

Estado del arte en el reconocimiento de emociones en textos de redes sociales

Cristian C. Segura, Hiram Calvo

Instituto Politécnico Nacional,
Laboratorio de Ciencias Cognitivas Computacionales,
Centro de Investigación en Computación,
Mexico

{cseguram2021, hcalvo}@cic.ipn.mx

Resumen El reconocimiento de emociones es una tarea de gran importancia en las tareas de procesamiento de lenguaje natural con diferentes aplicaciones en varias áreas del conocimiento. La detección implícita de emociones en los textos principalmente en aquellas interacciones en las redes sociales es posiblemente uno de los problemas más desafiante a resolver debido a que las emociones a transmitir y expresar se encuentran ocultas en el texto y encontrarlas requiere comprender el contexto de los diferentes tipos de texto. En el presente documento se revisará el estado del arte de los métodos actuales de reconocimiento de emociones con diferentes fuentes y en diferentes contextos identificando los diferentes enfoques utilizados en el proceso para así proceder con la comparación entre la literatura identificada. Finalmente se presenta una propuesta basada en el procesamiento de textos distorsionados, debido a que éste no ha sido abordado a profundidad en los trabajos presentados anteriormente.

Palabras clave: Reconocimiento de emociones, procesamiento de lenguaje natural, detección implícita, redes sociales, análisis de texto, contexto.

A Survey of State of the Art in the Recognition of Emotions in Texts in Social Networks

Abstract. Emotion detection is a task of great importance in natural language processing tasks with different applications in various topics of knowledge. The implicit detection of emotions in the texts, mainly in those interactions in social networks, is possibly one of the most challenging problem to solve because the emotions to be transmitted and expressed are hidden in the text and finding them requires understanding the context of the different types of text . This document will review the state of the art of current emotion detection methods using different sources and in different contexts, identifying the different approaches used in the process and comparing them with the identified literature.

Finally, a proposal based on the text processing of distorted texts is presented, because this has not been researched in depth in the works previously presented.

Keywords: Emotion recognition, natural language processing, implicit detection, social media, text analysis, context.

1. Introducción

El uso de las redes sociales ha crecido de una manera imponente en los últimos años Leong2019. Las redes sociales son definidas como los sitios y las herramientas en línea que facilitan la interacción entre los diferentes usuarios permitiéndoles la oportunidad de intercambiar información, opiniones, intereses e incluso emociones. Dentro de los principales usos de las redes sociales se encuentran el entretenimiento, comunicación, intercambio de opiniones, búsqueda de información entre otros. Con el auge de estas plataformas sociales, una gran cantidad de horas diarias son invertidas de manera virtual en herramientas como Instagram, Facebook, Twitter, YouTube, WhatsApp entre muchas otras. Las generaciones actuales viven en un mundo donde lo primero que se ve en la mañana y lo último en la noche es un teléfono celular, por lo que de acuerdo con algunos autores sugieren que las redes sociales han cambiado la forma de la interacción entre individuos y el comportamiento de los usuarios de manera individual y colectiva en todo el mundo Dhir2018,Tateno2019.

El mayor uso de plataformas de redes sociales ayuda a incrementar el número de “conexiones” pueden generar diferentes tipos de emociones afectivas pudiendo llegar al punto de generar cierto tipo de confusión emocional. Existen dos tipos de lazos relacionales en las redes sociales, los denominados lazos fuertes, que son establecidos con personas allegadas de una manera muy personal e íntima, como parejas románticas, mejores amigos, familiares cercanos entre otros. [2] se refiere al segundo tipo de lazos denominados débiles que se encuentran en los círculos sociales con allegados, compañeros o conocidos. Según de la fortaleza de las relaciones entre los individuos, el uso de las redes sociales puede ser saludable o disfuncional. En ambos casos, estas relaciones afectan el bienestar de la persona brindando felicidad o placer al usuario en el mejor de los escenarios.

Existen diferentes investigaciones realizadas que indican que cuando se experimentan una fuerte cantidad de sentimientos positivos o negativos, hay una gran motivación al querer compartirlos con los demás Rime2009. De acuerdo con la teoría del intercambio social de emociones de Rimé, compartir las emociones ayuda en dos objetivos: la satisfacción de las necesidades socioemocionales que consiste en obtener apoyo de terceros para revalidar los sentimientos, normalizar las experiencias y sentir estados emocionales positivos. Y como segundo objetivo, se encuentran las necesidades cognitivas, que implica recibir consejos de los demás para ayudar a dar sentido a las experiencias vividas.

Las redes sociales les brindan a las personas oportunidades nunca antes vistas para lograr estos dos objetivos Brady2020,Rime2020 permitiendo una

conexión inmediata y en tiempo real con las personas cercanas, pero también con aquellas que se encuentran a una distancia geográfica considerable. Este tipo de interacciones generan al usuario un sistema de retroalimentación que satisface las necesidades socioemocionales y cognitivas incrementando los niveles de apoyo social percibido y recibido. En algunos casos, el apoyo social que se consigue en una red social es percibido de manera ampliamente superior en el mundo virtual que en el mundo real. Algunos estudios han evidenciado que ciertas personas que sufren de depresión reciben menos apoyo de otras personas que no sufren de esta condición en el mundo fuera de línea y han recibido más apoyo de las personas mediante la interacción en redes sociales Park2016, Starr2008.

De la misma manera que las redes sociales brindan nuevas oportunidades para buscar y brindar apoyo, también permite expresar los sentimientos de formas que pueden lastimar a otras personas. Las redes sociales son plataformas que facilitan en gran medida el ciberacoso y el matoneo, comportamientos antisociales que impactan de manera negativa al bienestar de las demás personas. Aunque las diferencias juegan un papel importante en quienes se involucran en este tipo de actitudes, hay características de las redes sociales que promueven este comportamiento; por ejemplo, al no contar con un contacto cara a cara, es difícil determinar mediante un medio electrónico la reacción física de una persona ante un comentario o acción y que permita limitar un comportamiento que pueda llegar a ser agresivo Lapidot-Lefler2012, Lieberman2020a, Suler2004. De igual manera la facilidad y accesibilidad a los dispositivos para acceder a una red social, permiten que sea más rápido compartir las emociones que se sienten cuando están en un punto máximo y son comunicadas sin detenerse a pensar en las posibles consecuencias de los comentarios precipitados.

Las redes sociales también desempeñan un papel vital en la difusión de sentimientos de indignación moral, lo cual contribuye a la deshumanización de otras personas y reducir acciones que se realicen de manera colectiva y aumentar las brechas existentes entre diversos puntos de opinión o características sociales. Se evidenció en estudios que ciertas personas son más propensas a realizar comentarios de manera intensa sobre indignación moral en línea que estando fuera de ella Turel2018.

Estas redes en general demuestran ser una plataforma que permite satisfacer el deseo de compartir emociones con las demás personas algunas veces con resultados positivos y de bienestar que generan nuevas oportunidades para que las personas puedan tener un apoyo emocional y superar ciertas circunstancias en las que se puedan encontrar; pero también pueden ser promotoras de acciones negativas y ser una plataforma masificadora de ciberacoso, *trolleo* y difusión masiva de odio.

A continuación presentaremos los trabajos actuales que se han realizado, partiendo de modelos para representar emociones (sección 2), recursos utilizados en el reconocimiento de emociones (sección 3), posteriormente presentamos las técnicas utilizadas en el reconocimiento de emociones en textos (sección 4). Para concluir presentamos un esbozo de lo que hemos encontrado como uno de los múltiples casos que no han sido atacados dentro de este problema en la sección 6.

2. Modelos para representar emociones

Los modelos de emociones son fundamentales para los sistemas detectores de las mismas ya que definen cómo pueden ser representadas. Estos modelos afirman que las emociones pueden presentarse en varios estados por lo cual nace la necesidad de distinguir cada uno de ellos; por ello cuando se realiza cualquier tipo de actividad concerniente a la detección de emociones es importante realizar la definición del modelo a utilizar. Existen diferentes modelos que representan emociones, pero se hará énfasis en los modelos de emociones discretos y dimensionales.

2.1. Modelos de emociones discretos (*Discrete emotion Models - DEMs*)

Clasifican las emociones en diferentes categorías, donde se toman los siguientes tres modelos a consideración:

Modelo de Paul Ekman Kowalska²⁰¹⁷. Realiza diferenciación de emociones con un conjunto de seis categorías básicas. En su modelo afirma que existen seis emociones originadas desde diferentes sistemas neuronales como el resultado de cómo se percibe una situación, por ello son independientes. Estas emociones son ira, felicidad, sorpresa, tristeza, temor y desagrado. Adicionalmente, estas emociones en conjunto pueden generar otras más complejas como la vergüenza, orgullo, lujuria, culpa, avaricia, deseo, celos entre otras.

Modelo de Plutchik y Kellerman Kellerman¹⁹⁸⁹. Afirma también que existen emociones primarias las cuales ocurren en pares opuestos y se encargan de producir emociones más complejas con sus combinaciones. Nombra un total de ocho emociones primarias en pares fundamentales las cuales son alegría frente a tristeza, confianza frente a disgusto, ira frente a miedo y sorpresa frente a anticipación. De acuerdo con Plutchik, para cada una de las emociones existen varios grados de intensidad que ocurren de acuerdo como los eventos son contruidos por un individuo.

Modelo de Orthony, Clore y Collins (*OCC Model*) Colby¹⁹⁸⁹. Afirman que las emociones son el resultado de cómo un individuo percibe los eventos y esas emociones varían de acuerdo con el grado de intensidad en el que son percibidas. Realizan la diferenciación de 22 emociones definiendo su modelo de la siguiente manera: ira, felicidad, sorpresa, tristeza, temor, desagrado, envidia, desagrado, agrado, dolor, alivio, envidia, reproche, autorreproche, aprecio, vergüenza, lástima, admiración, decepción, esperanza, miedos reafirmados, agradecimiento.

2.2. Modelos de emociones dimensionales (*Dimensional emotion models DiEMs*)

El modelo dimensional supone que las emociones no son independientes unas de las otras y que existe una relación entre ellas, por lo que surge la necesidad de colocarlas de manera conjunta en un espacio dimensional que represente qué tan relacionadas están las emociones y realizando una representación de los dos

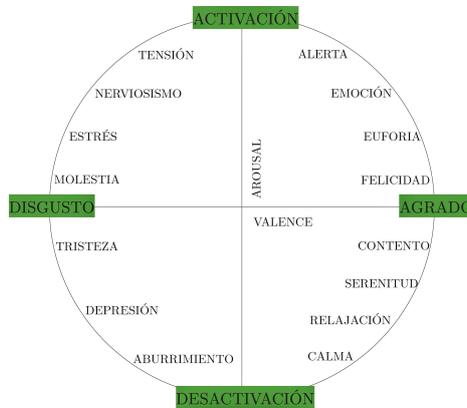


Fig. 1. Modelo de emociones de Russell. adaptado de Russell1980.

estados conductuales básicos fundamentales: bien y mal. Los siguientes son los dos modelos básicos dimensionales que se utilizarán:

El modelo de Russell (1980) consiste en una estructura de dos dimensiones denominada circunferencia de afecto. El modelo define dos valores de medida, *valence* que diferencia las emociones en un rango de positiva a negativa, y *arousal* la diferencia en un contexto de emoción o apatía. Esta relación se muestra en la figura 1.

El modelo de Plutchik presenta las emociones en un disco bidimensional con *valence* en el eje vertical y *arousal* en el horizontal. La figura muestra las emociones en capas, donde las más internas corresponden a derivaciones de las ocho emociones primarias, seguidas precisamente por estas emociones primarias y en la parte exterior combinaciones de las emociones primarias. Según la posición en la figura, se determina qué tan relacionadas se encuentran las emociones. En la figura 2 se presenta el esquema de Plutchik.

Según el problema a resolver se realiza la selección del modelo emocional a trabajar. Los modelos de emoción discretos usualmente son los seleccionados para los problemas de clasificación de emociones debido a la simplicidad, pero carecen de una gama más amplia de clase de emociones, la intensidad y grado de ocurrencia en contraste con los modelos dimensionales. Pero los modelos dimensionales son recomendables cuando se desea categorizar una mayor cantidad de emociones y se quiere diferenciar entre emociones que son muy similares entre sí.

3. Recursos utilizados en el reconocimiento de emociones

Comúnmente en el reconocimiento de emociones se utilizan dos tipos de conjunto de datos, los *corpora* que corresponden al conjunto de textos completos clasificados según la emoción a la que se encuentra asociado cada uno de ellos,

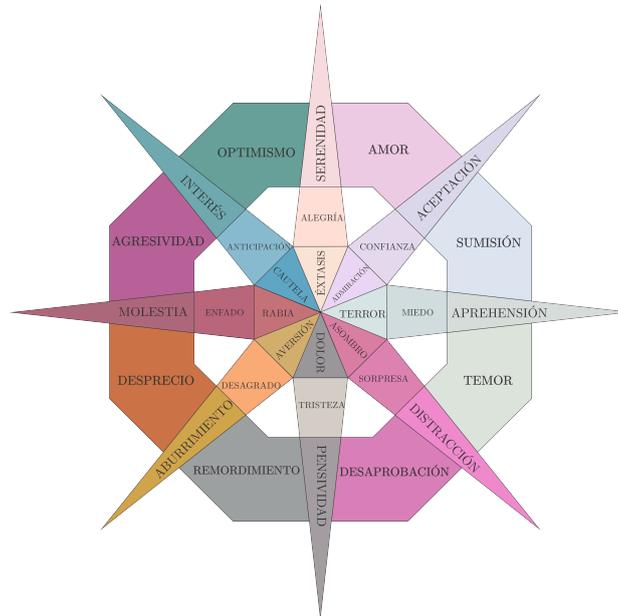


Fig. 2. Modelo de emociones de Plutchik. Adaptado de Kellerman1989.

y los *lexicons* que corresponden a un conjunto de listas ordenadas con palabras asociadas cada una de ellas a una emoción.

3.1. Corpora

[1]¹ presentan un total de 185 historias infantiles donde la emoción se clasifica a nivel de las oraciones que se encuentran en cada una de las historias y se asigna una etiqueta a cada una de las frases de acuerdo con el cuadro 1.

Para la clasificación se realiza la asignación en pares de las mismas historias y con el objetivo de evitar sesgos cada uno de los evaluadores trabaja de manera separada y realiza la asignación de una sigla de manera independiente. Cuando existe un desacuerdo en la etiqueta asignada, el evaluador principal elige una de las etiquetas disputantes.

El corpus de Aman2007 cuenta con un conjunto de publicaciones en diferentes blogs que son obtenidas mediante el uso de palabras semilla para asignar a cada emoción. Por ejemplo, palabras como miedo, susto, o pánico son clasificados en la categoría de temor. La clasificación se realiza a nivel de oración en una de las ocho categorías clasificadoras (felicidad, tristeza, enojo, desagrado, sorpresa, temor, emociones combinadas y sin emoción).

¹ people.rc.rit.edu/~coagla/affectdata/index.html

Tabla 1. Emociones básicas usadas en la clasificación.

Sigla	Clase de emoción
A	Enojo
D	Disgusto
F	Temor
H	Felicidad
Sa	Tristeza
Su+	Sorpresa positiva
Su-	Sorpresa negativa

[3]² describen el resultado de la aplicación de una encuesta internacional de antecedentes de emociones y reacciones (Internations Survey on Emotion Antecedents and Reactions – ISEAR); donde un selecto grupo de sicólogos de todo el mundo trabajaron en este proyecto y hubo una participación aproximada de 3000 estudiantes indicando situaciones del día a día en las que han experimentado emociones de alegría, miedo, temor, tristeza, desagrado, vergüenza y culpa.

En SemEval-2007 Strapparava2007³ se cuenta con titulares de noticias obtenidos de diferentes ejemplares, como BBC, CNN, New York Times y el motor de búsqueda de noticias de Google. La estructura con la que cuenta un titular de noticias permite realizar la clasificación a nivel de oraciones. Cada una es clasificada de acuerdo con una o más de las siguientes emociones: enojo, disgusto, temor, alegría, tristeza y sorpresa. La implementación cuenta con dos conjuntos de datos uno para entrenamiento con 250 titulares, y uno de prueba con un total de 1000 titulares.

SemEval-2018 Mohammad2018⁴ cuenta con la recopilación de tuits, donde se clasifican de tipo neutral o si expresan una o más emociones dentro de las que se encuentran enojo, desagrado, temor, alegría, amor, optimismo, pesimismo, tristeza, sorpresa y confianza. De igual manera existe diferenciación entre los conjuntos de entrenamiento. Los tuits fueron obtenidos en tres idiomas: inglés, árabe y español.

SemEval-2019 Chatterjee2019⁵ presenta diálogos textuales entre dos personas donde el primero de ellos es quien inicia las conversaciones, luego el segundo responde y de nuevo el primero vuelve conversar en un esquema de turnos. Cada una de las conversaciones es etiquetada de acuerdo con la emoción que expresa; alegría, enojo, tristeza entre otras. La clasificación de la emoción es realizada con base en la interacción en el tercer turno de la conversación.

SemEval-2020 Patwa2020⁶ aborda la clasificación de tuits en tres grupos principales de emociones: positivos que expresan felicidad, exaltar a una persona,

² github.com/PoorvaRane/Emotion-Detector/blob/master/ISEAR.csv

³ web.eecs.umich.edu/~mihalcea/affectivetext/\#resources

⁴ competitions.codalab.org/competitions/17751

⁵ www.humanizing-ai.com/emocontext.html

⁶ paperswithcode.com/paper/semEval-2020-task-9-overview-of-sentiment

grupo, país o celebrar algo. Negativos que expresan ataques a alguna persona, grupo, producto, o país, también disgusto, o tristeza hacia algo y crítica. Y neutrales que corresponden a noticias o publicidad.

En el cuadro 4 se realiza la clasificación de los conjuntos de datos especificados indicando el número de emociones según el tipo de emoción detectado y el número de ocurrencias de cada una de las emociones encontradas.

3.2. Lexicones

Existen diversos lexicones, también conocidos como diccionarios, que son utilizados comúnmente para el análisis de emociones. A continuación se describen los recursos existentes.

WordNet⁷ se trata de una base de datos léxica en inglés en línea. Agrupa verbos, sustantivos, adjetivos, y adverbios en conjuntos de sinónimos denominados *synsets*.

WordNet-Affect⁸ es una extensión de WordNet; un subconjunto de los synsets de éste con palabras que expresan emociones directas o indirectas, se utilizan etiquetas semánticas para su asignación.

SentiWordNet⁹ asigna una de los tres posibles valores de clasificación positivos, negativos, u objetivos a cada uno de los synsets de WordNet

AFINN¹⁰ consta de palabras calificadas de manera manual para el valor valence con un número entero entre menos cinco (negativo) y más cinco (positivo).

NRC Word-Emotion Association Lexicon¹¹ fue creado de manera manual utilizando el sistema de turcos mecánicos de Amazon. Ocho emociones, que son enojo, anticipación, disgusto, miedo, alegría, tristeza, sorpresa y confianza, y las categorías de sentimientos, positivos y negativos están incluidos.

NRC Affect Intensity Lexicon¹² proporciona puntuaciones de intensidad con valores reales por las emociones enojo, miedo, tristeza y alegría.

NCR Valence, Arousal y Dominance Lexicon¹³ incluye una lista de más de 20.000 palabras y sus valores de *valence*, *arousal* y *dominance*. Las puntuaciones van de 0 a 1.

NRC Hashtag Emotion Lexicon¹⁴ es un lexicón generado de manera automática a partir de tuits que incluyen *etiquetas* de palabras de emoción, como #happy. Asocia las palabras con las emociones ira, disgusto, miedo, tristeza, anticipación, sorpresa, alegría y confianza.

⁷ wordnet.princeton.edu/download/current-version

⁸ multiwordnet.fbk.eu/english/home.php

⁹ github.com/aesuli/SentiWordNet

¹⁰ github.com/fnielsen/afinn

¹¹ saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm

¹² www.saifmohammad.com/WebPages/AffectIntensity.htm

¹³ saifmohammad.com/WebPages/nrc-vad.html

¹⁴ saifmohammad.com/WebPages/lexicons.html\#NRCTwitter

NRC Hashtag Sentiment Lexicon¹⁵ fue creado de manera automática a partir de tuits que incluyen etiquetas de palabras de sentimiento como **#amazing**. Asocia palabras con un sentimiento positivo o negativo.

Sentiment140 Lexicon¹⁶ fue generado también automáticamente a partir de tuits con emoticonos.

Finalmente, Spanish Emotion Lexicon (SEL)Sidorov2013¹⁷ Contiene un total de 2036 palabras las cuales se encuentran asociadas con el factor de probabilidad de uso afectivo (FPA) con respecto a por lo menos una de las siguientes emociones básicas: alegría, ira, miedo, tristeza, sorpresa y disgusto. Rangel2014

4. Técnicas utilizadas en el reconocimiento de emociones en textos

En trabajos previos realizados normalmente se realiza la distinción entre cuatro enfoques utilizados para reconocer las emociones en el texto que son los basados en palabras clave, basados en reglas, basados en el aprendizaje y basados en aprendizaje profundo. Los artículos se clasifican con base en el enfoque propuesto, el uso de dicha clasificación ayudará a evaluar estos enfoques en función de su desempeño, fortalezas y limitaciones y establecer una comparación entre ellos. Las emociones explícitas son reconocidas principalmente con enfoques basados en palabras clave. Los otros tres se utilizan principalmente para reconocer las emociones implícitas en el texto, a pesar de que también han sido utilizados para el reconocimiento de emociones explícitas.

4.1. Enfoque basado en palabras clave

Este enfoque realiza la búsqueda de ocurrencias de palabras claves en un texto dado y realiza la asignación de una etiqueta de emoción de acuerdo con la palabra clave que ha sido detectada. La aproximación más utilizada es la técnica denominada *keyword-spotting*. Como inicio en el proceso, se realiza la definición de una lista de palabras para el etiquetado de las clases mediante el uso de los léxicos. Luego al texto se le realiza un preprocesamiento, que consiste en realizar la tokenización, eliminación de palabras gramaticales y su posterior lematización. Luego se procede a detectar las palabras clave de la emoción que se encuentran en el texto y compararlas contra el léxico utilizado. Posteriormente se analiza la intensidad de la emoción, se verifica si la emoción detectada corresponde a una negación y finalmente se realiza la asignación de la etiqueta de emoción para cada una de las oraciones en el texto. En la figura 3 se presenta el esquema general del enfoque en palabras clave.

¹⁵ saifmohammad.com/WebPages/lexicons.html\#NRCTwitter

¹⁶ help.sentiment140.com/for-students

¹⁷ <https://www.cic.ipn.mx/~sidorov/>

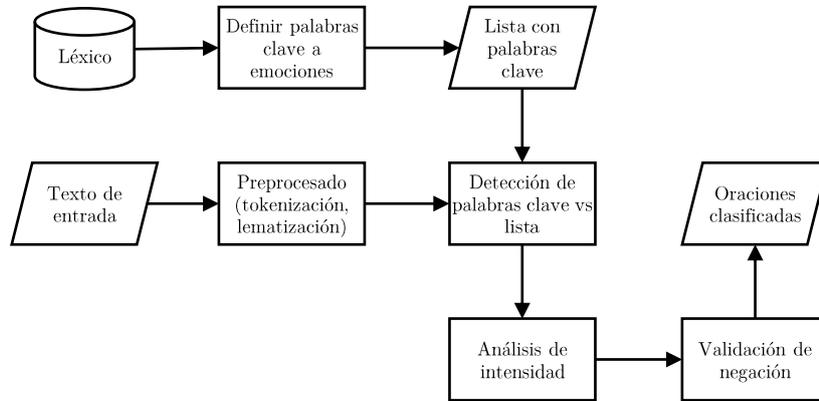


Fig. 3. Enfoque basado en palabras clave.

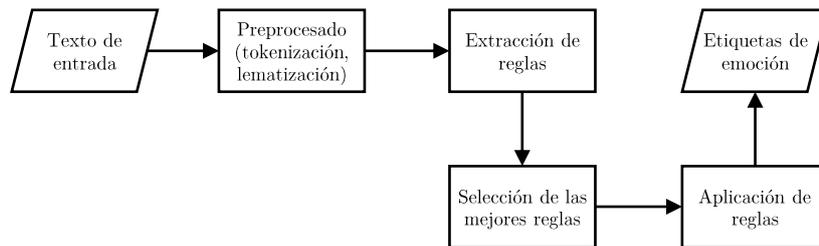


Fig. 4. Enfoque basado en reglas.

4.2. Enfoque basado en reglas

Se basa en la manipulación de diferentes fuentes de conocimiento para realizar la interpretación de la información de entrada y realizar su correspondiente clasificación. En primer lugar, se realiza el preprocesamiento del texto de entrada, donde se aplica tokenización, eliminación de palabras gramaticales, lematización, POS tagging y análisis de dependencias. Luego, las reglas de emoción son extraídas utilizando conceptos lingüísticos, estadísticos y computacionales. Las mejores reglas son seleccionadas y finalmente son aplicadas al texto de entrada para determinar cada una de las etiquetas de emoción. En la figura 3 se representa el enfoque basado en reglas.

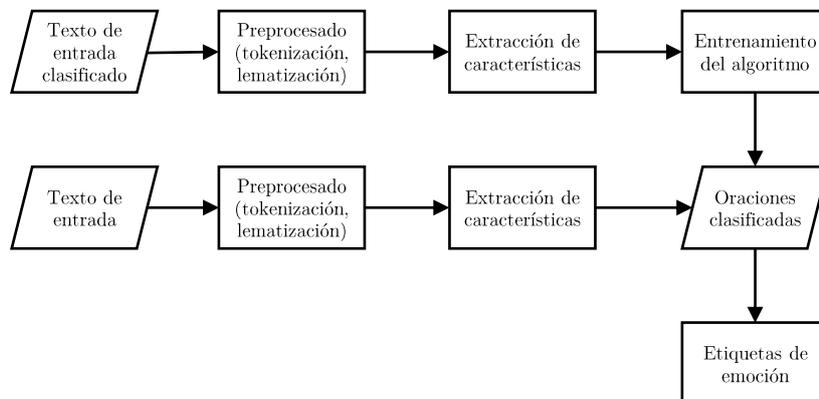


Fig. 5. Enfoque basado en aprendizaje.

4.3. Enfoque basado en aprendizaje

Este enfoque proporciona al sistema clasificador la habilidad de aprender de manera automática y mejorar su desempeño con base en la experiencia adquirida en el proceso. Los algoritmos de *Machine Learning* son usualmente clasificados con tipo de aprendizaje supervisado y no supervisado. Uno de los algoritmos más utilizados en la clasificación de emociones en textos es el algoritmo de Máquinas de Vectores de Soporte por sus siglas en inglés SVM el cual es un algoritmo del tipo supervisado. En este algoritmo el funcionamiento consiste en igual manera de iniciar con el preprocesamiento del texto donde se realiza la tokenización, remoción de palabras gramaticales, lematización y POS tagging. El siguiente paso es extraer funciones útiles y se seleccionan aquellas con la mayor ganancia de información. Dado el conjunto de características y las etiquetas de emoción, el algoritmo SVM genera un hiperplano óptimo. Finalmente, el modelo SVM entrenado se utiliza para clasificar emociones en el texto de entrada para realizar la clasificación. En la figura 5 se representa el enfoque basado en aprendizaje.

4.4. Enfoque basado en aprendizaje profundo

El aprendizaje profundo es una rama del *Machine Learning* en donde los diferentes programas implementados manejan el concepto de aprender de la experiencia donde el mundo es interpretado en conceptos de jerarquía de conceptos; realizando la definición de cada uno de los conceptos en términos de la relación con conceptos más sencillos. Esta aproximación permite que un programa adquiera conceptos complicados realizando su construcción a partir de conceptos más simples Goodfellow2016. Dentro de los modelos de aprendizaje profundo más utilizados se encuentra el de *Long Short Term Memory* (LSTM), el cual consiste en una forma especial de red neuronal recurrente con la capacidad

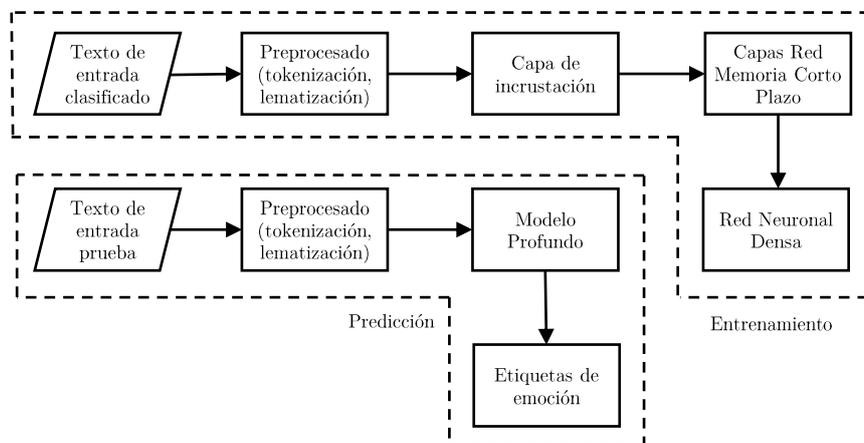


Fig. 6. Enfoque basado en aprendizaje profundo.

de manejar dependencias a largo plazo. Este método cuenta con la ventaja que supera el problema de desvanecimiento del gradiente. En este procedimiento se inicia con el preprocesamiento del texto, donde se realiza la tokenización, eliminación de palabras vacías y lematización. Después de eso, la capa de incrustación se construye y se alimenta a una o más capas de LSTM. Luego, la salida se alimenta a una red neuronal densa (DNN) con unidades iguales al número de etiquetas de emoción y una función de activación sigmoide para realizar la clasificación correspondiente. En la figura 6 se representa el enfoque basado en aprendizaje profundo.

4.5. Medidas de evaluación

Cada uno de los trabajos identificados en el estado del arte realizan la medición de su desempeño mediante diferentes técnicas, para así realizar la medición de la similitud entre los conjuntos pronosticados contra los datos reales y también identificar el nivel de precisión en el pronóstico. Algunas medidas identificadas en los documentos son las siguientes.

Índice de Jaccard:

$$\mathbf{IJ} = \frac{1}{|S|} \sum_{s \in S} \frac{|G_s \cap P_s|}{|G_s \cup P_s|}, \quad (1)$$

donde S es el conjunto de frases a clasificar, G_s el conjunto de todas las etiquetas disponibles para clasificación y P_s el conjunto de todas las etiquetas predecidas para la frase.

Precisión:

$$\mathbf{P}_e = \frac{VP_e}{VP_e + FP_e}, \quad (2)$$

donde e corresponde a la etiqueta de emoción, VP a Verdadero Positivo y FP a Falso Positivo.

Recall :

$$\mathbf{R}_e = \frac{VP_e}{VP_e + FN_e}, \quad (3)$$

donde e corresponde a la etiqueta de emoción, VP a Verdadero Positivo y FN a Falso Negativo.

F_{score} :

$$\mathbf{F}_1 = 2 \times \frac{P_e \times R_e}{P_e + R_e}, \quad (4)$$

P_e corresponde al valor de la precisión y R_e al valor del recall calculados previamente.

Accuracy :

$$\mathbf{A} = \frac{\sum_{e \in E} VP + \sum_{e \in E} VN}{\sum_{e \in E} VP + \sum_{e \in E} VN + \sum_{e \in E} FP + \sum_{e \in E} FN}, \quad (5)$$

donde E corresponde al conjunto de etiquetas de emoción, VP corresponde al número de verdaderos positivos, VN es el número de verdaderos negativos, FP es el número de Falsos Positivos y FN es el número de falsos negativos.

4.6. Resumen del estado del arte

En el cuadro 5 se presenta el resumen del estado del arte revisado realizando una comparación entre la bibliografía mediante una tabla donde se indica el artículo mencionado, el idioma de los conjuntos de datos utilizados, el enfoque utilizado, la herramienta y/o algoritmo utilizados en la implementación, el nombre del corpus utilizado y el desempeño del modelo implementado.

5. Trabajo propuesto

Hemos identificado diversas soluciones que se han presentado para resolver el problema de reconocimiento de emociones en distintos conjuntos de datos. Una parte importante que identificamos es que no se ha estudiado a profundidad la posible distorsión en las formas de escribir texto. Como ejemplo de ello, podemos considerar la escritura *Leet* o *1337 5p34k*. Ésta consiste en la combinación de caracteres alfabéticos y numéricos con el objetivo de reemplazar letras por dígitos e incluso símbolos que tienen parecido con la letra que se desea sustituir. Usualmente este tipo de escritura es utilizada para omitir los filtros de moderación de lenguaje en las redes sociales y principalmente expresar emociones

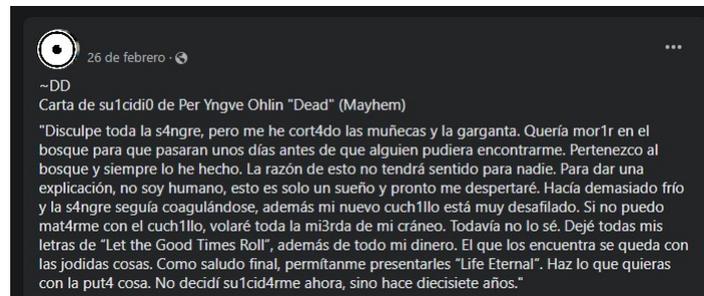


Fig. 7. Ejemplo de escritura *leet* en mensajes depresivos.

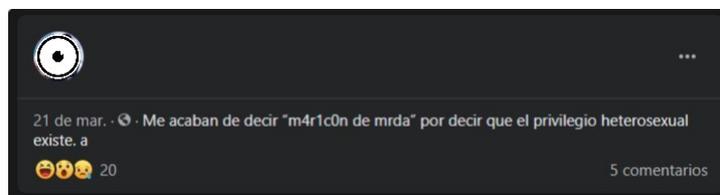


Fig. 8. Ejemplo de escritura *leet* en mensajes de odio.

radicales de odio o tristeza profunda como se puede evidenciar en las figuras 7 y 8. La identidad de los autores de las publicaciones fue eliminada de manera intencional para su protección.

En la documentación revisada y verificando en los conjuntos de datos utilizados, todos los textos tienen en común que son frases o escritos completos donde su gramática se encuentra configurada completamente con palabras y no tienen en cuenta este tipo de cambios en las letras, por lo que en los preprocesamientos de los textos antes de los métodos de entrenamiento, estas palabras pueden ser descartadas en el proceso de limpieza y cambiar la emoción que se transmite. Si se retoma la frase de la figura 8 y se descartan las palabras escritas en forma leet y las palabras con letras faltantes, la frase resultante es la siguiente:

"Me acaban de decir de por decir que el privilegio heterosexual existe"

Con esta simple observación se puede evidenciar que, del mensaje de odio agresivo identificado en el escrito original se convierte en uno que carece de emoción y podría ser clasificado con la etiqueta de emoción neutra lo cual obviamente es una mala asignación.

Varios de los usuarios de las redes sociales han identificado que escribir mensajes con este tipo de escritura no genera repercusiones en sus perfiles que usualmente consisten en el bloqueo de sus cuentas, lo cual ha incentivado que esta técnica se haya masificado en más de una de las redes sociales actuales. Realizando una búsqueda sencilla con la palabra "su1cidio" en Google, arroja un aproximado de 950,000 resultados y revisando algunos de ellos son bastante perturbadores. Por ello realizar el análisis de emociones en textos incluyendo

esta nueva variable, podría generar resultados más precisos en la clasificación de emociones extremistas que usualmente suelen ser las más peligrosas y que desencadenan en acciones que atentan contra la integridad física de las personas.

5.1. Algunas investigaciones en la detección de emociones negativas

Realizando la consulta a trabajos de investigación centrados en el análisis de emociones negativas, se realizó un análisis de las tecnologías implementadas para dicha tarea. En el trabajo de Villatoro-Tello2017 el objetivo principal consiste en realizar una predicción de manera anticipada de casos de depresión en las publicaciones realizadas por los usuarios. Como preprocesamiento del texto deciden por abandonar las técnicas comunes como las bolsas de palabras y realizar el modelamiento de los datos al igual que los modelos de predicción mediante el uso de grafos y clasificar las publicaciones en dos clases de resultados, si la persona que la escribió se encuentra deprimida o no deprimida. Las ejecuciones del modelo se realizaron con 5 configuraciones diferentes: LyRA donde el modelo es construido mediante los términos como palabras independientes. LyRB construye el modelo mediante modelos de 3-gramas. LyRC es una configuración híbrida generando un vector de $2 \times 4 \times N$ extraídas de las características de las dos configuraciones anteriores. LyRD es un método de ensamblaje conservador que utiliza las salidas de las tres configuraciones anteriores y realiza la asignación de deprimido si las tres primeras configuraciones coinciden en asignar dicha categoría. Finalmente, LyRE es similar a la configuración anterior, pero donde la clasificación se realiza con base en la clasificación de la mayoría de los tres modelos iniciales en un esquema 2 de 3.

En la investigación de Hosseini-Saravani2020 la meta consiste en medir manifestaciones conductuales asociadas a la depresión con un total de 21 categorías de síntomas y actitudes que de acuerdo con estudios clínicos son características que poseen aquellas personas con algún nivel de depresión. Cada categoría es graduada con 4 o 5 frases para validar con cual se puede sentir identificado el paciente y se asigna un rango de 0 a 3 para reflejar el rango de la severidad asociada a cada una de las categorías. Como método de clasificación se utilizó el método de *Naïve Bayes*.

Los siguientes dos trabajos ya se enfocan en tendencias más fuertes que corresponden a la identificación de tendencias suicidas en personas que ya pueden contar con un nivel bastante profundo de depresión. Tadesse2020 realiza la identificación de publicaciones con tendencias suicidas en la red social Reddit. La metodología implementada consistió en el diseño e implementación de un modelo híbrido de Redes Neuronales Convolucionales (CNN) y redes recurrentes del tipo *Long Short Term Memory* (LSTM), donde las entradas al sistema son procesadas por la capa LSTM y sus salidas son el insumo para la sección CNN. Como resultado del estudio se obtuvieron aquellas palabras que son muy frecuentes y presentan una alta ocurrencia en aquellas publicaciones con tendencias suicidas como los son "suicidio", "quiero morir", "morir maldita sea", "deseo suicida", "deseo morir", "me quiero ir." entre otras. La medida de *accuracy* para esta implementación fue del 91.7%.

Tabla 2. Tabla de equivalencia propuesta para vocales y algunas consonantes.

Letra	Alfanumérica	Alfabéticos	Caractéres especiales
A	4	ÁĂÄÅ	@
E	3	ÊË	€
I	1	Ï	i
O	0	ÔÕÖ	Ø
U	-	ÛÜ	-
L	1	ŁŁ	!£
R	-	ŔŖř	®
S	5	ŚŜŞȘ	§
T	7	Ť	-

Sawhney2020 de igual manera realiza el análisis de tendencias suicidas en la red social Twitter mediante un conjunto de tuits en idioma inglés, pero basando la recopilación de datos en un usuario en específico y considerando su comportamiento de publicaciones en un intervalo de tiempo para así poder detectar con mayor precisión si existe un patrón en cada una de las publicaciones y poder llegar así a considerar un punto de alerta máxima en la que el usuario pueda llegar a intentar a hacer algo en contra de su vida. Como metodología implementada se utilizaron transformadores entrenados previamente, en particular el denominado BERT. En una primera capa de este modelo realiza la tokenización de todas las publicaciones históricas para así realizar la creación de un vector emocional inicial. Para la capa subsiguiente se realizó la implementación de capas *Time-Aware Long-Short Term Memory* (T-LSTM) debido a que precisamente el comportamiento del usuario a analizar puede ser cambiante en los intervalos de tiempo por lo que este sistema resulta más eficaz para comparar datos anteriores en el tiempo con comportamientos actuales. La medida de *accuracy* para esta implementación fue del 85.1 %.

5.2. Propuesta inicial de trabajo

De acuerdo con el trabajo realizado por Tundis2021 en el que se pueden evidenciar varias formas en las que se realiza la representación de frases como combinación entre diferentes tipos de caracteres e idiomas. Para la presente propuesta se utilizará la categoría denominada *single row - single language* y como subcategorías se realizará uso de aquellas implementaciones alfanuméricas, mezcla de diferentes caracteres alfabéticos y el uso de caracteres especiales en la representación de letras.

Por ello, en el cuadro 2 se realiza una definición inicial para las vocales en cada una de las tres subcategorías y para algunas consonantes, la cual posteriormente debe ser modificada para incluir todas las letras del alfabeto.

Se realizó un ejercicio de prueba de concepto con las palabras "deprimirse", "d3pr1m1rse", "dep®imirse", "d3p®imirse" como insumo de entrada para que

Tabla 3. Resultado de evaluación de palabras con Stanza.

Palabra	Etiqueta		Lema 1	Etiqueta		Etiqueta
	Gramatical	Lematización		Gramatical	Lema 2	
				Lema 1	Lema 2	Lema 2
deprimirse	-	Si	Deprimir	Verbo	Él	Pronombre
d3pr1m1rse	Numeral	-	-	-	-	-
depⓂimirse	-	Si	dep<UNK>imir	Sustantivo	Él	Pronombre
d3pⓂimirse	Verbo	No	-	-	-	-

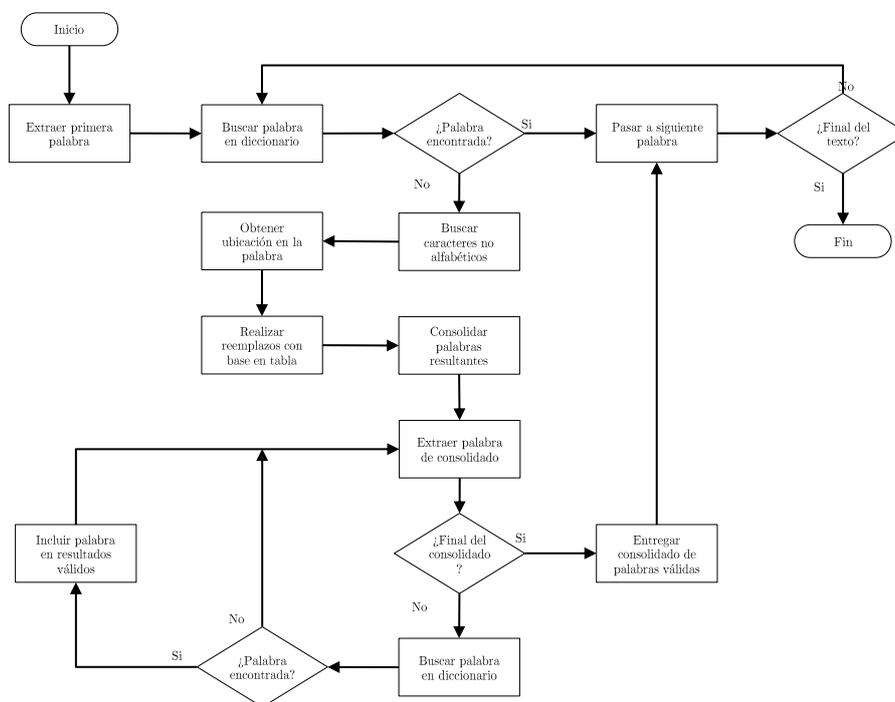


Fig. 9. Diagrama de flujo reglas de preprocesamiento.

fueran analizados por la librería de procesamiento de lenguaje natural Stanza para verificar cómo realiza la clasificación de estas palabras. El resultado se presenta en el cuadro 3 y se puede evidenciar que cada una de las cuatro palabras fue categorizada de manera diferente y para aquella que pudo ser lematizada la librería no pudo reconocer el carácter especial Ⓜ detectándolo como algo desconocido. Y en el último caso detectó la palabra como un verbo el cual no lematizó, lo cual también es inválido.

Tabla 4. Corpora para reconocimiento de emociones en textos.

Emoción	Alm et al	Aman	ISEAR	SemEval 2007		SemEval 2018		SemEval 2019		SemEval 2020	
				Entr	Prueba	Ent	Prueba	Ent	Prueba	Ent	Prueba
Alegría	445	536	1094	35	113	2477	1442	4243	284	-	-
Amor	-	-	-	-	-	700	516	-	-	-	-
Anticipación	-	-	-	-	-	978	425	-	-	-	-
Confianza	-	-	-	-	-	357	153	-	-	-	-
Culpa	-	-	1093	-	-	-	-	-	-	-	-
Disgusto	-	172	1096	9	12	2602	1099	-	-	-	-
Enojo	-	179	1096	20	21	2544	1101	5506	298	-	-
Enojo-Disgusto	218	-	-	-	-	-	-	-	-	-	-
Miedo	166	115	1095	32	92	1242	485	-	-	-	-
Neutral	-	2800	-	146	650	2207	75	-	-	3974	206
Optimismo	-	-	-	-	-	1984	1143	-	-	6005	3061
Otros	-	-	-	-	-	-	-	14948	4677	-	-
Pesimismo	-	-	-	-	-	795	375	-	-	2023	522
Sorpresa	114	115	-	8	42	361	170	-	-	-	-
Tristeza	264	173	1096	41	104	2008	960	5463	250	-	-
Vergüenza	-	-	1096	-	-	-	-	-	-	-	-
Total	1207	4090	7666	291	1034	18255	7944	30160	5509	12002	3789

Con base en el análisis realizado, se realiza como propuesta inicial la secuencia de reglas presentada en la figura 9 como un proceso previo para el preprocesamiento del texto con las características mencionadas y cuyo resultado ya podría pasar a la etapa de lematización para el proceso posterior de la detección de emociones en el texto.

En el mejor de los escenarios como palabras resultantes del proceso se va a generar una sola palabra válida. para el ejemplo de la palabra "d3p®imirse" se realizaría el reemplazo del valor 3 por E y ® por R de acuerdo con lo definido en el cuadro 2. Pero si se realiza una modificación y ahora la palabra a procesar es cop|a, y al realizar la transformación es posible tener dos resultados en donde el caracter | puede ser reemplazado por L o I obteniendo como resultados las palabras copia y copla las cuales son perfectamente válidas y seguramente van a existir en los diccionarios. Por ello el análisis del contexto será un factor determinante al momento de seleccionar la palabra final que en este caso reemplazará a cop|a.

6. Conclusiones

Los usuarios de las redes sociales se han dado cuenta que las publicaciones realizadas en las diferentes redes sociales empiezan a contar con una mayor restricción donde estas son revisadas y con base en la detección de palabras de alerta pueden advertir que la publicación puede atentar contra las normas de la comunidad de la plataforma y que la red social se encuentra en la potestad de tomar las medidas necesarias para evitar este tipo de comportamientos. Debido a esto los usuarios empiezan a recurrir a nuevas técnicas para realizar sus publicaciones con métodos de ofuscación en las palabras con varios niveles de complejidad según los conocimientos con los que posea el usuario para que así los filtros de restricción no generen ningún tipo de alertamiento y la publicación sea aceptada sin ningún inconveniente. Como se pudo evidenciar en la prueba de

Tabla 5. Resumen de trabajos sobre reconocimiento de emociones en textos en las redes sociales.

Referencia	Idioma	Método	Herramienta - Algoritmo	Dataset	Desempeño	
					Medida	Valor (%)
Bandhakavi2017	Inglés	Aprendizaje clásico	Máquinas de vectores de soporte	SemEval-07 Twitter Blogs ISEAR	F-Score	40.99
						49.55
						58.32
						31.21
Yasmina2016	Inglés	Aprendizaje clásico	Máquinas de vectores de soporte WEKA	Comentarios en YouTube	Presición	92.75
					Recall	72.72
					Accuracy	68.82
Wu2018	Inglés	Aprendizaje Profundo	Redes Neuronales Convolucionales	SemEval2018	Pearson	81.00
Tao2020	Chino	Aprendizaje Profundo	Redes de Memoria a corto plazo LDA (Latent Dirichlet Allocation) Redes de Memoria a corto plazo	Reseñas Hoteles	Presición	88.47
					Recall	79.88
Ghanbari-Adivi2019	Inglés	Híbrido: Reglas y Aprendizaje clásico	Máquinas de vectores de soporte Vecino más cercano Perceptrón multicapa Árboles de decisión	ISEAR	Accuracy	87.51
					Presición	82.00
					Recall	92.88
					Accuracy	88.49
					F-Score	81.00
Sailunaz2019	Inglés	Aprendizaje clásico	Naive Bayes Máquinas de vectores de soporte Bosques aleatorios	ISEAR Twitter	Presición	82.00
					Accuracy	66.86
Ibrahiem2020	Inglés	Aprendizaje clásico	Clasificación binaria Redes Neuronales Convolucionales	SemEval2018 NRC Emotion NRC Hashtag	Presición	63.00
					Recall	82.00
Baziotis2018	Inglés	Aprendizaje Profundo	Clasificación binaria Redes Neuronales Recurrentes Redes de Memoria a Corto Plazo	SemEval2017	NRC Hashtag	F-Score 69.00
					Jaccard	57.88
Ge2019	Inglés	Aprendizaje Profundo	Redes Neuronales Convolucionales	SemEval2019	Presición	75.98
					Recall	74.88
					F-Score	75.42
Patwa2020	Hinglish Spanglish	Aprendizaje Profundo	Redes Neuronales Convolucionales	SemEval2020	F-Score	80.6

concepto, realizar la modificación de una palabra que puede generar un alto nivel de alerta cambia la forma en que esta es interpretada por las herramientas de procesamiento de lenguaje natural actuales, lo cual ocasiona que la emoción real a transmitir llegue a ser clasificada de manera errónea con el sentimiento real que se desea transmitir y esto conlleva a que no se puedan prevenir situaciones de riesgo en la vida de las personas, por lo que el estudio de investigación propuesto presenta un punto de vista antes no investigado e innovador y que está adquiriendo relevancia debido a que las redes sociales en los últimos años realizan un mayor control sobre aquello que publican los usuarios en sus plataformas.

7. Trabajo futuro

Como trabajo futuro, se debe realizar una muy amplia, cuidadosa y profunda definición de la tabla de equivalencias en la sustitución de cada uno de los diferentes caracteres especiales debido a que su definición y la aplicación de las reglas sobre ésta consistirá en el factor principal en la cantidad de palabras válidas que saldrán como resultado del algoritmo de preprocesamiento en los textos. Adicionalmente, también de acuerdo con esos resultados, se deben analizar y finalmente seleccionar la mejor herramienta tecnológica para realizar la clasificación de cada uno de los textos.

Referencias

1. Alm, C. O., Roth, D., Sproat, R.: Emotions from text: Machine learning for text-based emotion prediction. HLT/EMNLP 2005 - Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference, , no. October, pp. 579–586 (2005)
2. Haythornthwaite, C.: Social networks and internet connectivity effects. *Information, Communication & Society*, vol. 8, no. 2, pp. 37–41 (2005)
3. Scherer, K. R., Wallbott, H. G.: Evidence for universality and Cultrual Variation, vol. 66, no. 2, pp. 310–328 (1994)