

SVM Based Learning System for the Detection of Depression in Social Networks

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

Instituto Politécnico Nacional,
Escuela Superior de Ingeniería Mecánica y Eléctrica Culhuacán,
Departamento de Ingeniería en Computación,
Mexico

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

Abstract. Depression represents a problem of public concern that is now prioritized in many health care agendas with the intention of preventing future suicides, which have devastating impact not only because of tragic loss of life, but also for the grieving family and friends. Investigations in each country reveal a reduction in physical and mental well-being; for this reason, the proposal presented in this article comprises an attempt to detect the feelings expressed in text sentences presented in social networks.

Keywords. Social networks, depression, machine learning.

1 Introduction

Today, the world is going through a time of transformation; daily life has made a 360 degree turn, where the protagonist, a viral strain known as SARS-CoV-2 has caused somewhat more than four million deaths. Apart from the economic consequences, social confinement is usually an unpleasant experience, which can lead to different stress factors that generate mental health outcomes [1]. Because this situation is new and very much expanding, it is still premature to estimate the emotional consequences of the epidemic outbreak. However, research carried out in [2,3] suggests that fear of the unknown and uncertainty can lead to the evolution of different mental health diseases such as: stress disorders, anxiety, depression, somatization and degenerative behaviors, resulting in an increase in the consumption of alcohol, tobacco and other substances harmful to health [4]. In particular,

people with chronic illnesses are expected to have higher levels of psychological symptoms [5]. Older people are also predicted to be more psychologically vulnerable than young people in this crisis [6].

This project has emerged because of the great problem of suicide cases in our country among young people [7,8,9]. For this reason, we have decided to develop a tool that is capable of alerting possible cases so that they can be contained.

1.1 Depression and Artificial Intelligence

Notably in Mexico, it was found that young adults (that is, between 15 and 25 years old) present suicidal ideas and show greater depressive states, meaning depression is evident in 67.3% of those who have attempted suicide and in 81.1% of those who manifest suicidal ideas [10]. Likewise, people with mental illness tend to disclose their mental condition on social media, as a way of seeking relief [11].

However, research on employing social media, as a means to understanding behavioral health disorders, is still in its infancy.

In [12], web activity patterns of university students were analyzed, as they may indicate depression.

Similarly, in [13] they showed that Facebook status updates may reveal symptoms of depressive episodes. Some differences have been noted, such as the fact that depressed users more frequently use first person pronouns, [14] as well as words indicating negative emotions and anger.

For this reason, depression has been associated with the use of linguistic markers such as greater use of first person pronouns. Many other studies of language and depression have been limited to clinical settings, and therefore to analyzing spontaneous speech or written essays.

Following this lead, some research [15, 16] has proposed innovative methodologies to amass textual content shared by people diagnosed with depression. However, there are no publicly available sources. This is because the text is often taken from social networking sites such as Twitter or Facebook that do not permit redistribution [17]. Hence, these previous studies direct us towards detecting depression in social networks as the first step against suicide. The main procedure in mental health studies using social networks has traditionally been carried out by applying surveys, where the number of users is limited by those who manage to complete the survey.

For example, in [18] Twitter users were requested to take a Center for Epidemiological Studies Depression Scale (CES-D) and share their profile with the public. This type of study has produced high quality data; however it is limited in size and scope. Therefore, in this research we will examine depression by considering automatically obtained samples from large amounts of Twitter data. The Internet has allowed us to follow the evolution of language and is providing us with a very accessible medium for people to express their feelings anonymously.

Hence, we have adapted the method in [15] for the construction of this data set in Spanish, we will proceed to identify self-expressions of diagnoses of mental illnesses and we take advantage of these messages to build our data set.

1.2 Analysis of Sentiments

Generally, the word feeling refers to a way of thinking (opinion) or sentiment (emotion) about something [20]. One of the best known tasks involved in sentiment analysis (SA) is the detection of polarity that can be synthesized in the following classification problem: "Given a text T as input, depending on its content, determine whether T contains a positive/negative/neutral opinion (and eventually determine a force parameter that indicates how much of the con- tent is

positive/negative). This sentiment analysis task is widely used in contexts such as the review of products or services, political predictions, among others [19].

Recently, the main focus of another line of research is medical and psychological, where the task of emotion recognition is carried out on forums, chats, and social networks. Data on social networks such as Twitter and Facebook, where users post their reactions and comments in real time, pose new and different challenges.

First of all, it seems that studies can be divided into supervised and lexical-based methods. The supervised methods are based on training classifiers, such as Naive Bayes, Vector Support Machines, Random Forest, whilst, lexicon-based methods determine the tendency of the sentiment in a text based on the use of pre-established lexicons of previously weighted words, in terms of the feelings expressed [20].

Similarly, the analysis of real-time data from social networks (SMA), has received considerable attention in recent years in the context of analysis for the detection of abnormal events/activities. Today, we can say that there are millions of Twitter posts, millions of Facebook posts and billions of forums on web pages and various documents that can be reviewed in order to determine the opinions expressed behind the words.

2 Related work

Originally, SA was associated with business intelligence [21, 23], but it has spread to other areas such as politics [23, 24], medicine [25], education [26], recommendations [27, 28], screening for plagiarism [29], news influences [30], deception detection [31, 32], irony detection [33] and account classification [34], among others.

In particular, SA is a prominent research topic in the field of computational linguistics. Tasks include classifying the polarity of sentiment expressed in text (e.g. positive, negative, and neutral), identifying the target/theme of sentiment, and identifying the sentiment in terms of various aspects of a theme.

The sentiment polarity classification problem is often modeled as two-way (positive / negative) or three-way (positive / negative / neutral) [35]. It is

Table 1. Regular expressions for tweet detection

Word	Regular expression
Depression	(depress[ion ed ive ant ing])
Associated phrases	(problem[s] disturbance[s])(mental psychological[s] psychiatrical[s])(die[d] (day[s]) (sad+problem[s])

important to note that this task of detection and classification is not easy, firstly because tweets are short messages where the indicators of depression tend to manifest themselves in a very subtle way.

Due to the widespread adoption of social media and the availability of large-scale data from social media, approaches to using this data for screening for depression are receiving increased attention from researchers.

In [22], it is shown that college students show symptoms that indicate depression on Facebook.

In [27], the differences between Twitter users with and without depression are analyzed by analyzing their activities.

In [28], a similar analysis is performed by analyzing Facebook data, using multiple regression analysis.

Some recent studies have submitted encouraging results for the detection of users suffering from depression, but more studies are still required [36,37].

In our research, we have formulated models in a bidirectional way (positive/negative); leaving to the future, the task of evaluating greater intensities of depressive sentiment: strong positives, strong negatives, mild positives, and mild negatives.

Detecting sentiments from the phrases is a complicated task, for example: "life is like jazz, better if it is improvised"; the sentiment of the opinion is positive because the word "life" implies something good.

However, the same word in another context, as shown in the following statement: 'my life is meaningless', implies a negative feeling - it is bad because negation reduces the positive in the word 'life'. Thus the problem involves the use of language, which is a very complex and huge problem.

3 Methodology

To solve this problem, a three-phase model was proposed:

3.1 Collection phase

During this phase, we took advantage of the large amount of data provided by Twitter. The collection method is based on two main stages: first, the tweets are filtered out from regular expressions and then these are classified into: negative and positive.

To acquire the tweets for this study, we developed an application that uses the Twitter search API [21].

To filter tweets that are not written in Spanish, we used the free language detection library [22]. This library is based on Bayesian filters and has a precision of 0.99 in the detection of the 53 languages to which it permits entry. The tweets were acquired over 180 days (from December 1, 2020 to June 1, 2021), producing data sets with approximately 3800 tweets for Spanish.

To generate data from a set of tweets with depressive traits, we considered tweets from people who declared to have been diagnosed with the disease of depression.

Table 1 shows the regular expressions used to detect people who refer to depression in their tweets; however the main intention is to identify people who make a direct and open statement that they were diagnosed with the disease of depression.

Subsequently, the tweets are extracted from the list of people who asserted through a tweet that they suffer from this disease.

3.2 Preprocessing phase

Data preprocessing is an often neglected but important stage in the process. It involves techniques to transform the raw data into a more understandable format. The main ones are data cleansing, data integration, data transformation and data reduction.

As apparent in Figure 1, our preprocessing mechanism includes:

- a. Extraction,

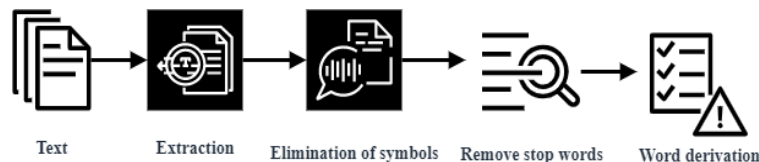


Fig. 1. Preprocessing mechanism

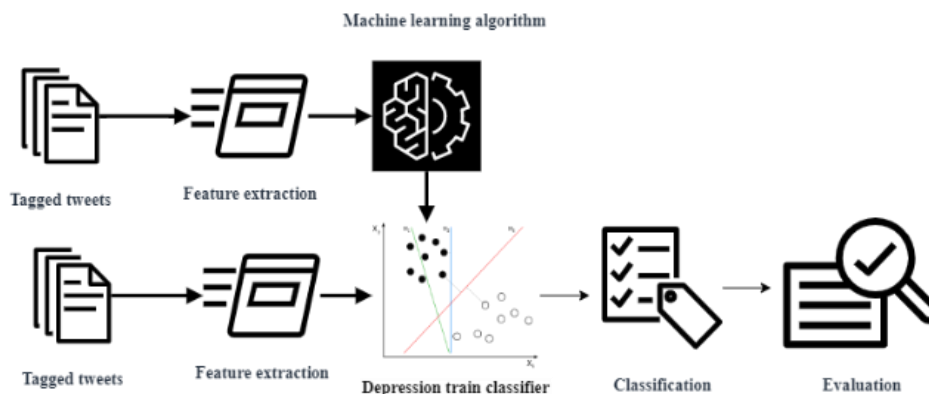


Fig. 2. Identification phase

- b. Elimination of numbers and URLs that may have an effect on our analysis but do reduce noise and our efficiency [23].
- c. Elimination of stop words such as articles, pronouns, and prepositions [24].
- d. Word derivation, which is used to transform different word forms into a standard root form [25].

In this phase, in addition to these techniques, we incorporate a weighting step using the Term Frequency-Inverse Document (TF-IDF) algorithm.

The TF-IDF reflects the importance of a word in a document; and this level of importance increases when the word appears many times, to the point that we can determine the themes that are trending [26].

Term Frequency (TF) is the frequency with which words appear in a document. The term t_i in a document can be formulated as follows:

$$Tf_{ij} = n_{ij} \quad (1)$$

In (1), we have that n_{ij} is the number of the word t_i occurs in the document d_j . In contrast, Inverted Document Frequency (IDF) measures the overall

importance of a word in a document. We can formulate this in the following way:

$$idf_{ij} = \log D/df_{ij} \quad (2)$$

In (2), we have that D is the total number of text documents and df_{ij} is the number of documents d_j which contain the term t_i .

Finally, we have that TF-IDF is a combination of TF and IDF; the formula would look like this:

$$Tf-idf_{ij} = tf_{ij} \times idf_{ij} \quad (3)$$

3.3 Identification/classification phase

The classification algorithm based on vector support machines (SVM) is a supervised learning machine, which requires training data and test data. This consists in finding an optimal hyperplane as the function that separates two types of data. The classification with the lowest error is the one obtained from the hyperplane that maximizes the margin, that is, whose distance between the plane and the support vectors is as large as possible. Despite its simplicity, this has proven to be a robust

Table 2. Highest positive and negative frequency

Positives	Negatives
life	like
happy	alone
better	bad
funny	shit
world	nobody
win	sad
love	cry
work	sleep
effort	feel
positive	Time

Table 3. Performance metrics

Tweets	Class
Life is like jazz; better when improvised	Positive
Magic is believing in yourself	Positive
Life is a waste of time	Negative
Life is meaningless	Negative

Table 4. Performance comparison

Metrics	%
Positive sensitivity	84
Negative sensitivity	84
Positive precision	87
Negative precision	84

Table 5. Performance comparison

Algorithm	%
Bayesian classifier	84
SVM classifier	86

algorithm that generalizes well to real-life problems [44-48].

4 Results

The proposed method involves the classification and identification of tweets that allows us to have

an accurate and direct visualization, it can determine whether the phrase that was extracted from Twitter is indicative of depression or not, making it possible to help the person who is in need.

In Figure 2, we can visualize the problem of analyzing the messages posted on Twitter in terms of the sentiments these messages express. Here, our first task was to tag a set of tweets in Spanish, obtained using the methodology described. Likewise, when we label, it is important to consider the presence of negation, because negation plays a very important role when detecting the polarity of a message (positives become negative and vice versa).

This classification is not a trivial task and one of the characteristics of Twitter is that it is a type of informal communication, with limitations in length. This makes it different from previous research on sentiment analysis, using conventional texts.

Table 2 shows the ten words with the highest positive and negative frequency, respectively. Notably, the word 'life' appears both on the positive and negative sides. Later, in Table 3, we explain this change in polarity.

Because of the limitations of this work, it will be necessary to do more studies to reduce dispersion, in order to apply semantic smoothing techniques, among others [49].

The results obtained using the Bayesian classifier and vector support machines were compared using the following metrics: accuracy, precision and sensitivity where,

Accuracy (4) is a percentage measure that is calculated as follows:

$$\text{Accuracy} = (Tp + Tn) / (Tp + Tn + Fn + Fp) \quad (4)$$

Positive sensitivity (5) and negative sensitivity (6) is the sensitivity ratio and is calculated as follows:

$$\text{Sensitivity } p = Tp / (Tp + Fn), \quad (5)$$

$$\text{Sensitivity } n = Tn / (Fp + Tn). \quad (6)$$

Positive precision (7) and negative precision (8) is the precision ratio and is calculated as follows:

$$\text{Precision } p = Tp / (Tp + Fp), \quad (7)$$

$$\text{Precision } n = Tn / (Fn + Tn). \quad (8)$$

In Table 3, four tweets extracted from the data set are presented; we can see that negation plays a very important role in detecting the polarity of a sentence (positives become negative and vice versa), in addition to negation, adjectives that accompany the noun and that change their quality must be considered.

Table 4 shows the performance comparison between the Bayesian classifier and vector support machines, respectively, in terms of precision and sensitivity. Similarly, Table 5 shows the performance of the classifiers in terms of accuracy.

5 Conclusions

The increasing trend for depression and suicide represents a serious public health problem. Undoubtedly, this is a problem that the Mexican health system must face with urgency, firstly considering that the country is in a stage of economic uncertainty (derived from the current pandemic), and secondly, because there are evident mental health care needs.

Our method can provide the basis for more social computing studies and opens the doors to future research on AI algorithms that make use of other training data of the multifactorial and multilevel type, such as social, economic and political variables. In order to explore mental health, the central idea of this research starts from the principle of classifying a text as positive, or negative, using AI algorithms. As a first step, we describe a methodology from which to generate a data set in Spanish and using this, some essential steps are established for the classification of depressive traits.

We have applied the Bayesian classifier and the vector support machine classifier to classify texts with depressive features, obtaining very good results.

In future works, we will proceed to increase the size of the data set using the methodology described. We will also analyze different techniques for representing texts, for example we will incorporate a dimensionality reduction through a bag of words model (BOW). We could also combine our algorithms with multimodal information so as to offer a new dimension to traditional text analysis, where we could take into

account different modalities such as visual and audio data, among others [50, 52]. Likewise, we can incorporate deep learning techniques into our method, using hierarchical architectures to increase scalability and precision [53, 54].

References

1. **McGuine, T., Biese, K., Hetzel, S., Petrovska, L., Kliethermes, S., Reardon, C., Bell, D., Brooks, A., Watson, A. (2021).** Changes in the health of adolescent athletes: a comparison of health measures collected before and during the CoVID-19 pandemic. *Journal of Athletic Training*, Vol. 56, No. 8, pp. 836–844. DOI: 10.4085/1062-6050-0739.20.
2. **Wang, Q., Su, M. (2020).** A preliminary assessment of the impact of COVID-19 on environment - a case study of China. *The Science of the total environment*, Vol. 728, pp. 138915–138915.
3. **Liu, C., Zhang, E., Wong, G., Hyun, S. Hahm, H. (2020).** Factors associated with depression, anxiety, and PTSD symptomatology during the COVID-19 pandemic: Clinical implications for US young adult mental health. *Psychiatry research*, Vol. 290. DOI: 10.1016/j.psychres.2020.113172.
4. **Shigemura, J., Kurosawa, M. (2020).** Mental health impact of the COVID-19 pandemic in Japan., *Psychological Trauma: Theory, Research, Practice, and Policy*, pp. 478–79. DOI: 10.1037/tra0000803.
5. **Martínez, A. (2020).** Pandemias, COVID-19 y salud mental: ¿qué sabemos actualmente?. *Revista Caribeña de Psicología*, Vol. 4, No. 2, pp.143–52. DOI: 10.37226/rcp.v4i2.4907.
6. **Landry, D., Van den Bergh, G., Hjelle, K., Jalovcic, D., Tuntland, H. (2020).** Betrayal of trust? The impact of the COVID-19 global pandemic on older persons. *Journal of Applied Gerontology*, Vol. 39, No. 7, pp. 687–89. DOI: 10.1177/0733464820924131.
7. **Rodríguez, L., Barraza, D., Salazar, J., Vargas, R. (2019).** Index of suicide risk in Mexico using twitter. *Journal of Social Researches*, Vol. 5, No. 15, pp. 1–13. DOI: 10.35429/JSR.2019.15.5.1.13.

8. **Cabello, H., Márquez, M., Díaz, L. (2020).** Suicide Rate, Depression and the human development index: an ecological study from Mexico. *Frontiers in Public Health*, Vol. 8, pp. 5611966. DOI: 10.3389/fpubh.2020.561966.
9. **Dávila, C., Pardo, A. (2020).** Estudio de la carga de la mortalidad por suicidio en México 1990-2017. *Revista Brasileira de Epidemiologia*, Vol. 23. DOI: 10.1590/1980-549720200069.
10. **Cañón, S., Carmona, J. (2018).** Ideación y conductas suicidas en adolescentes y jóvenes. *Rev Pediatr Aten Primaria*, Vol. 20, No. 80.
11. **Benítez, E. (2021).** Suicidio: El impacto del Covid-19 en la salud mental. *Revista de Medicina y Ética*, Vol. 32, No. 1, pp. 15–39. DOI: 10.36105/mye.2021v32n1.01.
12. **Katikalapudi, R., Chellappan, S., Montgomery, F., Wunsch, D., Lutzen, K. (2012).** Associating internet usage with depressive behavior among college students. *IEEE Technology and Society Magazine*, Vol. 31, pp. 73–80. DOI: 10.1109/MTS.2012.2225462.
13. **Egan, K., Moreno, M. (2011).** Alcohol references on undergraduate males, facebook profiles. *American Journal of Men's Health*, Vol. 5, No. 5, pp. 413–20. DOI: 10.1177/1557988310394341.
14. **Chung, C., Pennebaker, J. (2007).** The psychological functions of function words. *Social Communication*, pp. 343–359.
15. **Coppersmith, G., Dredze, M., Harman, C., Hollingshead, K., Mitchell, M. (2015).** CLPsych shared task: depression and PTSD on twitter. *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, pp. 31–39. DOI: 10.3115/v1/W15-1204.
16. **Martínez, R., Pichel, J., Losada, D. (2020).** 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, pp. 4752. DOI: 10.3390/ijerph17134752.
17. **Zivanovic, S., Martinez, J., Verplanke, J. (2020).** Capturing and mapping quality of life using twitter data. *GeoJournal*, Vol. 85, pp. 237–55. DOI: 10.1007/s10708-018-9960-6.
18. **De Choudhury, M., Gamon, M., Counts, S., Horvitz, E. (2013).** Predicting depression via social media. *Seventh international AAAI conference on weblogs and social media*, Vol. 7, No. 1.
19. **Alsaeedi, A., Khan, M. (2019).** A study on sentiment analysis techniques of twitter data. *International Journal of Advanced Computer Science and Applications*, Vol. 10, No. 2, pp. 361–374. DOI: 10.14569/IJACSA.2019.0100248.
20. **Sidorov G., Miranda S., Viveros, F., Gelbukh A., Castro-Sánchez, N., Velásquez, F., Díaz-Rangel, I., Suárez-Guerra, S., Treviño, A., Gordon, J. (2013).** Empirical study of machine learning based approach for opinion mining in tweets. *Mexican International Conference on Artificial Intelligence*, Vol. 7629. DOI: 10.1007/978-3-642-37807-2_1.
21. **Chaturvedi, S., Mishra, V., Mishra, N. (2017).** Sentiment analysis using machine learning for business intelligence. *2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI)*, pp.2162–2166. DOI: 10.1109/ICPCSI.2017.8392100.
22. **García, F., Batyrshin, I., Gelbukh, A. (2018).** Analysis of relationships between tweets and stock market trends. *Journal of Intelligent & Fuzzy Systems*, Vol. 34, No. 5, pp. 3337–3347. DOI: 10.3233/JIFS-169515.
23. **Bernábe, M., Espinoza, E., González, R., Cerón C. (2020).** Algorithm for collecting and sorting data from twitter through the use of dictionaries in python. *Computación y Sistemas*, Vol. 24, No. 2, pp. 719–724. DOI: 10.13053/cys-24-2-3405.
24. **Rill, S., Reinel, D., Scheidt, J., Zicari, R. (2014).** 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. DOI: 10.1016/j.knosys.2014.05.008.

25. **Pavan, C., Dhinesh, L. (2021).** 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. DOI: 10.3233/JIFS-202874.
26. **Gutiérrez, G., Canul-Reich, J., Ochoa, A., Margain, L., Ponce, J. (2018).** 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.
27. **Gupta, V., Singh, V., Mukhija, P. Ghose, U. (2019).** Aspect-based sentiment analysis of mobile reviews. *Journal of Intelligent & Fuzzy Systems*, Vol. 36, No. 5. Pp. 4721–4730. DOI: 10.3233/JIFS-179021.
28. **Wang, J., Zhang, X., Zhang, H. (2018).** Hotel recommendation approach based on the online consumer reviews using interval neutrosophic linguistic numbers. *Journal of Intelligent and Fuzzy Systems*, Vol. 34, pp. 381–394. DOI: 10.3233/JIFS-171421.
29. **González, O., Tapia, J., Salas, S. (2021).** 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.
30. **Maldonado, C., Sidorov, G., Kolesnikova, O. (2021).** Improved twitter virality prediction using text and RNN-LSTM. *International Journal of Combinatorial Optimization Problems and Informatics*, Vol. 12, No. 3, pp. 50–62.
31. **Hernández, A., García, R., Ledeneva, Y., Millán, C. (2020).** The impact of key ideas on automatic deception detection in text. *Computación y Sistemas*, Vol. 24, No. 3. DOI: 10.13053/cys-24-3-3483.
32. **Posadas, J., Gómez, H., Sidorov, G., Escobar, J. (2019).** Detection of fake news in a new corpus for the Spanish language. *Journal of Intelligent & Fuzzy Systems*, Vol. 36, No. 5, pp. 4869–4876. DOI: 10.3233/JIFS-179034.
33. **Calvo, H., Gambino, O. J., García, C. (2020).** Irony detection using emotion cues. *Computación y Sistemas*, Vol. 24, No. 3. DOI: 10.13053/cys-24-3-3487.
34. **Daouadi, K., Rebaï, R., Amous, I. (2019).** Organization, bot, or human: Towards an efficient twitter user classification. *Computación y Sistemas*, Vol. 23, No. 2, pp. 273–279. DOI: 10.13053/cys-23-2-3192.
35. **Zimbra, D., Abbasi, A., Zeng, D., Chen, H. (2018).** 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. DOI: 10.1145/3185045.
36. **Zucco, C., Calabrese, B., Cannataro, M. (2017).** Sentiment analysis and affective computing for depression monitoring. 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 1988–1995. DOI: 10.1109/BIBM.2017.8217966.
37. **Vázquez, M., Villaseñor L., Montes, M. (2020).** Identificación y pesado de términos para la detección de depresión en twitter. *Research in Computer Science*, Vol. 149, No. 8, pp. 465–474.
38. **Trupthi, M., Suresh P., Narasimha, G. (2017).** Sentiment analysis on twitter using streaming API. 7th IEEE International Advance Computing Conference (IACC), pp.915–19. DOI: 10.1109/IACC.2017.0186.
39. **Balazevic, I., Braun, M., Müller, K. (2016).** Language detection for short text messages in social media. *ArXiv:1608.08515*.
40. **Khader, M., Awajan, A., Al-Naymat, G. (2019).** The impact of natural language preprocessing on big data: sentiment analysis, Vol.16, No. 3, pp. 8.
41. **Saif, H., Fernández, M., He, Y., Alani, H. (2014).** On stop-words, filtering and data sparsity for sentiment analysis of twitter. 9th International Conference on Language Resources and Evaluation.
42. **Jabbar, A., Iqbal, S., Tamimy, M., Hussain, S., Akhunzada, A. (2020).** Empirical evaluation and study of text stemming algorithms. *Artificial Intelligence Review*, Vol.

- 53, No. 8, pp. 5559–5588. DOI: 10.1007/s10462-020-09828-3.
43. **Zhu, Z., Liang, J., Li, D., Yu, H., Liu, G. (2019).** Hot topic detection based on a refined TF-IDF algorithm. *IEEE Access*, Vol. 7, pp. 26996–27007. DOI: 10.1109/ACCESS.2019.2893980.
 44. **Khanna, D., Sahu, R., Baths, V., Deshpande, B. (2015).** 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*. DOI: 10.7763/ijmlc.2015.v5.544.
 45. **Lopez, C., Banitaan, S., Garcia, A., Yanez, C., (2017).** Support vector regression for predicting the enhancement duration of software projects. *16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 562–567. DOI: 10.1109/ICMLA.2017.0-101.
 46. **Ríos, G., Castro, N., Sidorov, G., Posadas, J. (2019).** 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*, No. 46, pp. 102-128.
 47. **Ramírez, J., Ibarra, R., Arguelles, A. (2020).** Tweets monitoring for real-time emergency events detection in smart campus. *Mexican International Conference on Artificial Intelligence*, pp. 205–213. DOI: 10.1007/978-3-030-60887-3_18.
 48. **Nieto, K., Castro, N., Jiménez, H. (2020).** 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.
 49. **Altinel, B., Ganiz, M. (2018).** Semantic text classification: a survey of past and recent advances. *Information Processing & Management*, Vol. 54, No. 6, pp. 1129–1153. DOI: 10.1016/j.ipm.2018.08.001.
 50. **Poria, S., Cambria E., Gelbukh, A. (2015).** Deep convolutional neural network textual features and multiple kernel learning for utterance-level multimodal sentiment analysis. *2015 Conference on Empirical Methods in Natural Language Processing*, pp. 2539–2544. DOI: 10.18653/v1/D15-1303.
 51. **Krishnamurthy, G., Majumder, N., Poria, S., Cambria, E. (2018).** A deep learning approach for multimodal deception detection. *arXiv:1803.00344*.
 52. **Banerjee, T. Yagnik, N. Hegde, A. (2021).** Impact of cultural-shift on multimodal sentiment analysis. *Journal of Intelligent & Fuzzy Systems*. Vol. 41, pp.1–10. DOI:10.3233/JIFS-189870.
 53. **Kastrati, Z., Imran, A., Yayilgan, S. (2019).** The impact of deep learning on document classification using semantically rich representations. *Information Processing & Management*, Vol. 56, No. 5, pp. 1618–1632. DOI: 10.1016/j.ipm.2019.05.003.
 54. **Amjad, M., Voronkov, I., Saenko, A., Gelbukh, A. (2019).** Comparison of text classification methods using deep learning neural networks. *Proceedings of the 20 th International Conference on Computational Linguistics and Intelligent Text Processing (CICLing)*.
Article received on 29/07/2021; accepted on 30/09/2021. Corresponding author is Eduardo Vázquez Fernández.