ISSN 1870-4069

# Aprendizaje automático para la detección de la depresión en redes sociales

Alma Partida-Herrera, Geovani Peña-Ramírez,
Eduardo Vázquez-Fernández, Arturo Pérez-Cebreros

Instituto Politécnico Nacional,
ESIME Culhuacán,
México

{partidaherreraalma, geovpe, perezcebreros}@gmail.com
eduardovf@hotmail.com

**Resumen.** La depresión es un problema de importancia pública que ahora se prioriza en muchas agendas de atención médica con el objetivo de prevenir futuros suicidios. Esto tiene un impacto devastador no sólo por la trágica pérdida de vidas, sino también por los familiares y amigos en duelo. Las investigaciones de cada país revelan una reducción del bienestar físico y mental, por ello la propuesta presentada en este artículo pretende detectar el sentimiento de enunciados de texto mencionados en redes sociales. En particular, nosotros examinamos los tuits mediante un clasificador bayesiano y máquinas de soporte vectorial lo que nos permite dar un paso hacia adelante para identificar el estado de salud emocional.

**Palabras clave:** Redes sociales, depresión, aprendizaje automático.

## Machine Learning for the Detection of Depression in Social Networks

**Abstract.** Depression is a problem of public importance that is now prioritized in many health care agendas with the goal of preventing future suicides. This has a devastating impact not only for the tragic loss of life, but also for bereaved family and friends. The investigations of each country reveal a reduction in physical and mental well-being, for this reason the proposal presented in this article aims to detect the feeling of text statements mentioned in social networks. In particular, we examine the tweets using a Bayesian classifier and support vector machines, which allows us to take a step forward to identify the state of emotional health.

**Keywords:** Social networks, depression, machine learning.

## 1. Introducción

Hoy en día, el mundo está viviendo momentos de transformaciones, la vida diaria ha tenido un giro de 360 grados en donde el protagonista ha sido una cepa viral

denominada SARS-CoV-2, que ha causado hasta el día un poco más de cuatro millones de muertes. Más allá de las consecuencias económicas, el confinamiento social suele ser una experiencia desagradable, que puede llevar a distintos estresores que generen afecciones de la salud mental [1].

Debido a que esta situación es nueva y se encuentra en plena expansión, es aún prematuro estimar las consecuencias emocionales del brote epidémico. Sin embargo, las investigaciones realizadas en [2,3] apuntan a que el miedo a lo desconocido y la incertidumbre pueden llevar a evolucionar distintas enfermedades de salud mental como: los trastornos de estrés, ansiedad, depresión, somatización y conductas que degeneran en aumento de consumo de alcohol, tabaco y otras substancias nocivas para la salud [4].

En particular, se espera que las personas con enfermedades crónicas presenten niveles más altos de síntomas psicológicos [5]. También las personas mayores se pronostican que sean psicológicamente más vulnerables que los jóvenes en esta crisis [6]. Este proyecto surge debido a la gran problemática sobre casos de suicidio en nuestro país en jóvenes [7, 8, 9] por ello se decide desarrollar una herramienta que sea capaz de alertar sobre posibles casos y que permita evitarlos.

### 1.1. La depresión y la inteligencia artificial

Cabe destacar que, en México, se encontró que adultos jóvenes (es decir, entre 15 y 25 años) presentan ideas suicidas y muestran mayores estados depresivos, es decir, la depresión aparece en el 67.3% de quienes han intentado suicidarse y en el 81.1% de quienes manifiestan ideas suicidas [10]. Y las personas con enfermedades mentales tienden a revelar su condición mental en las redes sociales, como una forma de alivio [11].

Sin embargo, la investigación sobre el aprovechamiento de redes sociales para comprender trastornos de la salud del comportamiento aún está en su infancia. En Katikalapudi et al. [12] analizaron patrones de actividad web de estudiantes universitarios que podrían indicar depresión. De manera similar, en Katie et al. [13] demostraron que las actualizaciones de estado en Facebook podrían revelar síntomas de episodios depresivos.

Aunque algunas diferencias han sido observadas, como que los usuarios deprimidos usan con más frecuencia pronombres en primera persona [14] así como palabras de emociones negativas e ira. Por ello, la depresión ha sido asociada al uso de marcadores lingüísticos tales como el uso elevado de pronombres de primera persona.

Muchos otros estudios del lenguaje y la depresión se han limitado a entornos clínicos, y, por lo tanto, a analizar discursos espontáneos o ensayos escritos. En esa dirección, algunas investigaciones [15, 16] propusieron metodologías innovadoras para recopilar contenidos textuales compartidos por personas diagnosticadas con depresión. Sin embargo, no hay colecciones disponibles públicamente.

Esto se debe a que a menudo el texto se extrae de sitios de redes sociales, como Twitter o Facebook que no permiten la redistribución [17]. De ahí que estos estudios previos, nos impulsen a la detección de la depresión en las redes sociales como primer paso contra el suicidio. El punto central de los estudios de la salud mental en redes sociales ha sido, tradicionalmente, llevado a cabo mediante el uso de encuestas. En

donde el número de usuarios está limitado por aquellos que puedan completar la encuesta. Por ejemplo, en Choudhury et al. [18] solicitó a los usuarios de Twitter que hicieran la Escala de Depresión del Centro de Estudios Epidemiológicos (CES-D) y compartieran su perfil al público.

Este tipo de estudios ha producido datos de alta calidad, sin embargo, está limitado en tamaño y alcance. Por ello, en esta investigación examinaremos la depresión considerando muestras obtenidas automáticamente, de grandes cantidades de datos de Twitter. El Internet ha permitido seguir la evolución del lenguaje y nos está proporcionando un medio muy accesible para que las personas expresen sus sentimientos de forma anónima.

De ahí que, nosotros hemos adaptado el método en Coppersmith et al. [15] para la construcción de este conjunto de datos en español, procederemos a identificar auto expresiones de diagnósticos de enfermedades mentales y aprovechamos estos mensajes para construir nuestro conjunto de datos.

### 1.2. Análisis de sentimientos

Entre las distintas plataformas de redes sociales, Twitter ha experimentado una adopción particularmente generalizada de usuarios; es una plataforma de microblogueo donde los usuarios crean tuits que se transmiten a sus seguidores o que son enviados a otro usuario.

A partir del 2016, Twitter tiene más de 313 millones de usuarios [19], y junto con este tremendo crecimiento, Twitter también ha sido objeto de muchas investigaciones de análisis de sentimientos (SA), ya que los tuits a menudo expresan la opinión de un usuario sobre un tema de interés. La palabra sentimiento se refiere a una forma de pensar (opinión) o sentir (emoción) sobre algo [20].

Originalmente, el SA apareció con la inteligencia de negocios [21, 23], pero se ha extendido a otras áreas como la política [23, 24], medicina [25], educación [26], recomendaciones [27, 28], detección de plagios [29], influencia en las noticias [30], detección del engaño [31, 32], detección de ironías [33], clasificación de cuentas [34], entre otras.

En particular, el SA es un tema destacado de investigación en el campo de la lingüística computacional. Las tareas incluyen la clasificación de la polaridad de sentimiento expresado en el texto (por ejemplo, positivo, negativo y neutral), identificación del objetivo/tema de sentimiento e identificación del sentimiento por varios aspectos de un tema.

El problema de clasificación de la polaridad de sentimiento es a menudo modelado como bidireccional (positivo/negativo) o tridireccional (positivo/negativo/neutral) [35]. Es importante resaltar que esta tarea de detección y clasificación no es sencilla, en primer lugar, debido a que los tuits son mensajes cortos en donde los indicadores de depresión suelen manifestarse de forma muy sutil.

A pesar de ello, algunos trabajos recientes han reportado resultados alentadores en la detección de usuarios que padecen de depresión, pero aún se requieren más estudios [36,37].

En nuestra investigación, modelaremos de manera bidireccional (positivo/negativo); dejando para un próximo trabajo, evaluar más intensidades del sentimiento de depresión: fuertes positivos, fuertes negativos, leves positivos, y leves

**Tabla 1.** Expresiones regulares para la detección de tuits.

| Palabra | Expresión regular |
|---|---|
| Depresión | (depresi[a-z]+) |
| Deprimido | (deprimi[a-z]+) |
| Frases asociadas | ((problema[s]\|disturbio(s)) *(mental\|psicologico(s)\|psiquiatrico(s))) (quiero) * (morir\|morir[a-z]+) (todo(s))*(dia(s)*(trist[a-z] + problema(s)) |

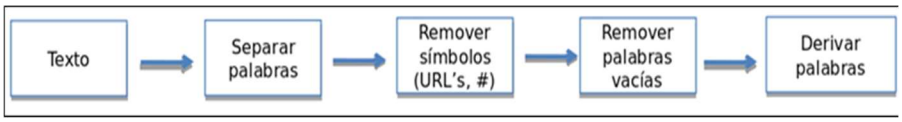


**Fig. 1.** Escenario del pre-procesamiento de tuits.

negativos. Detectar el sentimiento de las frases es una tarea complicada, por ejemplo: 'la vida es como el jazz, mejor si es improvisada'; el sentimiento de la opinión es positiva porque la palabra 'vida' implica algo bueno.

Sin embargo, la misma palabra en otro contexto, como se muestra en el siguiente enunciado: 'mi vida no tiene sentido', implica un sentimiento negativo – es malo porque la negación reduce lo positivo de la palabra 'vida'. Entonces el problema implica el uso del lenguaje, lo cual es un problema muy complejo y vasto.

## 2. Desarrollo

Para darle solución a este problema se propuso un modelo de tres fases.

### 2.1. Fase de recolección

Durante esta fase, se tomó ventaja de la gran cantidad de datos proporcionados por Twitter. El método de recolección se basa en dos pasos principales: primero, los tuits se filtran mediante expresiones regulares posteriormente son clasificados: en negativos y positivos.

Para adquirir los tuits para este estudio, desarrollamos una aplicación que utiliza la API de búsqueda de Twitter [38]. Para filtrar los tuits que no están escritos en español, utilizamos la biblioteca de detección de idiomas disponible gratuitamente [39]. Esta librería se basa en filtros bayesianos y tiene una precisión de 0.99 en la detección de los 53 idiomas que admite.

Los tuits se adquirieron durante 210 días (del 01 de diciembre del 2020 al 01 de julio del 2021), produciendo conjuntos de datos con aproximadamente 4000 tuits para español. Para generar el conjunto de datos de tuits con rasgos depresivos, nosotros consideramos tuits de personas que declararon haber sido diagnosticados con la enfermedad de la depresión.

En la Tabla 1, se muestran las expresiones regulares usadas para detectar a las personas que hacen referencia de la depresión en sus tuits; pero el principal objetivo es encontrar personas que hagan una declaración directa y abierta de que fueron

diagnosticados con la enfermedad de la depresión. Posteriormente, se procede a extraer los tuits de la lista de las personas que aseveraron, por medio de un tuit, tener dicha enfermedad.

## 2.2. Fase de preprocesamiento

El preprocesamiento de datos es un paso, a menudo, descuidado pero importante en el proceso. Implica técnicas para transformar los datos sin procesar en un formato más comprensible. Las principales son limpieza del dato, integración de datos, transformación de datos y reducción de datos.

Como podemos ver la Figura 1, nuestro mecanismo de preprocesamiento incluye:

a) Separar palabras del texto,
b) Eliminación de números y URLs que involucra un efecto sobre nuestro análisis, pero si reduce el ruido y nuestra eficiencia [40],
c) Eliminación de palabras vacías como artículos, pronombres, y preposiciones [41],
d) Derivación de las palabras, el cual se utiliza para transformar diferentes formas de palabras en una forma raíz estándar [42].

En esta fase, además de estas técnicas incorporamos un paso de ponderación mediante el algoritmo Term Frequency-Inverse Document (TF-IDF). El TF-IDF refleja la importancia de una palabra en un documento; y este nivel de importancia se incrementa cuando la palabra aparece muchas veces, al punto que podemos determinar temas de tendencia [43].

La Frecuencia de Términos (TF) es la frecuencia con la que las palabras aparecen en un documento. Para un término $t_i$ en un documento, podemos formularlo de la siguiente manera:

$$Tf_{i,j} = n_{i,j.} \tag{1}$$

En (1), tenemos que $n_{ij}$ es el número de ocurrencias de cada palabra $t_i$ en el documento $d_j$. Por otro lado, la Frecuencia del Documento Invertido (IDF) mide la importancia general de una palabra en un documento. La podemos formular de la siguiente forma:

$$idf_i = log\ (D/df_{i,}) \tag{2}$$

En (2), tenemos que D es el número total de documentos de texto y $df_i$ es un número de documentos el cual contiene el termino $t_i$ por lo menos una vez. Finalmente, tenemos que TF-IDF es una combinación de TF y de IDF, la formula quedaría así:

$$Tf - idf_{i,j} = tf_{i,j} \times idf_{i.} \tag{3}$$

## 2.3. Fase de Identificación/Clasificación

El algoritmo de clasificación basado en máquinas de soporte vectorial (SVM) es una máquina de aprendizaje supervisado, que requiere de datos de entrenamiento y datos de prueba. Consiste en encontrar un hiperplano óptimo como la función que separa dos clases de datos.

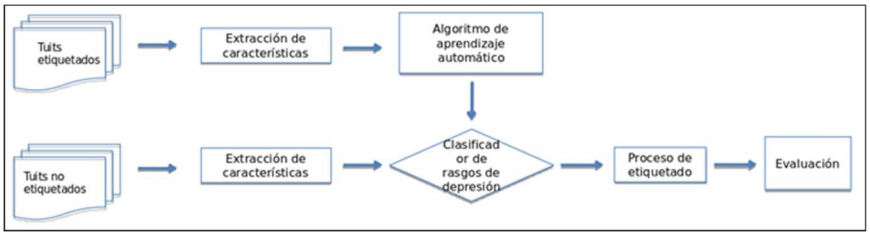


**Fig. 2.** Flujo del proceso de clasificación de rasgos depresivos.

La clasificación con menor error es la que se consigue con el hiperplano que maximiza el margen, esto es, cuya distancia entre el plano y los vectores soporte, sea la mayor posible. A pesar de su sencillez ha demostrado ser un algoritmo robusto y que generaliza bien en problemas de la vida real [44-48].

## 3. Resultados

El método propuesto, realiza una clasificación e identificación de tuits que nos permite tener una visualización precisa y directa, se puede determinar si la frase que fue extraída de Twitter tiene sentido de depresión o no y así poder ayudar a la persona que se requiera. En la Figura 2, podemos visualizar el problema de analizar los mensajes posteados en Twitter en términos de los sentimientos que estos mensajes expresan.

En donde primero nos dimos a la tarea de etiquetar un conjunto de tuits en español obtenidos bajo la metodología descrita. Adicional, cuando etiquetamos es importante considerar la presencia de la negación, debido a que la negación juega un papel muy importante en la detección de la polaridad de un mensaje (positivos se hacen negativos y viceversa).

Esta clasificación no es una tarea trivial y una de las características de Twitter es que es un tipo de comunicación informal, y con limitaciones de longitud. Esto lo hace diferente a otras investigaciones previas de análisis de sentimientos de textos convencionales. En la Tabla 2 se muestra las diez palabras con mayor frecuencia positivas y negativas respectivamente. Cabe resaltar que la palabra 'vida' aparece tanto en el lado de positivas como en el de negativas, más adelante, en la Tabla 3 explicamos el cambio de la polaridad.

Como parte de las limitaciones que tiene el presente trabajo es que, falta hacer más estudios para reducir la dispersión en donde podríamos aplicar técnicas de suavizado semántico entre otras [49].

Los resultados obtenidos por el clasificador bayesiano y las máquinas de soporte vectorial, se compararon mediante las siguientes métricas: exactitud, precisión y sensibilidad donde:

– Exactitud es una medida en porcentaje que se calcula de la siguiente forma:

$$Exactitud = (Tp + Tn) / (Tp+Tn+Fn+Fp). \quad (4)$$

– Sensibilidad positiva (5) y sensibilidad negativa (6) es el ratio de sensibilidad y es calculada de la siguiente forma:

**Tabla 2.** Listado de las palabras con mayor frecuencia.

| **Positivas** | **Negativas** |
|---|---|
| vida | vida |
| feliz | solo(a) |
| mejor | mal |
| contento | mierda |
| mundo | nadie |
| ganar | triste |
| amor | llorar |
| trabajo | dormir |
| esfuerzo | sentir |
| positivo | tiempo |

– La precisión positiva (7) y la precisión negativa (8) es el ratio de precisión y se calcula de la siguiente forma:

$$Precisión\ p = Tp/(Tp+Fp), \tag{7}$$

$$Precisión\ n = Tn/(Fn+Tn). \tag{8}$$

En la Tabla 3, se presentan cuatro tuits extraídos del conjunto de datos; podemos observar que la negación juega un papel muy importante en la detección de la polaridad de una frase (positivos se hacen negativos y viceversa), además de la negación se tiene que considerar adjetivos que acompañan al sustantivo y que cambian su cualidad.

$$Sensibilidad\ p = Tp\ /\ (Tp + Fn), \tag{5}$$

$$Sensibilidad\ n = Tn\ /\ (Fp + Tn). \tag{6}$$

## 4. Conclusiones

En la Tabla 4 se muestra la comparación del desempeño entre el clasificador bayesiano y las máquinas de soporte vectorial respectivamente en términos de precisión y sensibilidad. De manera similar, la Tabla 5 muestra el desempeño de los clasificadores en términos de exactitud.

La tendencia creciente de la depresión y del suicidio son un grave problema de salud pública. Sin duda, este es un problema que el sistema de salud mexicano debe enfrentar de manera urgente, por un lado, debe considerar que el país se encuentra en una etapa de incertidumbre económica (derivado de la actual pandemia), y por otro lado, que existen necesidades de atención de salud mental.

**Tabla 3.** Fragmentos de tuits extraídos del conjunto de datos.

| Tuits | Clase |
|---|---|
| La **vida** es como el **jazz**, mejor si es **improvisada** | Positiva |
| La **magia** es **creer** en **ti** mismo | Positiva |
| La **vida** es una **mierda** | Negativa |
| La **vida no** tiene **sentido** | Negativa |

**Tabla 4.** Métricas de desempeño.

| Métricas | % |
|---|---|
| Sensibilidad positiva | 84 |
| Sensibilidad negativa | 84 |
| Precisión positiva | 87 |
| Precisión negativa | 84 |

**Tabla 5.** Comparación de exactitud.

| Métodos | % |
|---|---|
| Clasificador bayesiano | 84 |
| Máquinas de soporte vectorial | 86 |

Nuestro método propuesto puede ser una base para más estudios de computación social y abre las puertas a futuras investigaciones sobre algoritmos de IA que hagan uso de otros datos de entrenamiento del tipo multifactorial y multinivel, como lo son las variables sociales, económicas y políticas.

Con el fin de explorar la salud mental, la idea central de la presente investigación parte del principio de clasificar un texto como positivo o negativo mediante algoritmos de IA.

Como primer paso, se describe una metodología con la que se genera un conjunto de datos en español y con ello, se establecen algunos pasos esenciales para la clasificación de rasgos depresivos.

Nosotros hemos aplicado el clasificador bayesiano y el clasificador de máquinas de soporte vectorial para la clasificación de textos con rasgos depresivos obteniendo muy buenos resultados. En futuros trabajos, procederemos a incrementar el tamaño del conjunto de datos mediante la metodología descrita, analizaremos diferentes técnicas para la representación de textos, por ejemplo, incorporaremos una reducción de dimensionalidad mediante un modelo de bolsa de palabras (BOW).

También podríamos combinar nuestros algoritmos con información de tipo multimodal que ofrece una nueva dimensión a los análisis tradicionales sobre texto, en donde podríamos tomar diferentes modalidades como datos visuales, de audio, entre otros [50-52]. Así como combinarla con otras técnicas de aprendizaje profundo mediante arquitecturas jerárquicas para incrementar la escalabilidad e incrementar la precisión de nuestro método [53, 54].

## Referencias


1. McGuine, T. A., Biese, K. M., Petrovska, L., Hetzel, S. J., Reardon, C. L., Kliethermes, S., Bell, D. R., Brooks, A., Watson, A. M.: Changes in the health of adolescent athletes: A comparison of health measures collected before and during the CoVID-19 pandemic. In: Proceedings of Journal of Athletic Training (2021)
2. Wang, Q., Su, M. A.: Preliminary assessment of the impact of COVID-19 on environment – A case study of China. Science of The Total Environment, vol. 728 (2020) doi: 10.1016/j.scitotenv.2020.138915
3. Liu, C. H., Zhang, E., Wong, G. T., Hyun, S., Hahm, H. C.: Factors associated with depression, anxiety, and PTSD symptomatology during the COVID-19 pandemic: Clinical implications for U.S. young adult mental health. Psychiatry Research, vol. 290 (2020) doi: 10.1016/j.psychres.2 020.113172
4. Shigemura, J., Kurosawa, M.: Mental health impact of the COVID-19 pandemic in Japan. Psychological Trauma: Theory, Research, Practice, and Policy, vol. 12, no. 5, pp. 478–79 (2020) doi: 10.1037/tra0000803
5. Martínez-Taboas, A.: Pandemias, COVID-19 y salud mental: ¿Qué sabemos actualmente? Revista Caribeña de Psicología, vol. 4, no. 2, pp. 143–152 (2020) doi: 10.37226/rcp.v4i2.4907
6. Landry, M. D., van den Bergh, G., Hjelle, K. M., Jalovcic, D., Tuntland, H. K.: Betrayal of Trust? The Impact of the COVID-19 Global Pandemic on Older Persons. Journal of Applied Gerontology, vol. 39, no. 7, pp. 687–689 (2020) doi: 10.1177/0733464820924131
7. Rodríguez Esparza, L. J., Barraza Barraza, D., Salazar Ibarra, J., Vargas Pasaye, R. G.: Index of suicide risk in Mexico using Twitter. Journal of Social Researches, pp. 1–13 (2019) doi:10.35429/JSR.2019.15.5.1.13
8. Cabello-Rangel, H., Márquez-Caraveo, M. E., Díaz-Castro, L.: Suicide rate, depression and the human development index: An ecological study from Mexico. Frontiers in Public Health, vol. 8 (2020) doi: 10.3389/fpubh.20 20.561966
9. Cervantes, C. A., Montaño, A. M.: Estudio de la carga de la mortalidad por suicidio en México 1990-2017. Revista Brasileira de Epidemiologia, vol. 23 (2020) doi: 10.1590/1980-549720200069
10. Cañón, Buitrago, S. C., Carmona Parra, J. A.: Ideación y conductas suicidas en adolescentes y jóvenes. Revista Pediatría de Atención Primaria, vol. 20, pp. 471–489 (2018)
11. Benítez Camacho, E.: Suicidio: El impacto del COVID-19 en la salud mental. Revista de Medicina y Ética, vol. 32, no. 1, pp. 15–39 (2021) doi: 10.36105/ mye.2021v32n1.01
12. Katikalapudi, R., Chellappan, S., Montgomery, F., Wunsch, D., Lutzen, K.: Associating internet usage with depressive behavior among college students. IEEE Technology and Society Magazine, vol. 31, no. 4, pp. 73–80 (2012)
13. Katie, G., Megan Moreno, A.: Alcohol references on undergraduate males' Facebook profiles. American Journal of Men's Health, vol. 5, no. 5, pp. 413–420 (2011) doi: 10.1177/1557988310394341
14. Chung, C., Pennebaker, J.: The psychological functions of function words. Social Communication, pp. 343–359 (2007)
15. Coppersmith, G., Dredze, M., Harman, C., Hollingshead, K., Mitchell, M.: CLPsych 2015 Shared Task: Depression and PTSD on Twitter. In: Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality, pp. 31–39 (2015) doi:10.3115/v1/W15-1204

16. Martínez-Castaño, R., Pichel, J. C., Losada, D. E.: A big data platform for real time analysis of signs of depression in social media. International Journal of Environmental Research and Public Health, vol. 17, no. 13 (2020) doi: 10.33 90/ijerph17134752
17. Zivanovic, S., Martinez, J., Verplanke, J.: Capturing and mapping quality of life using twitter data. GeoJournal, vol. 85, no. 1, pp. 237–255 (2020) doi: 10.1007/s10708-018-9960-6
18. Choudhury, M., Gamon, M., Counts, S., Horvitz, E.: Predicting depression via social media. In: Proceedings of the International AAAI Conference on Web and Social Media, vol. 7, no. 1, pp. 128–1973
19. Alsaeedi, A., Zubair, M.: A study on sentiment analysis techniques of Twitter data. International Journal of Advanced Computer Science and Applications, vol. 10, no. 2, pp. 361–374 (2019)
20. Sidorov, G., Miranda-Jiménez, S., Viveros-Jiménez, F., Gelbukh, A., Castro-Sánchez, N., Velásquez, F., Díaz-Rangel, I., Suárez-Guerra, S., Treviño, A., Gordon, J.: Empirical study of machine learning based approach for opinion mining in tweets. Batyrshin, I., González Mendoza, M. (eds) Advances in Artificial Intelligence. MICAI 2012, Lecture Notes in Computer Science, vol. 7629, Springer (2013) doi: 10.1007/978-3-642-37807-2_1
21. Chaturvedi, S., Mishra, V., Mishra, N.: Sentiment analysis using machine learning for business intelligence. In: IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI), pp. 2162–2166 (2017) doi: 10.1109/ICP CSI.2017.8392100
22. Garcia-Lopez, F. J., Batyrshin, I., Gelbukh, A.: Analysis of relationships between tweets and stock market trends. Journal of Intelligent & Fuzzy Systems, vol. 34, no. 5, pp. 3337–3347 (2018) doi: 10.3233/JIFS-169515
23. Bernábe Loranca, M. B., González Velázquez, E. E., Cerón Garnica, C.: Algorithm for collecting and sorting data from twitter through the use of dictionaries in Python. Computación y Sistemas, vol. 24, no. 2 (2020) doi: 10.13053/cys-24-2-3405
24. Rill, S., Reinel, D., Scheidt, J., Zicari, R. V.: PoliTwi: Early detection of emerging political topics on Twitter and the impact on concept-level sentiment analysis. Knowledge-based Systems, vol. 69, pp. 24–33 (2014)
25. Pavan Kumar, C. S., Dhinesh Babu, L. D.: Fuzzy based feature engineering architecture for sentiment analysis of medical discussion over online social networks. Journal of Intelligent & Fuzzy Systems, vol. 40, no. 6, pp. 11749–11761 (2021)
26. Gutiérrez, G., Canul-Reich, J., Ochoa Zezzatti, A., Margain, L., Ponce, J.: Mining: Students comments about teacher performance assessment using machine learning algorithms. International Journal of Combinatorial Optimization Problems and Informatics, vol. 9, no. 3, pp. 26–40 (2018)
27. Gupta, V., Singh, V. K., Mukhija, P. Ghose, U.: Aspect-based sentiment analysis of mobile reviews. Journal of Intelligent & Fuzzy Systems, vol. 36, pp. 4721–4730 (2019)
28. Wang, J., Zhang, X., Yu Zhang, H.: Hotel recommendation approach based on the online consumer reviews using interval neutrosophic linguistic numbers. Journal of Intelligent & Fuzzy Systems, vol. 34, pp. 381–394 (2018)
29. González Brito, O., Tapia Fabela, J. L., Salas Hernández, S.: Method of extraction of feature in the classification of texts for authorship attribution. International Journal of Combinatorial Optimization Problems and Informatics, vol. 12, no. 3, pp. 87–97 (2021)
30. Maldonado-Sifuentes, C. E., Sidorov, G., Kolesnikova, O.: Improved Twitter virality prediction using text and RNN-LSTM. International Journal of Combinatorial Optimization Problems and Informatics, vol. 12, no. 3, pp. 50–62 (2021)

31. Hernández Castañeda, Á., García Hernández, R. A., Ledeneva, Y., Millán Hernández, C. E.: The impact of key ideas on automatic deception detection in text. Computación y Sistemas, vol. 24, no. 3 (2020)
32. Posadas-Durán, J. P., Gómez-Adorno, H., Sidorov, G., Escobar, J. J.: Detection of fake news in a new corpus for the Spanish language. Journal of Intelligent & Fuzzy Systems, vol. 36, no. 5, pp. 4869–4876 (2019)
33. Calvo, H., Gambino, O. J., García Mendoza, C. V.: Irony detection using emotion cues. Computación y Sistemas, vol. 24, no. 3 (2020)
34. Daouadi, K. E., Rebaï, R. Z., Amous, I.: Organization, bot, or human: Towards an efficient twitter user classification. Computación y Sistemas, vol. 23, no. 2, pp. 273–279 (2019)
35. Zimbra, D., Abbasi, A., Zeng, D., Chen, H.: The state-of-the-art in Twitter sentiment analysis: A review and benchmark evaluation. ACM Transactions on Management Information Systems (TMIS), vol. 9, no. 2, pp. 1–29 (2018)
36. Zucco, C., Calabrese, B., Cannataro, M.: Sentiment analysis and affective computing for depression monitoring. In: IEEE International Conference on Bioinformatics and Biomedicine, pp. 1988–1995 (2017)
37. Vázquez-Hernández, M., Pineda, L. V., Montes-y-Gómez, M.: Identificación y pesado de términos para la detección de depresión en Twitter. Research in Computer Science, vol. 149, no. 8, pp. 465–474 (2020)
38. Trupthi, M., Pabboju, S., Narasimha, G.: Sentiment analysis on twitter using streaming API. In: IEEE 7th International Advance Computing Conference (IACC), pp. 915–919 (2017) doi: 10.1109/IACC.2017.0186
39. Balazevic, I., Braun, M., Müller, K. R.: Language detection for short text messages in social media (2016) doi: 10.48550/arXiv.1608.08515
40. Khader, M., Awajan, A. A., Al-Naymat, G.: The impact of natural language preprocessing on big data sentiment analysis. vol.16, no. 3, pp. 506–513 (2019)
41. Saif, H., Fernandez, M., He, Y., Alani, H.: On stopwords, filtering and data sparsity for sentiment analysis of Twitter. In: Proceedings of the Ninth International Conference on Language Resources and Evaluation, pp. 810–817 (2014)
42. Jabbar, A., Iqbal, S., Tamimy, M. I., Hussain, S., Akhunzada, A.: Empirical evaluation and study of text stemming algorithms. Artificial Intelligence Review, vol. 53, no. 8, pp. 5559–5588 (2020) doi: 10.1007/s10462-020-09828-3
43. Zhu, Z., Liang, J., Li, D., Yu, H., Liu, G.: Hot topic detection based on a refined TF-IDF algorithm. IEEE Access, vol. 7, pp. 26996–27007 (2019) doi: 10.1109/ACCESS. 2019.2893980
44. Khanna, D., Sahu, R., Baths, V., Deshpande, B.: Comparative study of classification techniques (SVM, Logistic Regression and Neural Networks) to Predict the Prevalence of Heart Disease'. International Journal of Machine Learning and Computing, vol. 5, no. 5, pp. 414–419 (2015) doi: 10.7763/IJM LC.2015.V5.544
45. Lopez-Martin, C., Banitaan, S., Garcia-Floriano, A., Yanez-Marquez, C.: Support vector regression for predicting the enhancement duration of software projects. In: 16th IEEE International Conference on Machine Learning and Applications (ICMLA), pp. 562–567 (2017)
46. Toledo, G. R., Sánchez, N. A., Sidorov, G., Durán, J. P.: Identificación de cambios en el estilo de escritura literaria con aprendizaje automático. Onomázein: Revista de lingüística, filología y traducción de la Pontificia Universidad Católica de Chile, vol. 46, pp. 102–128 (2019)
47. Ramírez-García, J., Ibarra-Orozco, R. E., Cruz, A. J.: Tweets monitoring for real-time emergency events detection in smart campus. In: Mexican International Conference on Artificial Intelligence, pp. 205–213 (2020)

48. Nieto-Benitez, K., Castro-Sánchez, N. A., Jiménez-Salazar, H.: Reconocimiento de patrones para la clasificación de componentes argumentales en textos académicos en español. Research in Computing Science, vol. 149, no. 8, pp. 637-648 (2020)
49. Altinel, B., Ganiz, M. C.: Semantic text classification: A survey of past and recent advances. Information Processing & Management, vol. 54, no. 6, pp. 1129–1153 (2018)
50. Poria, S., Cambria, E., Gelbukh, A.: Deep convolutional neural network textual features and multiple kernel learning for utterance-level multimodal sentiment analysis. In: Conference on Empirical Methods in Natural Language Processing, pp. 2539–2544 (2015) doi: 10.18653/v1/D15-1303
51. Krishnamurthy, G., Majumder, N., Poria, S., Cambria, E.: A deep learning approach for multimodal deception detection (2018)
52. Banerjee, T., Yagnik, N., Hegde, A.: Impact of cultural-shift on multimodal sentiment analysis. Journal of Intelligent & Fuzzy Systems, pp. 5487–5496 (2021) doi: 10.3233/JIFS-189870
53. Kastrati, Z., Imran, A. S., Yayilgan, S. Y.: The impact of deep learning on document classification using semantically rich representations. Information Processing & Management, vol. 56, no. 5, pp. 1618–1632 (2019)
54. Amjad, M., Voronkov, I., Saenko, A., Gelbukh, A.: Comparison of text classification methods using deep learning neural networks. In: Proceedings of the 20th International Conference on Computational Linguistics and Intelligent Text Processing (2019)