ORIGINAL PAPER

Study of Text Patterns Found on Social Networks of Mental Health Reactions to COVID-19

Denis Cedeno-Moreno[1,2], Alan Delgado-Herrera[2], Nelson Montilla-Herrera[3], Miguel Vargas-Lombardo[1]

[1]Grupo de Investigación en Salud Electrónica y Supercomputación (GISES), Universidad Tecnológica de Panamá

[3]Facultad de Ingeniería de Sistemas Computacionales, Universidad Tecnológica de Panamá

[3]Facultad de Informática, Electrónica y Comunicación, Universidad de Panamá

Corresponding author:Miguel Vargas-Lombardo Ph.D, Universidad Tecnológica de Panamá, Panamá. Phone: , +507 63005543. E-mail: miguel.vargas@utp.ac.pa, ORCID ID: https://orcid.org/0000-0002-2074-2939.

doi: 10.5455/aim.2024.32.15-18.
ACTA INFORM MED. 2024, 32(1): 15-18
Received: JAN 15, 2024
Accepted: MAR 09, 2025

ABSTRACT

Background: SARS-CoV-2 is an infectious disease caused by the coronavirus that was first reported in December 2019 in China and immediately spread around the world causing a pandemic, which has caused countless deaths and cases in global health. Mental health has not gone untouched by this pandemic; due to the lockdown and the vast amounts of information disseminated, the Panamanian population has begun to feel the collateral effects. **Objective:** We propose classifying tweets using a machine learning (ML) and deep learning (DL) approach and pattern search to make recommendations to the emotional and psychological reactions of the Panamanian population. **Methods:** Our study has been carried out with a corpus in spanish extracted from X for the automatic classification of texts, from which we have categorized, through the ML&DL approach, the tweets about Covid-19 in Panama, in order to know if the population has suffered any mental health effects. **Results:** We can say that the ML models provide competitive results in terms of automatic identification of texts with an accuracy of 90%. **Conclusion:** X is a social network and an important information channel where you can explore, analyze and organize opinions to make better decisions. Text mining and patron search are a natural language processing (NLP) task that, using ML&DL algorithms, can integrate innovative strategies into information and communication technologies.

Keywords: NLP, SARS-CoV-2, ML&DL, opinion mining, mental health.

1. BACKGROUND

Since the beginning of the global crisis caused by the Covid-19 pandemic, many of our Central American countries have not been able to overcome the challenges that this pandemic has caused (1). For the population of Panama, these challenges are very traumatic and have implications mainly for physical health, but have also profoundly affected the mental health and well-being of patients. Today, with variants like Omicron, Delta and others, the outlook becomes even more discouraging. Many mental health problems have been reported around the world. However, mental health treatment for patients takes a backseat when resources are limited and other needs are prioritized (2).

Currently in the world, people live in fear and concern for their personal security and for the lack of effective treatment. To this is added the socioeconomic consequences caused by Covid-19, such as unemployment. These effects can be translated into a variety of emotional reactions (such as disgust, fear, sadness, anguish, worry), which are mental health conditions that represent a complex and multilevel problem that can have a negative impact on the Panamanian population, requiring the attention of health professionals or specialists and researchers (3).

There are investigations that have reported different mental health problems in the general population due to Covid-19. A study by Liang et al. (4) evaluated mental health among 584 young people using the General Health Survey (GHQ-12). In this study, 40% of participants had psychological problems, while 14% had symptoms of post-traumatic stress disorder (PTSD). A study by Lei et al. (5) used the Anxiety Self-Evaluation Scale (SAS) and the Depression Self-Evaluation Scale

(SDS) to evaluate the mental health status of 1593 surveyed 18-year-olds or more in southern China and found the prevalence of anxiety and depression. A study carried out by Lozano Vargas (6) to analyze the psychological impact of the virus in the city where it started, Wuhan, indicates that more than 50% of the population present symptoms related to anxiety and depression in different degrees and more than 70% of its people present symptoms of fear, worry and anxiety.

From the first moment, social networks have been loaded with opinions and messages about the coronavirus. Of these platforms X, formerly known as Twitter (7), has been the one with the most complete and relevant data and information (8). It is considered one of the most important data sources for infodemiology studies that involve the monitoring of public response (9). Analyzing a large amount of content on social networks is crucial as the opinions expressed by users online are a very valuable asset for organizations that seek to develop strategies based on public opinion. The activity of analyzing opinions from social networks is known as text mining or sentiment analysis, which deals with the processing of opinions, feelings and subjective expressions. Opinion mining is a task for NLP and computational linguistics (10).

2. OBJECTIVE

This study is carried out with the objective of examining the opinions, discussions, concerns and feelings of people, especially those who face problems or emotional reactions in the Panamanian population, using text analysis techniques.

3. MATERIAL AND METHODS

The methodology applied in this study is a mixed method with both quantitative and experimental aspects as a total of 150,000 opinions of Panamanians about the Covid-19 pandemic were extracted from the X platform. A software architecture based on several steps was developed: (1) obtaining a set of data; (2) carry out text preprocessing; (3) analyze the text; (4) apply classification algorithms to perform text polarity detection; and (5) patron search. Figure 1 shows the proposed architecture.

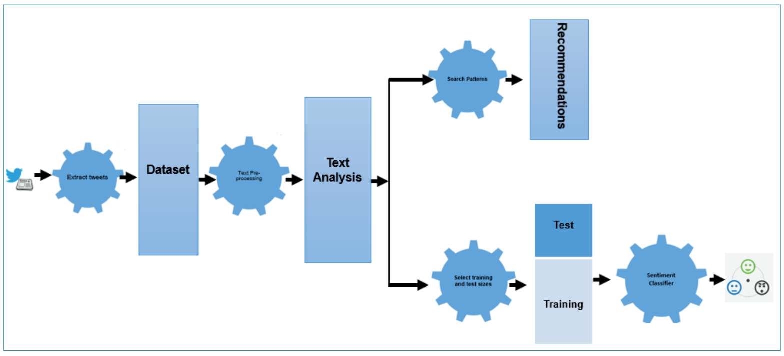

Figure 1. Overview of proposed Architecture

3.1. Dataset

We collected tweets from June 12, 2022, to November 30, 2022, using a Panama geolocation filter. To search for tweets on X, we used the keywords: "covid, coronavirus, disgust, fear, sadness, anguish, worry and pandemic". The corpus contains 150,000 tweets.

3.2. Text Analysis

The first 5,000 tweets of the dataset were preprocessed, labeled and manually annotated by two annotators to establish the sentiment labels and identify their polarity (positive or negative) in order to evaluate our proposed model. However, as this arduous task was time consuming, we proceeded to use SenticNet 5.0 (11), which is a knowledge base that classifies a text and detects its polarity with good results. Each tweet contains the following columns: sentiment_rate, prediction, sentiment and sentence. The sentence field includes the text of the tweet (12).

3.3. Sentiment Classifier

In this process, the positive or negative polarity of the opinion is determined. The performance of ML&DL models depends on the effectiveness of the chosen method for feature extraction (7). The next step in our pipeline was to extract the features to perform supervised classification. We rely on the bag of words (BoW) technique, which includes term counting functions used in traditional ML&ML classifiers (13) . We propose generating predictive models from ML&DL classifiers using techniques: Decision Trees (14), Support Vector Machines (15), Naive Bayes (16) and Logistic Regression classifiers (17). We also introduce a recurrent neuronal network (RNN) of the Long Short Term Memory (LSTM) type (18). To calculate the accuracy of the trained classifier, we have used the accuracy_score function of the Sklearn Python library, where we can also find methods to perform the metrics and be able to evaluate the models (19).

4. RESULTS

4.1. Polarity detection

The detection of the polarity of opinions on social networks has attracted increasingly greater attention in medical and social areas. In general, the results of our architecture are promising and confirm the effectiveness of our method to identify the polarity of sentiment in tweets about mental health conditions caused by Covid-19 in the Panamanian population. The metrics used for the evaluation were precision, recall and F-measure. This set of metrics is very common in the evaluation of NLP systems (20). Table 1. Classifier accuracy results.

Classifier	Precision	Recall	F-Measure
Support Vector Machines	87.14	86.234	86.131
Decision Trees	88.125	85.435	84.457
Naive Bayes	83.117	80.236	78.345
Logistic Regression	89.743	87.432	85.290
LSTM	90.102	88.323	87.986

Table 1. Classifier results

The analysis carried out was to detect polarity using ML&DL classifiers, obtaining percentages of sentiment from each sentence of tweets, in this case a 66.42% negative and 33.58% positive result was obtained. Regarding precision using metrics, we observed that using LSTM 90.102 we obtained the best results. However, we consider that all other algorithms were very good due to the percentage obtained.

4.2. Pattern Search

At the same time, in order to prove that the polarity detec-

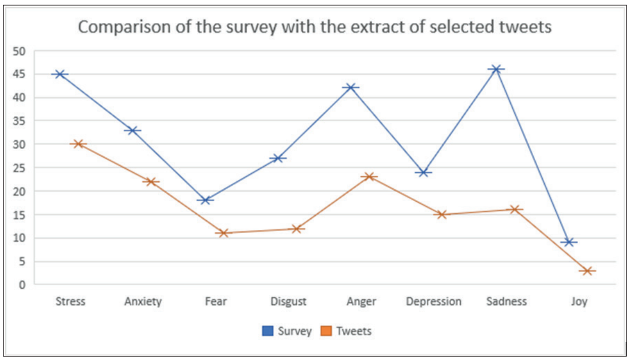

Figure 2. Results of the NLP analysis vs Survey.

tion process of our architecture was functional, we carried out a patron search process where we analyzed the text of tweets to find patrons that identified some mental health condition. We carried out an analysis of tweets with help from the Natural Language Understanding tool from IBM Watson, which is a free NLP-based artificial intelligence platform to help with research (21). To evaluate this pattern of search process, we randomly used a part of the data set.

Parallel to these processes (polarity detection and patron search) we carried out a survey between colleagues, friends, students from the Republic of Panama, to determine whether the proposed architecture is congruent. The survey applied to a sample of 100 people, between 18 and 60 years old, 44 men and 56 women. Establishing direct questions, for example if you have had any suffering during the pandemic related to problems such as: stress, anxiety, fear, disgust, disgust, depression, sadness and joy. You can select what you want. Figure 2 we can see an NLP analysis versus the survey carried out.

The results obtained in the survey that the highest percentage is in stress, anger and sadness. Many of those surveyed also have a higher percentage in fear and disgust. In both processes the results are very similar, regardless of the technique used, which reflects that in the Panamanian population there is a growing problem of mental health and emotional reactions to Covid-19, which should be alerted to those taking decisions in the country regarding health, especially mental health.

5. DISCUSSION

The contributions of this investigation are several: First, we proposed and developed a software architecture to classify the feelings of a set of textual data related to COVID-19 based on the ML&DL approach using several proprietary classification algorithms. This model also allowed us to search for patrons in tweets and relate them to problems of psychological emotional reactions in the Panamanian population as a result of the pandemic. Secondly, we carried out several experiments to compare the performance of this model. Third, this model serves as a basis for comparing future work in the country and establishing some strategies on aspects of mental health.

6. CONCLUSION

This work aims to examine the opinions, discussions, concerns and feelings of users, especially patrons who delimit problems of psychological emotional reactions in the population, using text analysis techniques for this purpose and using this analysis to have a conceptual framework to establish strategies of mental health problems during and after the COVID-19 pandemic.

- **Acknowlledgment:** We thanks to the Sistema Nacional de Investigación (SNI-Senacyt Panamá).
- **Authors's contribution:** Principal Author DCM; Conceptualization DCM, MVL, ADH; methodology DCM, MVL, ADH, NMH; formal analysis DCM; research DCM, ADH, NMH.; original-writing DCM, MVL; writing—review and edition DCM, MVL; Corresponding author MVL.
- **Conflicts of interest:** There are no conflict of interest.
- **Financial support and sponsorship:** None.

REFERENCES

1. Velavan TP, Meyer CG. The COVID-19 epidemic. Trop Med Int Heal. 2020; 25(3): 278–280.
2. Ranjan R. Omicron Impact in India: Analysis of the Ongoing COVID-19 Third Wave Based on Global Data. medRxiv [Internet]. 2022; (January 2022): 2022.01.09.22268969. Available from: http://medrxiv.org/content/early/2022/01/12/2022.01.09.22268969.abstract
3. Savira M, Ghifari M. The Relationship Between Decrease in Vision Aquity and Gadget Use During Online Lectures and COVID-19 Pandemic in Students of the Faculty of Medicine, Universitas Sumatera Utara. Acta Inform Med. 2023; 31(4): 240-243. doi: 10.5455/aim.2023.31-240-243.
4. Liang L, Ren H, Cao R, Hu Y, Qin Z, Li C. The Effect of COVID-19 on Youth Mental Health. 2020; (1163): 841–852.
5. Lei L, Huang X, Zhang S, Yang J, Yang L, Xu M. Comparison of Prevalence and Associated Factors of Anxiety and Depression among People Affected by versus People Unaffected by Quarantine during the COVID-19 Epidemic in Southwestern China. Med Sci Monit. 2020; 26: 1-12.
6. Lozano A. Impacto de la epidemia del Coronavirus (COVID-19) en la salud mental del personal de salud y en la población general de China. Revista de Neuro-Psiquiatria [revista en Internet] 2020 [acceso 10 de junio de 2020]; 83(1): 51-56. Rev Neuropsiquiátrica [Internet]. 2020; 83(1): 51–56. Available from: http://www.scielo.org.pe/scielo.php?pid=S0034-85972020000100051&script=sci_arttext
7. Reveilhac M, Morselli D. Dictionary-based and machine learning classification approaches: a comparison for tonality and frame detection on Twitter data. Polit Res Exch. 2022; 4(1).
8. Behl S, Rao A, Aggarwal S, Chadha S, Pannu HS. Twitter for disaster relief through sentiment analysis for COVID-19 and natural hazard crises. Int J Disaster Risk Reduct [Internet]. 2021; 55(January): 102101. Available from: https://doi.org/10.1016/j.ijdrr.2021.102101
9. Jiloha RC. COVID-19 and Mental Health. 2020; 5(1): 9–11.
10. Heinze DT, Morsch ML, Holbrook J. Mining free-text medical records. Proc AMIA Symp. 2001; 254–258.
11. Sasikala P, Sheela LMI. Comparative Study of Sentiment Analysis Techniques in Web. 2017; 8(5): 125–129.
12. Cedeno-Moreno D, Vargas-Lombardo M, Moctezuma D. Approach Based on Text Processing and Ontology Applied to the Health Domain of Palliative Care in Panama. Commun Comput Inf Sci. 2021; 1460 CCIS: 3–17.
13. Yang Q, Zhang Y, Dai W, Pan SJ. Transfer Learning in Natural Lan-

guage Processing. Transf Learn. 2020; (2010): 234–256.
14. Izza Y, Ignatiev A, Marques-Silva J. On Tackling Explanation Redundancy in Decision Trees. J Artif Intell Res. 2022; 75: 261–321.
15. Bansal M, Goyal A, Choudhary A. A comparative analysis of K-Nearest Neighbor, Genetic, Support Vector Machine, Decision Tree, and Long Short Term Memory algorithms in machine learning. Decis Anal J [Internet]. 2022;3(November 2021):100071. Available from: https://doi.org/10.1016/j.dajour.2022.100071
16. Deepa N, Sathya Priya J, Devi T. Towards applying internet of things and machine learning for the risk prediction of COVID-19 in pandemic situation using Naive Bayes classifier for improving accuracy. Mater Today Proc [Internet]. 2022;62:4795–9. Available from: https://doi.org/10.1016/j.matpr.2022.03.345
17. Huang RJ, Kwon NS-E, Tomizawa Y, Choi AY, Hernandez-Boussard T, Hwang JH. A Comparison of Logistic Regression Against Machine Learning Algorithms for Gastric Cancer Risk Prediction Within Real-World Clinical Data Streams. JCO Clin Cancer Informatics. 2022; (6): 7–10.
18. Dou G, Zhao K, Guo M, Mou J. Memristor-based LSTM network for text classification. Fractals. 2023; 31(6): 1–12.
19. Nallamani R. An Automated Conversation System Using Natural Language Processing (NLP) Chatbot in Python. Cent Asian J Med Nat Sci [Internet]. 2022;3(4):314–36. Available from: http://www.centralasianstudies.orghttp//cajmns.centralasianstudies.org
20. Lavazza L, Morasca S. Common Problems With the Usage of F-Measure and Accuracy Metrics in Medical Research. IEEE Access. 2023;11(May):51515–26.
21. Liaw ST, Liyanage H, Kuziemsky C, Terry AL, Schreiber R, Jonnagaddala J, et al. Ethical Use of Electronic Health Record Data and Artificial Intelligence: Recommendations of the Primary Care Informatics Working Group of the International Medical Informatics Association. Yearb Med Inform. 2020;29(1):51–7.