BUAP_01 at MentalRiskES 2024: Detection of Depression and Anxiety Using Features Based on Tagging

Erick Barrios González[1,*,†], Mireya Tovar Vidal[1,†]

[1] *Benemerita Universidad Autonoma de Puebla, Av. San Claudio y 14 sur, Ciudad Universitaria, 72592 Puebla, Pue., México*

Abstract

This work presents the participation of team BUAP_01 in the MentalRiskES 2024 competition, covering Task 1, which consists of detecting depression and anxiety in Telegram group messages. The task was addressed using features related to sentiment classification along with probabilistic features derived from information such as POS tagging, lemmatization, and message sending time. The results place the best model at position 17 using the macro-F1 metric to evaluate participants' results.

Keywords

Depression detection, anxiety detection, sentiment features

1. Introduction

It is estimated that approximately one in eight people worldwide faces some form of mental disorder, which causes notable alterations in cognition, emotional regulation, and behavior, with many of them associated with functional disabilities in various areas [13]. Among the most common mental disorders are anxiety and depression, although there are also others such as bipolar disorder, post-traumatic stress disorder, schizophrenia, etc. [3] mentions that currently it is estimated that approximately one billion people live with a condition that affects their mental health worldwide.

Suicide is closely related to mental health problems such as anxiety, depression, and chronic stress, with chronic stress being a significant risk factor [2] and depression being able to lead to suicide as mentioned in the work of [4]. According to a recent report by the World Health Organization, suicide is the fourth leading cause of death among young people aged 15 to 29 years. The organization considers early identification to be a key effective intervention in preventing these problems.

IberLEF 2024, September 2024, Valladolid, Spain
* Corresponding author.
† These authors contributed equally.
erick.barrios@viep.com.mx (E. Barrios); mireya.tovar@correo.buap.mx (M. Tovar)
0000-0002-2077-7541 (E. Barrios); 0000-0002-9086-7446 (M. Tovar)

The potential of social networks in the context of depression is highlighted by the fact that patients with this condition may feel more comfortable sharing episodes of their life and communicating with others through these platforms [10].

Anxiety disorders are characterized by excessive fear and worry, as well as the onset of other related disorders [14]. Depression, unlike a common mood disturbance, is characterized by a depressed mood or a loss of enjoyment or interest in activities, most of the day, almost every day, for at least two weeks [15].

This responds to a demand from society due to the significant increase in these problems among the population, in several types of mental risks: eating disorders, dysthymia, anxiety, depression, suicidal ideation, and others. Currently, relevant evaluation campaigns such as the Cross-Lingual Evaluation Forum (CLEF) have hosted during the last years the Early-Risk Identification task (eRisk). Unfortunately, these campaigns have focused mainly on English, leaving aside other languages, like Spanish.

MentalRiskES 2024 [8] of IberLEF 2024 [1] is a task focused on the early identification of mental disorder risks in Spanish social media comments. This is the second edition, following last year's inaugural event at the IberLEF evaluation forum as part of the SEPLN 2023.

MentalRiskES 2024 [8] tackled the challenge of detecting mental disorders in an online setting, aiming to identify potential risks as early as possible within a continuous data stream. As a result, success depended not only on the accuracy of the systems but also on their speed in detecting issues. The second edition introduces three tasks: the first involves identifying the disorder itself, the second entails recognizing contextual cues associated with the disorder, and the third focuses on detecting suicidal ideation.

In this work, characteristics based on feelings are explored. There are works such as [5] and [11] in MentalRiskES 2023[6], which obtain good results when using Pysentimiento [12] to detect depression.

In this work, the methodology will be shown first, followed by the results, and finally the conclusions.

2. Task description and Methodology

This section will describe the task addressed and the methodology used.

2.1. Task 1 - Disorder detection

The task 1, "Disorder detection"[8], consists of detecting whether a user is suffering from depression, anxiety, or none of the above. It is a multiclass classification task, with the three classes mentioned previously.

- Depression: A user Is perceived as suffering from depression when he or she expresses everyday situations, desires, or actions related to the suffering of this pathology.
- Anxiety: A user Is recognized as suffering from anxiety when he or she expresses everyday situations, desires, or actions related to the suffering of this pathology.
- None: The user does not present evidence of suffering from any of the aforementioned disorders.

The corpus is divided into two collections: one for tasks 1 and 2, and another for task 3. Each subcorpus has an average of 50 messages per user. The corpus for task 1 includes a total of 894 users (20 for trial, 464 for training, and 400 for testing).

In Table 1, the number of users per class in each corpus is shown. It can be observed that the classes are not balanced.

Table 1
Number of users per class

Corpus	None	Depression	Anxiety
Trial	10	5	5
Train	223	169	93
Test	200	100	100

2.2. Methodology

The following stages have been carried out for the completion of this work:

2.2.1. Features

A set of numerical features was created to address the detection of depression and anxiety. For this purpose, information that might be relevant to the detection of depression was collected. This information is described below.

1. Sentiment information: This information involves labeling a text as "Positive", "Negative", or "Neutral" sentiment. To achieve this labeling, Pysentimiento [12], a Python toolkit for Sentiment Analysis and Social NLP tasks, is used.
2. Average nocturnal activity of a user: Messages sent between 10 PM and 4 AM are found, and the average number of nocturnal messages per day for a specific user is obtained.
3. Average daily activity on social networks: This is the average number of messages per day for a specific user.
4. Average message length: The average length of all messages for a specific user is obtained from the lengths of these messages.
5. Difference in days (between the user's first and last message): This is the number of days between the user's first and last message.
6. Spacy information: "Spacy" and the "es_core_news_lg" model were used to obtain the tokenization, POS (Part-of-speech) labels, and lemma labels of each text. Each word is counted according to the number of appearances it has in each class to estimate the probabilities of appearance of each label per class.

Table 2 shows the characteristics used for each run in the competition. The combination of features was obtained from a grid search algorithm.

Table 2
Characteristics per run

Run/Feature	Sentiment information	Average nocturnal activity	Average daily activity	Average message length	Difference in days	Spacy Information
0	✓	✓				✓
1	✓	✓				✓
2	✓	✓	✓	✓	✓	

2.2.2. Models Development

A multilayer perceptron neural network was used as the classifier model, and 3 combinations of characteristics were selected for three runs. To adjust the hyperparameters of the multilayer perceptron network, a grid search has been performed with the training corpus.

Run 0: This proposal uses an average of the quantities of each Pysentimiento label obtained per user. This feature is combined with the average nocturnal activity and Spacy information. The configuration of the multilayer perceptron neural network is as follows:

- Activation = "logistic"
- Learning_rate_init = 0.002,
- Hidden_layer_sizes = (25, 29)
- Max_iter = 175
- Solver = 'adam'

Run 1: Unlike Run 0, this proposal uses Pysentimiento's tagging but represents the information according to the number of appearances per class (a probability of occurrence). The configuration of the multilayer perceptron neural network is as follows:

- Activation = "logistic"
- Learning_rate_init = 0.004,
- Hidden_layer_sizes = (25,27)
- Max_iter = 233
- Solver = 'adam'

Run 2: Proposal 2 uses an average of the Pysentimiento labeling combined with the average nocturnal activity, average daily activity on social networks, average message length, and difference in days. The configuration of the multilayer perceptron neural network was as follows:

- Activation = "logistic"
- Learning_rate_init = 0.0035,
- Hidden_layer_sizes = (25, 27)
- Max_iter = 143
- Solver = 'adam'

2.2.3. Evaluation

Task 1 will be evaluated from different perspectives: multi-class classification using the macro-F_1-score metric, and the ERDE30 [9] metric aimed at evaluating how quickly depression symptoms are detected. Cross-validation was used to select the models to participate. Due to the limited number of classified users with anxiety (98 users, 5 trials + 93 training), the data corpus was divided into two sets (set 1 and 2, see Table 3) for the corresponding tests.

Table 3
Results of the different sets created, using macro-F_1-score and standard deviation (Devest)

Run	Macro-F_1 Set 1	Macro-F_1 Set 2	Devest
0	0.4526	0.4732	0.0145
1	0.5805	0.5776	0.0020
2	**0.6328**	**0.6314**	**0.0009**

In Table 3, it can be observed that Run 2 seems to have very good results, followed by Run 1, because the standard deviation is lower and the results of the macro-F_1 metric are higher. Run 0 performed poorly because the input to the multilayer perceptron network had two different types of data (average data and percentage data). Run 1 does not have this problem and only uses probabilistic data. Run 2, has a greater number of features and had the best performance on the training set.

3. Results

The results obtained are shown in Table 4, where it can be observed that Run 1 had the best results out of the Run 0 and 2, which is very different from what is shown in Table 3. This is probably because the characteristics of Run 2 overfit the model for the training set.

Table 4
Classification-based evaluation in Task 1. Metric ranking: Macro-F1

Rank	Team	Run	Accuracy	Macro_P	Macro_R	Macro_F1
1	ELiRF-UPV	2	**0.890**	**0.875**	**0.880**	**0.874**
3	BaseLine - Roberta Base	2	0.853	0.840	0.843	0.834
14	BaseLine - Roberta Large	1	0.670	0.786	0.708	0.682

17	**BUAP_01**	**1**	**0.620**	**0.692**	**0.662**	**0.632**
18	BaseLine - mDeberta	0	0.710	0.748	0.645	0.623
23	**BUAP_01**	**0**	**0.427**	**0.650**	**0.557**	**0.411**
24	**BUAP_01**	**2**	**0.393**	**0.348**	**0.352**	**0.348**

Table 5 shows that run 1 performs better compared to the previous table, where it managed to position itself in 14th place. This is due to the evaluation based on the ERDE30 metric.

Finally, the carbon emissions for run 1 have a mean duration of 132.464 with a mean emission of 9.27E-04 and a deviation of 9.23E-04.

Table 5
Latency-based evaluation in Task 1. Metric ranking: ERDE30

Rank	Team	Run	ERDE5	ERDE30	latencyTP	speed	latency-weightedF1
1	BaseLine - Roberta Base	2	0.162	0.042	3	0.969	0.909
2	ELiRF-UPV	2	0.405	0.045	8	0.891	0.845
9	BaseLine - mDeberta	0	0.211	0.102	1	1.000	0.891
13	BaseLine - Roberta Large	1	0.205	0.133	1	1.000	0.811
14	**BUAP_01**	**1**	**0.282**	**0.134**	**3**	**0.969**	**0.769**
25	BUAP_01	0	0.272	0.240	1	1.000	0.676
26	BUAP_01	2	0.363	0.359	1	1.000	0.522

4. Conclusion

This work has reviewed the 3 proposals for Task 1 of MentalRiskES 2024 [8], which addresses the task of detecting depression and anxiety in Telegram message texts.

The purpose of this work was to explore the combination of different features primarily aimed at detecting depression. The results show that only 1 of the 3 proposals has had an acceptable performance, which is probably due to the features being selected based on a corpus that contained more users from the "depression" and "none" classes, which would lead to overfitting in the final model, despite cross-validation being performed to obtain reliable results.

As mentioned earlier, an area for improvement would be to balance the classes before adjusting the model parameters. Similarly, no features related to the detection of depression were included, which could have affected the model. These aspects are expected to be addressed in future work.

5. Acknowledgments

The authors would like to thank Benemerita Universidad Autonoma de Puebla. The present work was funded by the research project 00082 at VIEP-BUAP 2024 and by the Consejo Nacional de Humanidades de Ciencia y Tecnologia (CONAHCYT) with CVU number 1067846.

References

[1] Chorizo, L., Jiménez-Zafra, S. M., Rangel, F., Overview of IberLEF 2024: Natural Language Processing Challenges for Spanish and other Iberian Languages, in: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2024), co-located with the 40th Conference of the Spanish Society for Natural Language Processing (SEPLN 2024), CEUR-WS.org, 2024

[2] García-Dolores, F., Tendilla-Beltrán, H., Flores, F., Carbajal-Rimoldi, L. A., Mendoza-Morales, R. C., Gómez-Mendoza, L. E., Vázquez-Hernández, A. J., De La Cruz, F., Genis-Mendoza, A. D., Nicolini, H., & Flores, G., Increased suicide rates in Mexico City during the COVID-19 pandemic outbreak: An analysis spanning from 2016 to 2021. Heliyon, 9(6), 2023, e16420. https://doi.org/10.1016/j.heliyon.2023.e16420

[3] Garg, M., Mental health analysis in social media Posts: a survey. Archives of Computational Methods in Engineering, 30(3), 2023, pp. 1819-1842. https://doi.org/10.1007/s11831-022-09863-z

[4] Gómez-Peresmitré, G., Acevedo, S. P., Hernández, R. C. L., Pineda-García, G., & Guzmán-Saldaña, R., vPsychometric assessment of an online Self-Test measuring risky eating behavior, depression, social Anxiety, and Self-Injury in Mexican adolescents. International Journal of Environmental Research and Public Health, 20(1), 2022, pp. 399 https://doi.org/10.3390/ijerph20010399

[5] González-Silot, S., Martınez-Cámara, E., & Ureña-López, L. A., SINAI at MentalRisk: Using Emotions for Detecting Depression. In *IberLEF (Working Notes). CEUR Workshop Proceedings*, 2023.

[6] Mármol-Romero, A. M., Moreno-Muñoz, A., Plaza-del-Arco, F. M., Molina-González, M. D., Martín-Valdivia, M. T., Ureña-López, L. A., & Montejo-Raéz, A, Overview of MentalriskES at IberLEF 2023: Early Detection of Mental Disorders Risk in Spanish. Procesamiento del Lenguaje Natural, 71, 329-350, 2023.

[7] Mármol-Romero, A. M., Moreno-Muñoz, A., Plaza-del-Arco, F. M., Molina-González, M. D., Martín-Valdivia, M. T., Ureña-López, L. A., & Montejo-Raéz, A., MentalRiskES: A New Corpus for Early Detection of Mental Disorders in Spanish. In Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), ELRA and ICCL, 2024.

[8] Mármol-Romero, A. M., Moreno-Muñoz, A., Plaza-del-Arco, F. M., Molina-González, M. D., Martín-Valdivia, M. T., Ureña-López, L. A., Montejo-Ráez, A., Overview of MentalRiskES at IberLEF 2024: Early Detection of Mental Disorders Risk in Spanish. Procesamiento del Lenguaje Natural, vol 73, 2024.

[9] Losada, D. and F. Crestani., A test collection for research on depression and language use. volume 9822, 2016, pp. 28–39, 09.

[10] Park, M., McDonald, D. W., & Cha, M., Perception differences between the Depressed and Non-Depressed users in Twitter, in: Proceedings of the International AAAI Conference on Web and Social Media, 7(1), 2021, pp. 476-485. https://doi.org/10.1609/icwsm.v7i1.14425

[11] Pan, R., Garcıa-Dıaz, J. A., & Valencia-Garcıa, R., UMUTeam at Mental-RiskES2023@ IberLEF: Transformer and Ensemble Learning Models for Early Detection of Eating Disorders and Depression. In IberLEF (Working Notes). CEUR Workshop Proceedings, 2023.

[12] pypi.org, pysentimiento, 2024. URL: https://pypi.org/project/pysentimiento/

[13] World Health Organization (WHO), 2022. Trastornos mentales. URL: https://www.who.int/es/news-room/fact-sheets/detail/mental-disorders
[14] World Health Organization (WHO), 2023a. Depresión. URL: https://www.who.int/es/news-room/fact-sheets/detail/depression
[15] World Health Organization (WHO), 2023b. Trastornos mentales. URL: https://www.who.int/es/news-room/fact-sheets/detail/mental-disorders