto de comparación para validar el desempeño del método propuesto.

- **Clasificación en cascada:** se refiere al método propuesto que consiste en un ensamble de clasificadores con una arquitectura en cascada, donde la salida del primer clasificador (distribución emocional) se utiliza para entrenar al segundo clasificador que en función de esa entrada determina si un texto es una nota de ideación suicida.

5.3.1 Intención suicida y ejecución: dos caras de la muerte

La clasificación entre las notas de ideación suicida obtenidas de Reddit y las notas de personas sentenciadas a muerte se realizó con la finalidad de conocer el desempeño del método propuesto cuando ambas notas tratan el mismo tema (muerte) pero difieren de circunstancia.

Teniendo en cuenta el ejemplo otorgado en la Sección 4.4.1 donde se explica cómo se obtienen las emociones de un documento, es importante resaltar que de cada documento se genera una distribución emocional diferente, dependiendo del contenido de dicha nota. Por lo tanto, las siguientes gráficas de barras exponen el promedio de las emociones plasmadas en los conjuntos de datos analizados.

La Figura 21 revela que las personas con ideación suicida (de acuerdo con el conjunto de datos utilizado) están más influenciadas por emociones negativas como: *pena* (0.42%), *desesperanza* (0.14%) e *ira* (0.09%). Sin embargo, también se muestra la presencia de emociones positivas en menor grado, tal como *esperanza* (0.09%) y *felicidad* (0.06%).

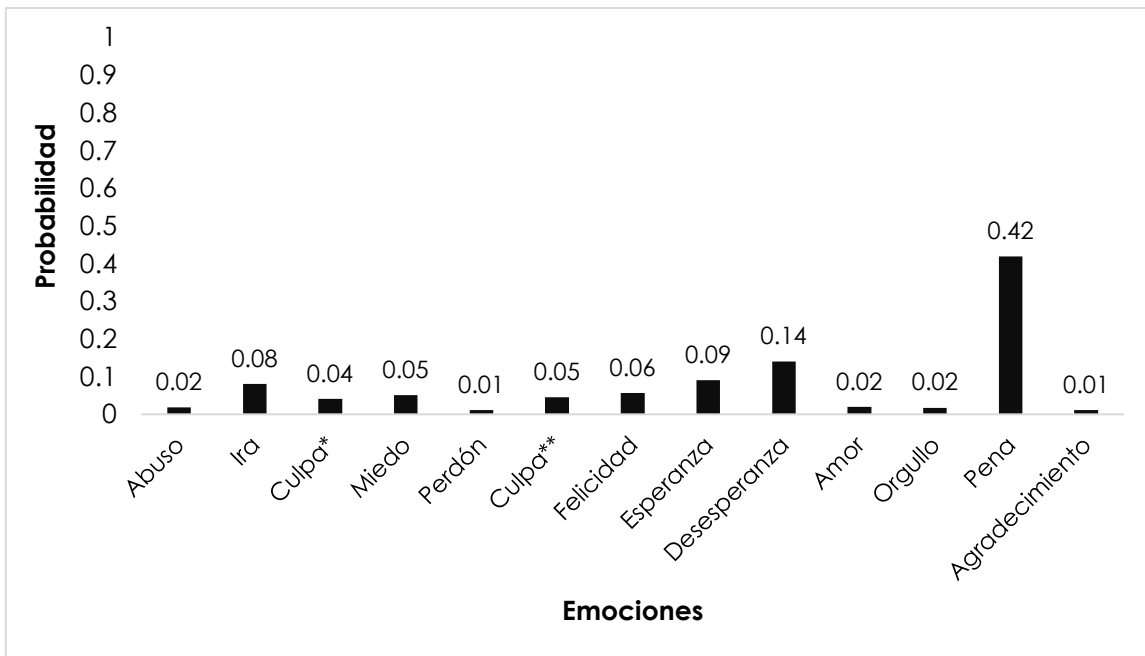


Figura 21. Emociones en notas de ideación suicida

En cambio, en la Figura 22 se muestra que en promedio las notas de personas sentenciadas a muerte reflejan mayor presencia de emociones con polaridad positiva. La emoción con mayor presencia (0.34%) es *esperanza*, seguida de las emociones *amor* y *agradecimiento* (0.14%).

En este caso, las emociones negativas están menos presentes, por ejemplo, *pena* (0.14%), *culpa*** (0.05%) y *desesperanza* (0.03%).

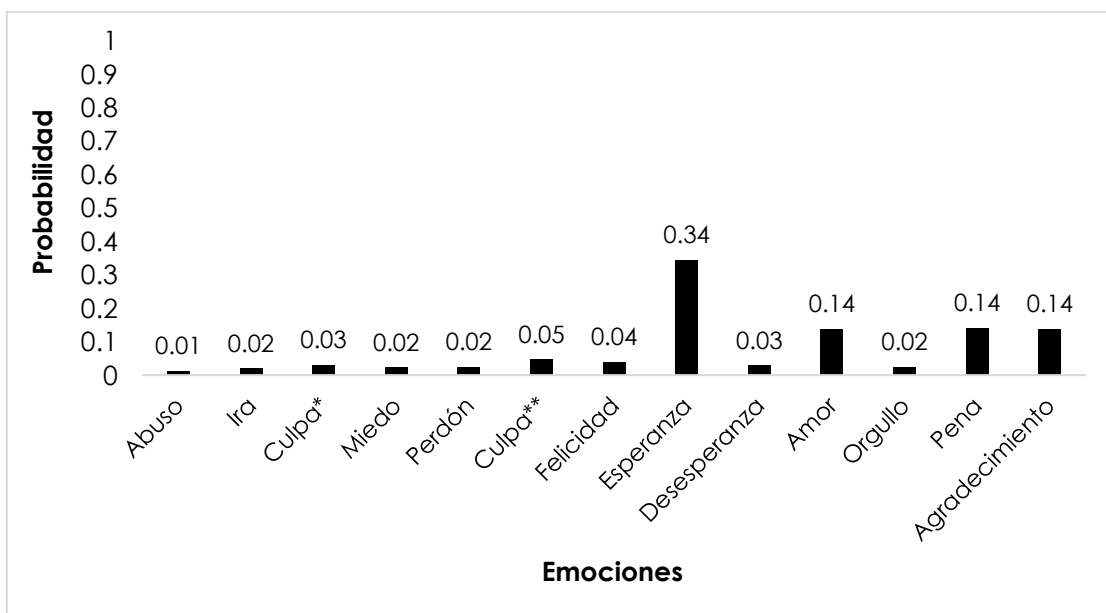


Figura 22. Emociones en notas de personas sentenciadas

En general, entre ambos tipos de notas existe una diferencia del 0.28% de la emoción *pena* y 0.25% de la emoción *esperanza*. Siendo la emoción *pena* la predominante en las notas de personas con ideación suicida, mientras que la emoción *esperanza* es la relevante en las notas de personas sentenciadas.

Los resultados obtenidos de la clasificación entre notas de ideación suicida y notas de personas sentenciadas se muestran en la Tabla 13. En esta tabla se compara el desempeño de dicha clasificación mediante el enfoque directo y en cascada.

Tabla 11. Resultados de la clasificación entre notas asociadas con la muerte

Aproximación	Método	P	E	F
Directa	FT- FDI + MVS	97.21%	98.46%	97.79%
	FT- FDI + PMC	98.03%	96.92%	97.9%
	FT- FDI + NB	95.97%	99.12%	97.46%
Cascada	D2V + MVS	96.37%	96.9%	96.57%
	ALD + PMC	51.16%	41.45%	92.04%
	FT- FDI + NB	79.59%	92.71%	84.18%

Como se observa en la tabla anterior, la *clasificación directa* obtiene el mejor resultado con un 97.9% de medida F al considerar el método de mapeo FT- FDI y el PMC; mientras que el mejor desempeño de la clasificación en cascada se obtuvo con las características obtenidas del método de mapeo D2V y el algoritmo de MVS, obteniendo el 96.57% de medida F.

Es importante remarcar que, en la clasificación directa la presencia de ciertas palabras es suficiente para determinar a qué clase pertenece cada nota. En comparación con la clasificación en cascada, donde la información semántica que considera el contexto de las palabras es más relevante.

Adicionalmente, se seleccionaron los diez términos más frecuentes de cada clase (Tabla 14) para analizar el vocabulario de cada conjunto de datos. Este análisis se llevó a cabo con el objetivo de conocer que términos utilizan más las personas suicidas y las sentenciadas para expresarse ante el tema de muerte, cuando la circunstancia que las puso en tal situación es diferente. Las personas que escribieron las notas de Reddit desean morir voluntariamente, en comparación con las personas sentenciadas quienes son obligadas.

Tabla 12. Vocabulario más frecuente en las notas relacionadas con la muerte

No.	Notas de ideación suicida		Notas de personas sentenciadas	
	Palabra	Frecuencia	Palabra	Frecuencia
1	im	600	love	816
2	want	416	family	371
3	life	366	know	356
4	like	324	thank	331
5	know	304	sorry	285
6	suicide	280	want	257
7	feel	238	like	255
8	even	236	god	247
9	people	203	say	200
10	die	193	life	174

Del análisis del vocabulario, observamos que las personas que comparten sus notas suicidas en la red social Reddit suelen expresar abiertamente sus intenciones de morir, por lo tanto, utilizan términos que refieren específicamente a dicho objetivo. Algunas de las palabras más empleadas en este tipo de notas son: *vida*, *suicidio* y *muerte*.

En los escritos de las personas sentenciadas predomina el uso de palabras con connotación positiva que podrían estar vinculadas a algún tipo de agradecimiento. Entre las palabras más relevantes se encuentran: *amor*, *familia*, *gracias*, *Dios* y *vida*.

Los resultados obtenidos de esta clasificación nos permiten plantear la hipótesis de que la clasificación directa obtuvo un alto rendimiento debido a la información explícita contenida en las notas de ideación suicida; es decir, existen palabras predominantes en una clase que no están en la otra. Esta característica permite al clasificador diferenciar directamente ambos conjuntos de datos, reflejándose en el desempeño obtenido.

Para confirmar la hipótesis planteada previamente, realizamos un segundo experimento en el cual se clasifican notas de suicidio consumado y notas no relacionadas con el suicidio. Se escogieron ambos conjuntos de datos porque, según su naturaleza, no presentan términos muy relacionados con determinado tema. Con esta clasificación se trata de evitar que las clases sean fácilmente diferenciadas por el clasificador, tal como sucedió en el caso previo.

5.3.2 Suicidio consumado: la necesidad de analizar más allá de las palabras

La Figura 23 muestra que en promedio las personas que enunciaron su intención suicida en alguna nota y que consumaron el acto presentan emociones de ambas polaridades (negativas - positivas). Por ejemplo, muestran emociones como *pena* (0.34%), *esperanza* (0.17), *desesperanza* (0.09%), *amor* (0.07%), *culpa* (0.06%) e *ira* (0.06%).

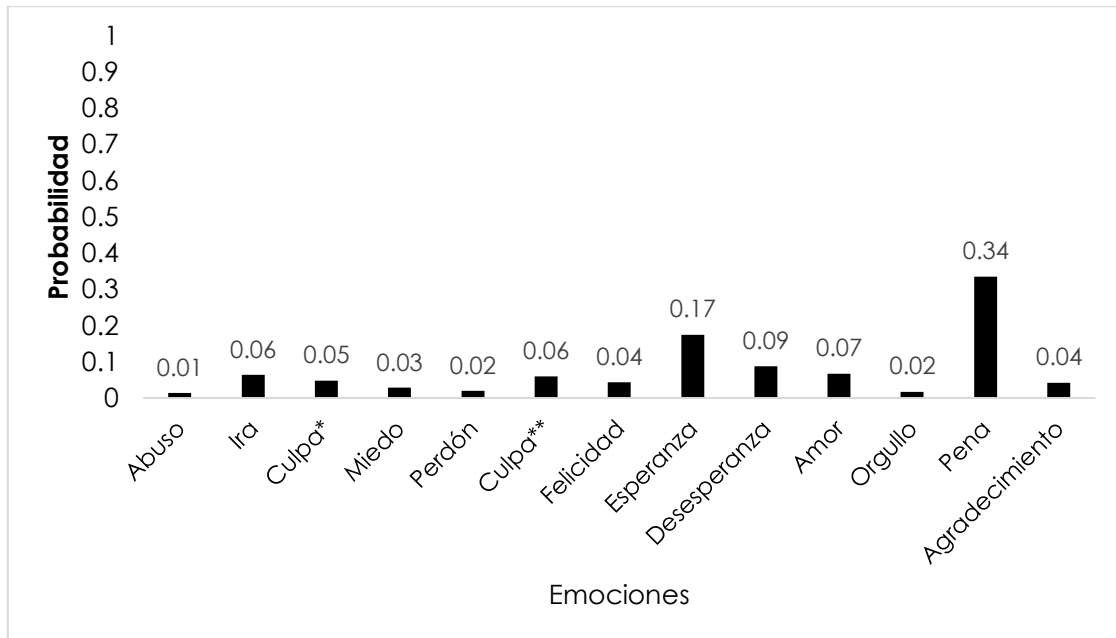


Figura 23. Emociones en notas de suicidio consumado

Las notas de temáticas, no asociadas al suicidio o al tema de muerte, muestran una distribución emocional mixta (Figura 24). En esta distribución se presentan mayormente las siguientes emociones: *pena* (0.38%), *esperanza* (0.17), *desesperanza* (0.08%), *amor* (0.07%), *culpa* (0.05%) e *ira* (0.05%).

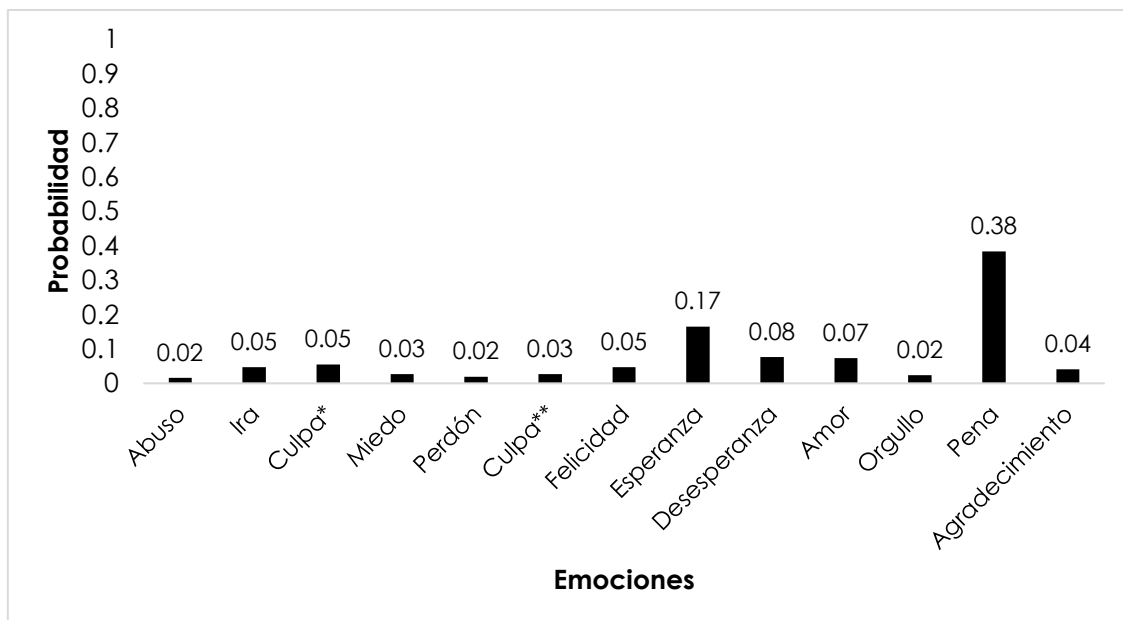


Figura 24. Emociones en notas no asociadas con el suicidio

La Tabla 15 muestra que el enfoque de clasificación directa obtiene un 90.85% de medida F en el mejor de los casos, mediante el uso de las características semánticas que consideran el contexto y el algoritmo de MVS. Mientras que con el enfoque en cascada obtiene el 94.84% de medida F al considerar información semántica contextual y usar el clasificador NB.

Tabla 13. Resultados de la clasificación entre notas suicidas y no suicidas

Aproximación	Método	P	E	F
Directa	D2V + MVS	85.76%	92.26%	90.85%
	D2V + PMC	83.33%	88.39%	89.19%
	D2V + NB	85.33%	90.97%	87.66%
Cascada	D2V + MVS	81.61%	90.71%	88.98%
	ALD + PMC	79.11%	88.69%	85.78%
	D2V + NB	95.96%	90.69%	94.84%

Posteriormente, se realizó el análisis de los diez términos más frecuentes de cada categoría. Esta comparación se llevó a cabo con el objetivo de observar que términos predominan en ambas clases y determinar si dichas palabras influyen en el desempeño de los clasificadores.

La Tabla 16 refleja que entre los dos tipos de notas existen palabras muy comunes que con facilidad podrían emplearse en otros contextos, estos términos son: *tiempo*, *amor*, *disculpa*, *juego* y *sentir*.

Tabla 14. Términos más frecuentes en las notas suicidas y ajenas al suicidio

No.	Notas de suicidio consumado		Notas de no suicidio	
	Palabra	Frecuencia	Palabra	Frecuencia
1	never	22	people	169
2	time	19	like	136
3	love	18	get	94
4	know	17	would	73
5	life	14	really	69
6	feel	13	time	65
7	would	12	good	63
8	see	12	make	62
9	sorry	11	know	62
10	want	9	game	58

Podemos destacar que, entre los diez términos más presentes en las notas de suicidio, no se muestran palabras asociadas a dicha temática, lo que indica que no todas las personas con pensamientos suicidas usan palabras al tema en cuestión para expresar su intención. Sin embargo, el que no expresen explícitamente dicho propósito, no es indicador de que el individuo no corre riesgo de atentar contra su propia vida.

Los resultados de esta clasificación nos muestran que la clasificación directa disminuye su rendimiento cuando no hay términos explícitos relacionados con el acto suicida. Por el contrario, el rendimiento de la clasificación en cascada es superior, aunque se carezca de términos asociados al suicidio. Por lo tanto, puede concluirse que el enfoque directo no es robusto para clasificar notas en las que los pensamientos suicidas no se exponen explícitamente, sino de forma latente.

Además, nos indica que detectar notas de suicidio en las que no se expone explícitamente el deseo o intención de morir, es una tarea significativa para prevenir intentos de suicidio e incluso actos consumados. Su relevancia se debe a que no todas las personas se sienten en la libertad de compartir sus pensamientos de muerte abiertamente.

5.3.3 Emociones finales: notas de suicidio consumado y personas sentenciadas

Para validar la robustez del modelo propuesto, se realizó la clasificación entre *notas de suicidio consumado* y *notas de personas sentenciadas a muerte*. Se seleccionaron estos conjuntos de datos, porque ambos tratan el mismo tema de interés (*muerte*), teniendo la certeza de que en ambos casos las personas murieron. Las de suicidio consumado murieron por decisión propia; mientras que los sentenciados fueron obligados.

Es importante resaltar que la diferencia de esta clasificación en contraste con la primera descrita en este capítulo, se centra en que ahora se tiene la certeza de muerte, mientras que en el caso anterior se desconoce si la persona con ideación suicida consumó el acto.

Cabe señalar que el proceso de clasificación es diferente a los anteriores; en esta ocasión decidimos concatenar los vectores de MTD con los vectores de otros métodos de generación de características (Sección 5.2) porque se ha demostrado que la combinación de características mejora los resultados.

El análisis de emociones muestra que en promedio las personas que consuman el acto están mayormente influenciadas por emociones negativas (Figura 25), tal como: *pena* (0.25%), *desesperanza* (0.09%), *culpa* (0.09) e *ira* (0.09%). Por otro lado, las emociones positivas más frecuentes son *amor* y *esperanza*, con 0.13% de presencia, seguida de *agradecimiento* (0.05%).

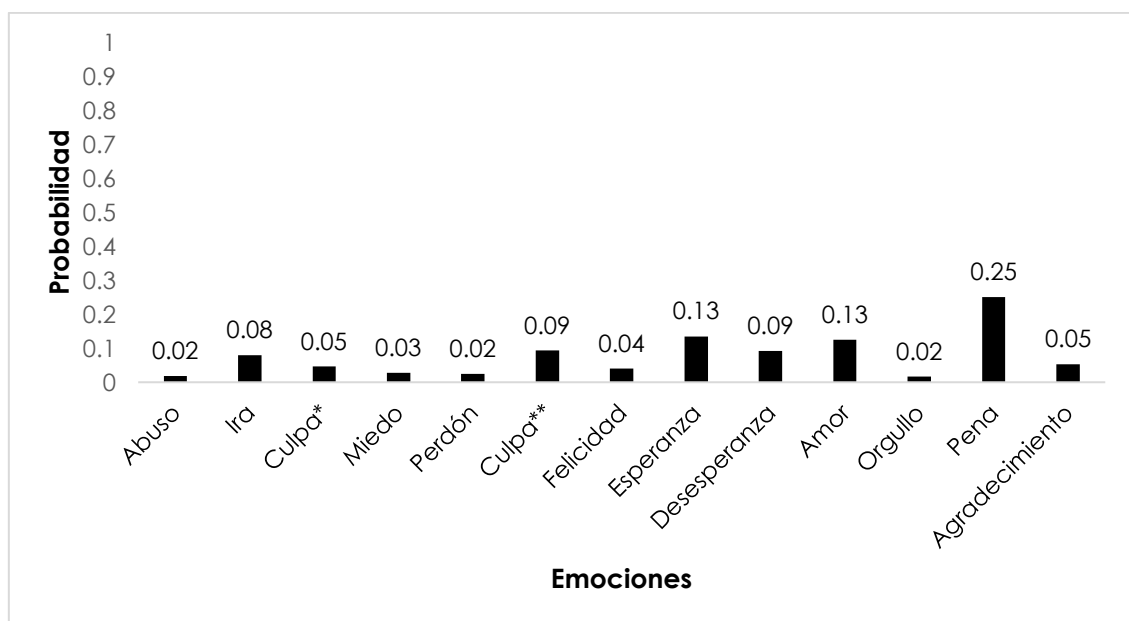


Figura 25. Emociones previas al suicidio consumado

En las notas de las personas sentenciadas a muerte, se muestra que en promedio este tipo de personas suelen tener un contenido emocional ligeramente más positivo. Como se observa en la Figura 26. Las emociones positivas más presentes son: amor (0.21%), esperanza (0.10%) y agradecimiento (0.08%). Entre las emociones negativas más relevantes se encuentra la pena (0.18%), desesperanza (0.09%), ira (0.08%) y culpa** (0.07%).

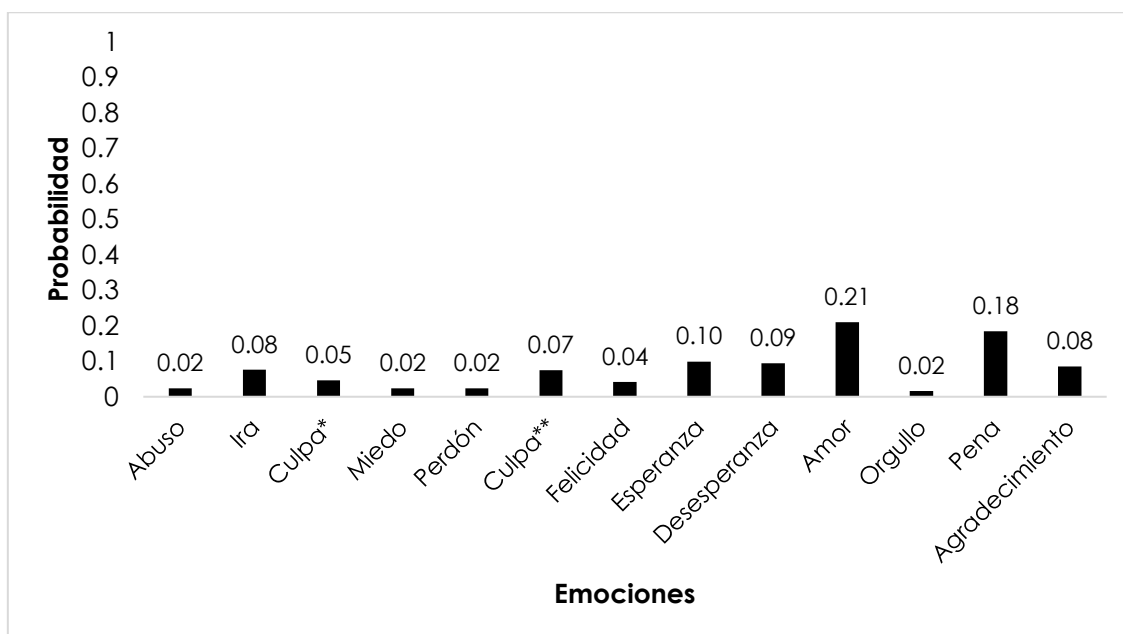


Figura 26. Emociones antes de ser ejecutados

Los resultados obtenidos de esta clasificación se visualizan en la Tabla 17, donde se observa que la clasificación directa alcanza un desempeño del 91.45% de medida F mediante el análisis de características obtenidas con FT- FDI y el algoritmo de MVS; mientras que la clasificación en cascada alcanzar una medida F igual a 95,37% con las características concatenadas de FT- FDI y MTD mediante el perceptrón multicapa.

Tabla 15. Resultados de la clasificación entre notas previas a la muerte

Aproximación	Método	P	E	F1
Directa	FT- FDI + MVS	95.2%	93.36%	91.45%
	FT- FDI + PMC	91.64%	95.14%	89.62%
	D2V + NB	94.07%	78.32%	80.82%
Cascada	ALD + MTD + MVS	96.85%	32.68%	94.69%
	FT- FDI + MTD + PMC	96.68%	94.68%	95.37%
	ALD + MTD + NB	88.29%	96.90%	92.35%

La Tabla 18 muestra el análisis del vocabulario más frecuente entre las notas de suicidio consumado y de las personas sentenciadas. En ambas clases se observa el uso de palabras coloquiales, por ejemplo: *amor*, *vida* y *perdón*. En este tipo de notas a pesar de estar estrechamente relacionadas con el tema de muerte, en ninguno de los casos se expresa de forma explícita.

Tabla 16. Vocabulario más frecuente antes de morir

No.	Notas de suicidio consumado		Notas de sentenciados	
	Palabra	Frecuencia	Palabra	Frecuencia
1	<i>never</i>	22	<i>love</i>	816
2	<i>time</i>	19	<i>family</i>	371
3	<i>love</i>	18	<i>know</i>	356
4	<i>know</i>	17	<i>thank</i>	331
5	<i>life</i>	14	<i>sorry</i>	285
6	<i>feel</i>	13	<i>want</i>	257
7	<i>would</i>	12	<i>like</i>	255
8	<i>see</i>	12	<i>god</i>	247
9	<i>sorry</i>	11	<i>say</i>	200
10	<i>want</i>	9	<i>life</i>	174



Capítulo 6

Conclusiones

El desarrollo de este estudio nos permitió comprobar la **hipótesis** planteada inicialmente, la cual se describe de la siguiente manera: Si se utiliza la distribución emocional que plasma una persona en un texto, entonces se podrá detectar automáticamente notas de ideación suicida.

Asimismo, nos permitió cumplir con el **objetivo general**: Detectar automáticamente notas de ideación suicida mediante el análisis emocional expresado en textos no explícitos.

Mencionamos que se cumplen ambos requerimientos dado que nuestros experimentos (descritos en el capítulo 5) logran alcanzar resultados competentes. Mencionando que el método propuesto basado en un ensamble de clasificadores es más robusto que la clasificación directa empleada en la mayoría de los estudios previos.

El método en cascada se considera más robusto porque toma en cuenta información adicional al contenido explícito, es decir, este enfoque no basa su proceso de clasificación exclusivamente en la presencia o ausencia de ciertas palabras claves relacionadas directamente al acto suicida, tal como sucede en el enfoque de clasificación directa.

El enfoque propuesto permite extraer automáticamente las emociones plasmadas en las notas, posteriormente utilizar dicho conocimiento para clasificar las notas. Esto se puede considerar una ventaja respecto al estado del arte, porque la clasificación que realiza este enfoque no se basa exclusivamente de la presencia o ausencia de ciertas

palabras claves relacionadas directamente al acto suicida. Lo que permite detectar notas suicidas con contenido no explícito.

Detectar notas con contenido no explícito es una tarea muy importante, porque no todas las personas con pensamientos suicidas expresan abiertamente (explícitamente) su intención de cometer suicidio. Este comportamiento puede observarse en las notas de suicidio consumado analizadas en este estudio. En este tipo de notas no todos los usuarios exponen explícitamente su deseo de morir, ya que en ocasiones carecen de la presencia de palabras asociadas al acto suicida; sin embargo, las personas que escribieron dichas notas consumaron el acto.

La experimentación llevada a cabo con el método en cascada no únicamente logra resultados competentes. Además, nos brinda el conocimiento de la relación existente entre las emociones y la ideación suicida. De modo que mediante este enfoque se pueden conocer automáticamente las distribuciones de emociones involucradas y su porcentaje de pertenencia.

De acuerdo con los experimentos realizados, también se llega a la conclusión de D2V es un método de mapeo óptimo para extraer las características semánticas-contextuales que permiten clasificar correctamente notas de comportamiento suicida. Como se indicó previamente, este método tiene la ventaja (respecto a los otros métodos analizados) de determinar el sentido de las palabras que conforman una nota y el contexto en que estas se relacionan, lo que brinda mayor conocimiento al clasificador.

Los clasificadores utilizados en este estudio se consideran de los más comunes en tareas de procesamiento del lenguaje natural, sin embargo, permiten hacer la discriminación correcta entre las notas de suicidio y otro tipo de notas no asociadas al tema de suicidio.

6.1 Aportaciones

Las aportaciones que brinda esta investigación son las siguientes:

- Se encontró que es posible reconocer automáticamente notas suicidas mediante el análisis emocional
- Se encontró que no todas las personas con ideación suicida expresan abiertamente su deseo de morir.
- Se demostró que la concatenación de características y de clasificadores mejora el proceso de clasificación de textos suicidas.

- Se demostró que D2V al considerar el contexto de las palabras mejora el desempeño del clasificador cuando los textos de estudio son de contenido implícito.
- Se demuestra que es posible obtener automáticamente las emociones que influyen en actos de muerte a partir del análisis textual, indicando en que porcentaje se encuentra cada una.
- Se encontró la concatenación de MTD y FT-FDI es relevante para detectar notas con contenido implícito.
- Se publicó un artículo en la revista: Avances en Inteligencia Computacional. García-Galindo, M.d.C., Hernández-Castañeda, Á., García-Hernández, R.A., Ledeneva, Y. (2022). Automatic Identification of Suicidal Ideation in Texts Using Cascade Classifiers. In: Pichardo Lagunas, O., Martínez-Miranda, J., Martínez Seis, B. (eds) Advances in Computational Intelligence. MICAI 2022. Lecture Notes in Computer Science(), vol 13613. Springer, Cham. https://doi.org/10.1007/978-3-031-19496-2_9

Los resultados del estudio realizado se expusieron en el Congreso Internacional Mexicano de Inteligencia Artificial (MICAI), que se llevó a cabo del 24 al 29 de octubre de 2022, en la Ciudad de Monterrey.

- Se publicó un artículo de divulgación en la revista Komputer Sapiens. García-Galindo, María del Carmen., Hernández-Castañeda, Ángel., García-Hernández, René Arnulfo., Ledeneva, Yulia., & Cruz-Reyes, Rafael. (2022). Detección de notas de ideación suicida. Komputer Sapiens, 15(2).

Referencias

- Abdulsalam, A., & Alhothali, A. (2022). Suicidal Ideation Detection on Social Media: A Review of Machine Learning Methods. *arXiv preprint arXiv:2201.10515*.
- Águila Tejada, A. (2011). *Suicidio. La última decisión*. México: Trillas.
- Akhtar, M. S., Ekbal, A., & Cambria, E. (2020). How Intense Are You? Predicting Intensities of Emotions and Sentiments using Stacked Ensemble [Application Notes]. *IEEE Computational Intelligence Magazine*, 15(1), 64-75.
- Álvarez Ramos, E., Heredia Ponce, H., & Romero Oliva, M. F. (2019). La Generación Z y las Redes Sociales. Una visión desde los adolescentes en España. *Revista Espacios*, 40(20), 1-13.
- Anguita, D., Ghelardoni, L., Ghio, A., Oneto, L., & Ridella, S. (2012). Validación cruzada 'K' in K-fold. *ESANN*, 441-446.
- Arroyo Fernández, A., & Bertomeu Ruiz, A. (2012). Métodos suicidas e Internet. *Revista Española de Medicina Legal*, 38(4), 143-148.
- Awwalu, J., Abdullahi, S. E., & Nonyelum, O. (2020). Naïve Bayes Emotion Classification of Final Statements from Death Row Youths before Execution. *Malaysian Journal of Computing and Applied Mathematics*, 3(2), 60-72.
- Barroso Martínez, A. A. (2019). Comprender el suicidio desde una perspectiva de género: una revisión crítica bibliográfica. *Revista de la Asociación Española de Neuropsiquiatría*, 39(135), 51-66.
- Berrar, D. (2019). Cross-validation. *Data Science Laboratory*, 542-545.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3, 993-1022.
- Bonet Mas, C., Palma Sevillano, C., & Gimeno-Santos, M. (2020). Riesgo de suicidio, inteligencia emocional y necesidades psicológicas básicas en adolescentes tutelados en centros residenciales. *Revista de Psicología Clínica con Niños y Adolescentes*, 7(1), 30-37.
- Caicedo, R. W. (2020). Assessment of supervised classifiers for the task of detecting messages with suicidal ideation. *Heliyon*, 6(8).
- Castillo-Sánchez, G., Marques, G., Dorrnzoro, E., Rivera-Romero, O., Franco-Martín, M., & De la Torre-Díez, I. (2020). Suicide risk assessment using machine learning and social networks: a scoping review. *Journal of medical systems*, 44(12).
- Chávez-Hernández, A. M., & Leenaars, A. A. (2010). Edwin S Shneidman y la suicidología moderna. *Salud mental*, 33(4), 355-360.
- Chowdhary, K. (2020). *Introducing Artificial Intelligence*. In: *Fundamentals of Artificial Intelligence*. New Delhi: Springer.
- Cortez, V. A., Vega, H. H., & Pariona, Q. J. (2009). Procesamiento de lenguaje natural. *Revista de investigación de Sistemas e Informática*, 6(2), 45-54.
- Cox, P. L. (2020). Aportes y daños derivados del uso de redes sociales en adolescentes. *Psicoterapia Integrativa*, 11, 35.

- Dalal, M. K., & Zaveri, M. A. (2011). Automatic text classification: a technical review. *International Journal of Computer Applications*, 28(2), 37-40.
- De la Cruz, L. N., & Zúñiga, T. (2017). *Ideación suicida y regulación emocional en adolescentes del municipio de Soledad, Atlántico*. Universidad de la Costa, C.U.C., Barranquilla.
- Desmet, B., & Hoste, V. (2013). Emotion detection in suicide notes. *Expert Systems with Applications*, 10(16), 6351-6358.
- Dimock, M. (17 de Enero de 2019). *Pew Research Center*. Recuperado el 11 de Abril de 2023, de Defining generations: Where Millennials end and Generation Z begins: <https://www.pewresearch.org/fact-tank/2019/01/17/where-millennials-end-and-generation-z-begins/>
- Durkee, T., Hadlaczky, G., Westerlund, M., & Carli, V. (2011). Internet Pathways in Suicidality: A Review of the Evidence. *International Journal of Environmental Research and Public Health*, 8(10), 3938-3952.
- Duval, M. A., Shulcloper, J., & Vega, S. (2012). Combinación de clasificadores supervisados: estado del arte. *Reporte técnico. RNPS*. Centro de aplicaciones de tecnología avanzada, Cuba.
- Echeburúa, E. (2015). The many faces of suicide in the psychological clinic. *Terapia psicológica*, 33(2), 117-126.
- Farzindar, A., & Inkpen, D. (2015). *Natural language processing for social media*. Toronto: Morgan & Claypool Publishers series Synthesis Lectures on Human Language Technologies.
- Fata, I. A., Yusuf, Y. Q., Kamal, R., & Namaziandost, E. (2021). The characteristics of linguistic features enfolded in suicide notes. *Journal of Language and Linguistic Studies*, 17(2), 720-735.
- Fehling, K. B., & Selby, E. A. (2021). Suicide in DSM-5: Current Evidence for the Proposed Suicide Behavior Disorder and Other Possible Improvements. *Frontiers in Psychiatry*, 11.
- Ghosh, S., Ekbal, A., & Bhattacharyya, P. (2020). CEASE, a Corpus of Emotion Annotated Suicide notes in English. *Proceedings of the Twelfth Language Resources and Evaluation Conference* (págs. 1618–1626). Marseille, France: European Language Resources Association. Obtenido de <https://aclanthology.org/2020.lrec-1.201/>
- Ghosh, S., Ekbal, A., & Bhattacharyya, P. (2022). Deep cascaded multitask framework for detection of temporal orientation, sentiment and emotion from suicide notes. *Scientific reports*, 12(1), 1-16.
- Gunn, J. F., & Lester, D. (2012). Twitter postings and suicide: An analysis of the postings of a fatal suicide in the 24 hours prior to death. *Suicidologi*.
- Han, J., & Kamber, M. (2006). *Data mining: concepts and techniques*. San Francisco CA: Morgan Kaufmann.
- Haykin, S. (2009). *Neural networks and learning machines*. Pearson Education.
- Hernández-Castañeda, Á. (2017). *Detección automática de texto engañoso mediante algoritmos de modelos basados en tópicos (Tesis de Doctorado)*. Instituto Politécnico Nacional; Centro de Investigación en Computación, México.

- Obtenido de <https://tesis.ipn.mx/bitstream/handle/123456789/25836/T1931.pdf?sequence=1&isAllowed=y>
- Hernández-Castañeda, Á., García Hernández, R. A., Ledeneva, Y., & Millán Hernández, C. E. (2020). The impact of key ideas on automatic deception detection in text. *Computación y Sistemas*, 24(3), 1229--1239.
- Jain, A. K., Mao, J., & Mohiuddin, K. M. (1996). Artificial neural networks: A tutorial. *Computer*.
- Jain, P., Srinivas, K. R., & Vichare, A. (2022). Depression and suicide analysis using machine learning and NLP. In *Journal of Physics: Conference Series*. 2161. Mumbai, India: IOP Publishing.
- Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. *Multimedia Tools and Applications*, 78, 15169--15211.
- Kulkarni, S., Kulkarni, R. S., & Lorenz, K. E. (2022). Artificial Intelligence (AI) in Psychiatry-A Summary. *Global Journal of Medical Research: A Neurology & Nervous System*, 22(3), 1-13.
- Le, Q., & Mikolov, T. (2014). Distributed Representations of Sentences and Documents. *Proceedings of the 31st International Conference on Machine Learning* (págs. 1188-1196). Proceedings of Machine Learning Research.
- Ledeneva, Y. N., & García Hernández, R. A. (2017). *Generación automática de resúmenes-Retos, propuestas y experimentos*. México: Universidad Autónoma del Estado de México.
- Lee, Y. H. (2019). Identifying suicide notes using forensic linguistics and machine learning. *언어학*, 27(2), 171-191.
- Lu, Y. (1996). Knowledge integration in a multiple classifier system. *Appl Intell*, 6.
- Lui, B. (2012). *Sentiment analysis and opinion mining*. Morgan & Claypool.
- Malini, N., & Tan, V. (2016). Forensic linguistics analysis of Virginia Woolf's suicide notes. *International Journal of Education*, 53-58.
- Mariñelarena-Dondena, L., Errecalde, M. L., & Castro Solano, A. (2017). Extracción de conocimiento con técnicas de minería de textos aplicadas a la psicología. *Revista Argentina de Ciencias del Comportamiento*, 9(2).
- Martínez Restrepo, S., & Jossa Jaramillo, J. (2021). *(App) braza la vida: prevención del riesgo de comportamiento suicida en adolescentes, por la influencia del contenido digital de algunas plataformas*. Grupo de investigación psicología, salud y sociedad – línea de investigación estudios (Universidad CES), Medellín.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems*, 26.
- Monwabisi, R. K., Russell, K. H., & Heydon, G. (2022). *Language and the Law: Global Perspectives in Forensic Linguistics from Africa and beyond*. Western: African Sun Media.

- Murcia, S. C., & Vargas, Y. K. (2020). Factores relacionados al suicidio en adolescentes. *Tesis de licenciatura en Psicología*. Institución Universitaria Politécnica Grancolombio, Colombia.
- Olivares Pérez, S. (2019). Uso de internet y conductas suicidas ne adolescentes de 14 a 18 años en México. *Visión criminológica-criminalística*, 1-18.
- Organización Mundial de la Salud. (17 de Junio de 2021). *Organización Mundial de la Salud*. Obtenido de Suicidio: <https://www.who.int/es/news-room/fact-sheets/detail/suicide>
- Ortega González, M. (2018). *Comportamiento suicida. Reflexiones críticas para su estudio desde un sistema psicológico*. México: Qartuppi.
- Pallarés, M. (2010). *Emociones y sentimientos. Dónde se forman y cómo se transforman* (1era ed.). Barcelona (España): Marge Books.
- Parrott, S., Britt, B. C., Hayes, J. L., & Albright, D. L. (2019). Social Media and Suicide: A Validation of Terms to Help Identify Suicide-related Social Media Posts. *J Evid Based Soc Work*, 624-634.
- Rajpura, N., Tejani, A., Upadhyay, S., & D'Britto, R. (2021). Suicidal Tendencies and Ideation Prediction using Reddit. *International Research Journal of Engineering and Technology*, 8.
- Ramírez, G. M., Carrillo, R. M., & Sánchez, L. A. (2015). Combinación de clasificadores para el análisis de sentimientos. *Benemérita Universidad Autónoma de Puebla, Facultad de Ciencias de la Computación*, 1-14.
- Rangel-Garzón, C. X., Suárez-Beltrán, M. F., & Escobar-Córdoba, F. (2015). Escalas de evaluación de riesgo suicida en atención primaria. *Revista de la Facultad de Medicina*, 63(4).
- Real Academia Española. (14 de Marzo de 2023). *Diccionario de la lengua española*. Obtenido de <https://dle.rae.es/suicidio>
- Rodríguez-Esparza, L. J., Barraza-Barraza, D., Salazar-Ibarra, J., & Vargas-Pasaye, R. G. (2021). Metodología de Análisis de Emociones para Identificar Riesgo de Cometer Suicidio Generado por el COVID-19. *Revista Lasallista de Investigación*, 18(2), 105-124.
- Rouhiainen, L. (2018). *Inteligencia artificial: 101 cosas que debes saber hoy sobre nuestro futuro*. Barcelona: Alienta Editorial.
- Russell, S. J., & Norvig, P. (2016). *Artificial Intelligence A modern approach*. England: Pearson Education Limited.
- Sarkar, D. (2016). *Text Analytics with Python: A Practical Real-World Approach to Gaining Actionable Insights from Your Data*. India: Apress.
- Sarsam, S., Al-Samarraie, H., Alzahrani, A., Mon, C., & Shibghatullah, A. (2022). Characterizing Suicide Ideation by Using Mental Disorder Features on Microblogs: A Machine Learning Perspective. *Int J Ment Health Addiction*, 1-14.
- Sawhney, R., Manchanda, P., Singh, R., & Aggarwal, S. (2018). A computational approach to feature extraction for identification of suicidal ideation in tweets. *Proceedings of ACL 2018, Student Research Workshop* (págs. 91-98). ACL.

- Seger, C. (2018). *An investigation of categorical variable encoding techniques in machine learning: binary versus one-hot and feature hashing*. STOCKHOLM, SWEDEN: School of Electrical Engineering and Computer Science.
- Shneidman, E. (1985). *Definition of suicide*. New York: John Wiley & Sons.
- Srivastava, A. N., & Sahami, M. (2009). *Text mining: Classification, clustering, and applications*. CRC press.
- Tadesse, M. M., Lin, H., Xu, B., & Yang, L. (2019). Detection of suicide ideation in social media forums using deep learning. *Algorithms*, 13(1), 7.
- Talib, R., Hanif, M. K., Ayesha, S., & Fatima, F. (2016). Text mining: techniques, applications and issues. *International Journal of Advanced Computer Science and Applications*, 7(11).
- The Center For Generational Kinestics. (2018). *The state of Gen Z 2018, research findings*. Recuperado el 11 de Abril de 2023, de New study finds Gen Z creates digital celebrities out of social influencers: <https://genhq.com/state-of-gen-z-2018-research-findings/>
- Troya-Capa, C., Mocha-Brito, L., & Cabrera-Sarango, M. (2020). Revisión sistemática de literatura: Inteligencia Artificial aplicada a la predicción de pensamientos suicidas. *Congreso Estudiantil de Inteligencia Artificial Aplicada a la Ingeniería y Tecnología* (págs. 1-8). Estado de México: UNAM, FESC.
- Vecina Jiménez, M. L. (2006). Emociones positivas. *Pap. psicol*, 9-17.
- Viera, A. F. (2017). Técnicas de aprendizaje de máquina utilizadas para la minería de texto. *Investigación bibliotecológica*, 31(71), 103--126.
- Weishaar, M. E., & Beck, A. T. (1992). Hopelessness and suicide. *International Review of Psychiatry*, 4(2), 177-184.
- Yao, H. R. (2020). Detection of suicidality among opioid users on reddit: machine learning-based approach. *Journal of medical internet research*, 22(11).
- Žižka, J., Dařena, F., & Svoboda, A. (2019). *Text mining with machine learning: principles and techniques*. CRC Press.