SINAI at MentalRisk: Using Emotions for Detecting Depression

Santiago González-Silot[1,†], Eugenio Martínez-Cámara[2,**,†] and Luis Alfonso Ureña-López[2,†]

[1]MSc Student, University of Granada, Spain
[2]Departamento de Informática, CEATIC, Universidad de Jaén, Spain

Abstract
Mental health is progressively reaching the attention of society because mental disorders provoke devastating consequences in the patients, their families and social circle. The preemptive detection of mental disorders, and in particular, depression disorders, may help its medical treatment. MentalRisk proposes three challenges related to the early detection of mental disorders. The SINAI team has participated in the task 2.a that consists in the detection of signals of depression in the Telegram messages of a set of users. We claim that the messages of people with depression have a relevant emotional burden. Accordingly, we follow a fine-tuning approach of the BETO Spanish language model pre-trained on a dataset of emotions. One of our submitted systems was reached the first position in the early detection of depression, and the fifth position according to an standard text classification evaluation. Hence, the use of emotional knowledge enhances the detection of depression.

Keywords
Depression detection, emotion analysis, adaptive fine-tuning, transformers-based language models

1. Introduction

Mental health has climbed ranks in social concerns in recent years, since the presence of any kind of mental disorder have personal and social costs [1], as well as disastrous consequences for the patients and their families [2]. Among the wide range of mental disorders, depression is the most prevalent mental disorder worldwide with an estimated population affected of 280 million people [3]. Moreover, depression may lead suicide ideation and attempt [3]. Accordingly, depression has become in a real public health problem worldwide.

People with depression often use social media to talk about their feelings, disorder, treatments, share information, reduce social isolation and manage their suffering [4, 5]. This use of social media may help in the early detection of depression, since there are signals of depression on the use of language of people with this mental disorder [6, 7]. In this context, the IberLEF evaluation

IberLEF 2023, September 2023, Jaén, Spain
** Corresponding author.
† These authors contributed equally.
sgonzalezsilot@gmail.com (S. González-Silot); emcamara@ujaen.es (E. Martínez-Cámara); laurena@ujaen.es (L. A. Ureña-López)
0000-0001-8378-5840 (S. González-Silot); 0000-0002-5279-8355 (E. Martínez-Cámara); 0000-0001-7540-4059 (L. A. Ureña-López)

campaign includes in its 2023 edition the MentalRisk task [8], which is focused on the detection of cues of mental disorders on messages from Telegram users written in Spanish.

MentalRisk proposes three tasks related to two mental disorders and the detection of any evidence that an user is suffering any kind of mental disorder. The first challenges ask for identifying users that express any evidence of any kind of eating disorders, like anorexia or bulimia. The second challenge is focused on detecting users with any susceptible cue of suffering depression. However, the real challenge of MentalRisk lies in the evaluation of the early detection of those disorders, which forces the systems to really discover the underlying signals of a mental disorder as soon as possible. This early detection is related to the fact that the diagnosis of depression requires some time of suffering this disorder, at least two weeks according to the Depression Symptom Detection (DSD) model [9]. Additionally there are some evidences that the alterations of the mood linked to depression are evident since several weeks before the diagnosis [10, 11].

The team SINAI has participated in the task of binary classification of depression detection. Albeit Leis et al. [5] show the linguistic features that featured depression in tweets written in Spanish, we argue that transformed-based language models (LM) may provide a more precise preemptive detection of depression, since they have proven their effectiveness in previous works [11, 12]. Likewise, we claim that messages from users susceptible of suffering depression have a substantial emotional burden, since depression is related to emotional disorders and feelings of sadness, grief and bereavement [13], and previous works have shown the effectiveness of using emotions to detect depression disorders [14]. Accordingly, we argue that transformed-based models further trained on texts that express emotions may enhance the early detection of depression.

We submitted two transformed-based language models for the task of binary classification of depression. The first system (SINAI-SELA-r0) is built upon the fine-tuning of the LM BETO Emotion Analysis [15], which is further pre-trained on the dataset for emotion analysis EmoEvent dataset [16]. Our second submitted system (SINAI-SELA-r1) attempts to wider the coverage of use of language of patients with depression of the BETO Emotion Analysis LM. Accordingly, we further pre-trained the LM with tweets written in Spanish by depression patients from the Depressive Users (DU) dataset [5] to adapt the LM to the depression domain. Subsequently, we fine-tuned the LM on the dataset of the task.

The official results of the 2.a task of MentalRisk put the SINAI-SELA-r0 system in the first position and the SINAI-SELA-r1 in the fifth position according to the early detection evaluation [8]. Likewise, both systems are in the fifth and eighth positions respectively according to an standard binary classification evaluation. These strong results means that our claim related to the emotional burden of messages written by susceptible depressive patients holds.

The rest of the paper is structured as follows: Section 2 presents the exploratory analysis performed of the data. Then, section 3 describes the failed attempts that helped us to arrive to our final claim and strong results. Section 4 presents the two submitted systems. We expose the results and their analysis in section 5, and we finally present the conclusions in Section 6.

2. Analysis of data

The organisation provided a dataset per each mental disorder. We have only worked with the dataset of depression disorder, which we called MentalRisk Depresion dataset. This dataset is composed of Spanish messages published on public groups of the Telegram platform related to mental disorders topics. The messages were annotated by ten non-specialised in mental health annotators. The gold standard for the classification tasks were calculated by majority vote, and the individual annotations of the annotators are not provided.

The MentalRisk Depression dataset is composed of 6248 messages from 335 users. Table 1 shows some statistics of the MentalRisk Depression dataset, where there is not any significant difference among depression and non-depression users. The messages were anonymised to preserve user privacy. The messages show some characteristics of the social media genre [17], such as misspelled words, colloquial language and informal abbreviations like "q". Likewise, the emoticons of the original messages were replaced by the organisation by their textual representation, which has constrained the use of emoticons as a classification signal.

Table 1
Statistics of the MentalRisk Depression dataset. The table does not incorporate the statistics of the 10 messages used for the evaluation competition trial.

		Users	**Messages**	**Average messages per user**
Depression	Train	94	3113	33.12
	Test	68	2339	34.40
Non-depression	Train	81	3135	38.70
	Test	81	2825	34.88

Before designing of the classification model, we conducted an exploratory analysis of the dataset consisted in a lexical frequency and length analysis that we exposed as what follows:

Lexical frequency analysis We studied the most frequent words of the dataset in order to study if there are any lexical features that may be leveraged in the early detection of depression. Accordingly, we show in Table 2 the twenty most salient words according to absolute frequency and TF-IDF. We see in the Table some words related to (1) the problem domain, as "depresion", "vida" or "personas"; (2) emotions, as "quiero", "siento" or "llorando" and (3) sentiments or feelings, as "mal", "sonriendo" or "gracias". If we compare the words of emotions and sentiments, they are very similar, so we see at lexical level that the messages of the users has some degree of emotional burden that matches our claim related to the emotional underlying meaning of messages related to depression. Additionally, we assert that the decision of the organisation of replacing the emoticons by their textual representation has distorted in some sense the lexical analysis, because some words are related to the textual representation of the emoticons, as "cara" (face) or "ojos" (eyes).

We finally want to also highlight that we see some expressions more common in the American versions of Spanish than the Spanish language spoken in Spain, which means that the dataset

Table 2
Most salient words of the MentalRisk Depression dataset.

Dataset		Depression		Non-depression	
Frequency	TF-IDF	Frequency	TF-IDF	Frequency	TF-IDF
cara (face)	cara (face)	cara (face)	cara (face)	cara (face)	cara (face)
q (what)	día (day)	q (what)	siento (feel)	q (what)	risa (laugh)
vida (life)	gracias (thanks)	siento (feel)	gracias (thanks)	risa (laugh)	llorando (crying)
día (day)	llorando (crying)	mal (bad)	día (day)	llorando (crying)	vida (life)
llorando (crying)	vida (life)	día (day)	depresión (depression)	vida (life)	día (day)
risa (laugh)	risa (laugh)	grupo (group)	grupo (group)	ojos (eyes)	sonriendo (smiling)
mal (bad)	mas (more)	quiero (want)	mal (bad)	día (day)	mas (more)
grupo (group)	mal (bad)	vida (life)	quiero (want)	personas (people)	mal (bad)
mas (more)	siento (so)	mas (more)	mas (more)	sonriendo smiling	alguien somebody
ojos (eyes)	grupo (group)	años (years)	vida (life)	gente (people)	personas (people)
personas (people)	alguien (somebody)	depresión (depression)	gente (people)	corazón (heart)	corazón (heart)
gente (people)	depresión (depression)	personas (people)	alguien (somebody)	mas (more)	gusta (like)
sonriendo (smiling)	gente (people)	gente (people)	años (years)	grupo (group)	gracias (thanks)
siento (so)	sonriendo (smiling)	alguien (somebody)	hola (hello)	mal (bad)	ojos (eyes)
alguien (somebody)	corazón (heart)	ojos (eyes)	sé (know)	alguien (somebody)	grupo (group)
años (years)	quiero (want)	sonriendo (smiling)	corazón (heart)	Hola (Hello)	gente (people)
corazón (heart)	hola (hello)	ayuda (help)	feliz (happy)	gusta (like)	hola (hello)
quiero (want)	personas (people)	sé (know)	personas (people)	persona (people)	frío (cold)
Hola (Hello)	años (years)	Hola (Hello)	favor (please)	años (years)	asi (so)

has a global representation of the Spanish language.

Length of the messages We use all the messages of each user to decide whether suffers of depression, instead of processing each individual message. Additionally, we argue that LMs may provide a more precise early detection of depression. Most LMs set a maximum number of

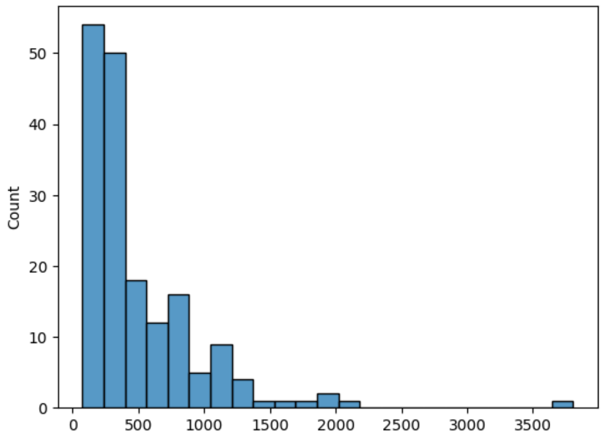

Figure 1: Histogram of the length of the messages of each user.

input tokens, which force us to study the length distribution of messages in order to assure the processing of most of them of each user.

Figure 1 shows the distribution of the length in number of tokens of the messages of the users. We see that there some users that the length of all their messages goes beyond of 500 tokens, which is usually the maximum allowed length (512). This constraint may lead to discard the processing of some messages, which may implies a loss of information and a reduction of the depression detection capacity. Accordingly, we study of filtering out messages with a reduce burden of emotional and sentiment meaning, which we present in Section 3.

3. Depression Detection needs all the user data

Since the length of all the messages of some user exceeds the limit of the standard number of tokens of LLMs, we studied whether discarding some specific messages may, on one hand, enhance the detection of depression, and, on the other hand, surpass the length processing constraint. Accordingly, we analyse to filter out the messages with neutral sentiment and to summarise the messages of each user.

Removing of the messages with neutral sentiment We claim that patients of depression tend to write messages with some grade of emotional or sentiment meaning. Hence, we use the `Pysentimento` toolkit [15] to classify the sentiment of each message. The sentiment classification model of `Pysentimento` is based on the BETO LM [18] further pre-trained on the Twitter Spanish sentiment analysis dataset InterTASS [19]. We used a LM models based on Twitter messages because the genre of tweets is similar to the Telegram messages of the MentalRisk Depression dataset. The `pysentimento` classifies tweets as positive, negative and neutral, and we decided to discard all the messages with a neutral sentiment.

The removal of the messages with a neutral meaning reduces the average length of user messages to 302 tokens surpassing the processing length constraint. We fine-tuned the sentiment `Pysentimiento` LM on the MentalRisk Depression dataset with and without neutral messages. The model is optimised using Adam with a learning rate of $2e{-}5$, an epsilon of 1e-8, max sequence length of 512 tokens and a batch size of 8. The model with neutral messages performed better than the one without neutral messages, in particular the model with neutral messages achieved 0.8571 of Accuracy and the one without neutral messages 0.7714 of Accuracy on the validation set. Therefore, the removal of neutral messages harmed the system, which may be due to breaking the contextual meaning of the entire discourse of the user.

Summarising the messages We also studied reducing the length of the messages by only using the most salient information. We thus summarised the messages of the users. We use a RoBERTa model fine-tuned for Spanish summarisation [20].

We again applied fine-tuning with the same hyperparameters as in the previous case to see the performance of this approach and obtained as a result 0.8286 of Accuracy on the validation set. This result is better than removing neutral messages, but it is still under the performance of using all the messages. We think that the lack of an strong connection among the messages of each user limits the existence of a real context that would allow to summarise the most prominent information.

After this first analysis, we conclude that we need to process all the messages of the users to identify whether they have signals of mental disorders. However, we still have the length constraint impose by most of LMs. Hence, we decide to use the first 512 tokens of all the messages of each user.

4. Using Emotions for detecting Depression

We present in this section the two models submitted to the evaluation of the task 2.a. Both models are based on our claim that the messages written by people with depressions or any initial symptoms of depression have a greater emotional burden than healthy people. Hence, we need to transfer knowledge from the task of emotion analysis to depression detection. We transfer that knowledge by fine-tuning the LM BETO Emotion Analysis [15], which is the BETO LM trained on the emotion analysis dataset of tweets written in Spanish EmoEvent dataset [16]. Likewise, we group the messages of each user, and we process them as they were a unique document. We use TensorFlow and the Huggingface libraries for the implementation of the models [21]. We subsequently present the specific details of the two models.

SINAI-SELA-r0 - Emotion knowledge The EmoEvent dataset is a dataset of 8409 tweets written in Spanish, which first adapt the BETO LM to a similar genre of text of the target dataset, Telegram messages. Likewise, the EmoEvent dataset is annotated at the six Ekamn's basic emotions (anger, disgust, fear, joy, sadness and surprise). As we show in Section 2, some of those emotions are prevalent in the MentalRisk Depression dataset.

The fine-tuning of the model on the target data was conducted using the Adam optimiser with a value of $2e^{-5}$ of learning rate, a weight decay value of 0.01, an epsilon value of $1e^{-8}$, an input maximum length sequence of 512 tokens and batch size of 8.

SINAI-SELA-r1 - Depression knowledge Additionally to incorporate emotion knowledge to the model, we also attained to incorporate depression knowledge to it. Accordingly, we followed an adaptative fine-tuning approach [22] to adapt the BETO Emotion Analysis LM to the language distribution of the target domain, i.e. depression speech. We use the Depressive User Dataset [5], which is composed of the timelines of tweets written in Spanish of 90 Twitter users that explicitly recognise that they are depression patients.

The adaptive fine-tuning was conducted using the same training goal as the original LM, i.e. a masked language goal. This unsupervised training consisted of 3 epochs using a value of learning rate of $2e^{-5}$ and a weight decay value of 0.01 (using the 10% of the dataset as validation set). Once the unsupervised learning was performed, the same fine-tuning process of the SINAI-SELA-r0 system was conducted.

Table 3 shows the results reached by the submitted systems and other attempts (see Section 3) on the validation set provided by the organisation. We see that the incorporation of emotion

Table 3
Evaluation of the models on the validation set according to Accuracy.

	Accuracy
Without neutral sentiment messages	77.14%
Summarising messages	82.86%
SINAI-SELA-r0	94.00%
SINAI-SELA-r1	91.43%

knowledge enhances the classification of the depression in the validation set.

5. Results and Analysis

We submitted the predictions of the two systems following the evaluation rules of the shared task. The evaluation of task 2.a consisted in the iterative prediction of the messages of users of the test set whether they give any evidence if the user is suffering depression. The systems have been ranked according to their global classification capacity and their ability of preemptive detection of depression.

Latency-based evaluation (ERDE-30). Table 4 shows the ranking of the systems according to the evaluation measure ERDE30 [8]. i.e. the early detection of depression. The SINAI-SELA-r0 has reached the first position, which means that the use of emotion knowledge is essential for the preemptive detection of depression. The adaptive fine-tuning applied in the SINAI-SELA-r1 was not performed as well as we expected, although it is in the top-five of best systems with the same ERDE30 value than one of the baselines provided by the organisation.

Table 4
Official results of the early detection of depression evaluation of task 2.a. The systems were ranked according to ERDE30.

Rank	Team	Run	ERDE5	ERDE30	latencyTP	speed	latency-weighted F1
1	**SINAI-SELA**	0	0,395	**0,140**	4,000	0,951	**0,720**
2	UNSL	1	0,567	0,148	14,000	0,791	0,609
3	BaseLine - Deberta	0	0,303	0,153	**2,000**	**0,984**	0,719
4	BaseLine - Roberta Large	1	0,290	0,159	4,000	0,951	0,704
5	SINAI-SELA	1	0,389	0,159	4,000	0,951	0,696
6	TextualTherapists	1	0,421	0,161	7,000	0,903	0,682
7	TextualTherapists	0	0,342	0,168	3,000	0,967	0,696
8	VICOM-nlp	2	**0,275**	0,173	**2,000**	**0,984**	0,706
9	CIMAT-NLP-GTO	0	0,423	0,175	5,000	0,935	0,665
10	BaseLine - Roberta Base	2	0,342	0,176	4,000	0,951	0,671

Table 5
Classification-based evaluation in Task 2.a. (10 first) Metric ranking: Macro-F1.

Rank	Team	Run	Acuracy	Macro-P	Macro-R	**Macro-F1**
1	UMUTeam	0	**0.738**	0.756	0.749	**0.737**
2	UNSL	1	**0.738**	**0.791**	**0.756**	0.733
3	UNSL	0	0.732	0.752	0.742	0.731
4	TextualTherapists	1	0.732	0.766	0.746	0.729
5	**SINAI-SELA**	0	0.725	0.775	0.742	0.720
6	UMUTeam	1	0.705	0.714	0.712	0.705
7	BaseLine - Roberta Large	1	0.698	0.759	0.718	0.690
8	SINAI-SELA	1	0.685	0.751	0.705	0.675
9	TextualTherapists	0	0.664	0.740	0.687	0.651
10	NLP-UNED	1	0.651	0.674	0.664	0.648

Standard classification-based evaluation (Macro-F1). Table 5 shows the ranking according to an standard evaluation based on Macro-F1 score. In this case, the SINA-SELA-r0 is ranked in the fifth position and the SINA-SELA-r1 in the eighth position. According to Macro-F1 the differences with the best ranked systems are not substantial, hence, again, the results show that our claim holds.

We highlight the fact that we failed in the first attempts of sending the predictions of the test data. That failed resulted in the loss of the opportunity of predicting the first messages of the users. Once the organisers have delivered the test data, we evaluated our two submitted systems with the entire test dataset and both of them achieved better results. In particular, the SINAI-SELA-r0 and SINAI-SELA-r1 reached a Macro-F1 of 0.7581 and 0.7223 respectively, which would have meant a higher position in the standard classification evaluation, namely the first and sixth position.

5.1. Analysis of the results

Since our two submitted systems are based on the incorporation of emotion knowledge into the classification system, we also evaluated the performance of the system without that knowledge, in a such a way of an ablation test. We predicted the test data with a system based only on the fine-tuning of the BETO LM. The result is 0.6691 according to Macro-F1, which is a worse performance than the models SINAI-SELA-r0 and SINAI-SELA-r1. Therefore, the incorporation of emotion knowledge contribute to the detection of depression.

5.2. Error analysis

We conduct an error analysis to learn the reasons behind the miss-classifications of the SINAI-SELA-r0 model. From the 150 test users, the SINAI-SELA-r0 model miss-classified 36 users. Likewise, those errors of the system are with a high probability value, which means that the system is sure of the decision although it is failing.

We read the messages of the miss-classified users, and we find out three main categories of errors:

Depression messages about other users Some users arrive to this kind of mental disorders social groups for sharing experiences and asking recommendations for helping people with depression among their relatives, friends or colleagues. Hence, they do not write in first person, but in third person. This is in line with the conclusions arrived in [5], where the authors stand out that patients of depression usually write in first-person singular. Therefore, the identification of depression may require at least a morphological analysis that allows to identify the pos-tags of the words and other lexical categories that may be cues of the presence of depression.

Misleading emotions Although the messages of users that are suffering depression has an emotional burden, not all the messages that express emotions are written by people with depression. This is evident in the MentalRisk Depression dataset, because there are users that express some events that provoke them negative emotions, but they are not suffering depression, at least according to the annotation of the dataset. Therefore, we suggest that the detection of depression has to take into account if the expression of negative emotions is maintained over time, which is in line with the definition of depression of the American Psychiatry Association [13].

Technical constraints Since we have use LM that limits the input size length to 512 tokens, we find errors in the classification in those users whose set of messages have a length greater than 512 tokens. Hence, we will have to work on strategies to process messages larger than the limitation of the standard LM.

Table 6 shows an example of each category of errors that we found.

The main insight of the error analysis is similar to other natural language processing tasks, or in other words, albeit LMs provide strong baselines, the incorporation of linguistic knowledge and contextual knowledge is still needed in other to really learn and understand the entire semantic meaning of a set of messages.

5.3. Efficiency Analysis

The strong results reached by our two submitted systems according to standard classification evaluation measures are not in line with the efficiency evaluation. Albeit, we did not use too much computational electronic components (only one Tesla T4 GPU), the different efficiency metrics show that our systems have a substantial carbon footprint. The reason behind this result is that the backbone of our systems are a transformer-based language model, which are known to a large consume of computational resources. Therefore, we need to make an effort to develop and use more efficient transformer-based models in order to reach the goal of green artificial intelligence systems [23].

6. Conclusions

The SINAI team participated in the task 2.a of MentalRisk shared-task with two classification system of depression. Both systems are built upon the claim that messages of people with depression have a relevant emotional burden. Accordingly, the two submitted systems are

Table 6
Examples of miss-classification users according to three error types identified.

Type of error	User	Messages	Prediction
Depression messages about other users.	9	Hola necesito consejos tengo una hija con depresion y estos días anda muy mal gracias. Es difícil esta enfermedad lo siento mucho. Y esta tomando medicamento pero la vd no vemos cambios. Esta unos días bien y otros días mal. Pero esta semana esta peor y mañana es su cumpleaños... Ánimo chicas ya les llegará su príncipe azul. Ustedes crean en ustedes como mujeres. Entonces el pelao arrumbalo y se feliz. 38 pero me siento de 30. B días que tengan un buen dia.	`Non-Depression`
Misleading emotions.	444	Ayer perdí un examen bro. Y era una materia que mas le entendia. Estoy en semana de exámenes. Fue el primero n. Si bro mañana será otro día. Acá escuchan canserbero no se pero el tiene tema que relajan. Solo en mi opinión. Es duro por lo que pasas. Y cuando salen mál las cosas es peor. Claro acá se trata de ayudarnos. Y expresar lo que sentimos. El rap es muy bueno. La mayoría de mis problemas está . En los estudios y la familia no se por que no hay comunicación.	`Non-Depression`
Technical constraints.	86	La soledad y el tiempo son el mejor aleado para encontrar paz. La soledad bueno la soledad refiere a que necesitas estar solo para poder encontrarte a ti mismo si no estás solo no lo logras. Si somos muy necesarios pues solos no podemos muchas gracias de verdad gracias por leer y entender. **The whole example cannot be displayed as the text is too large, 2500 tokens.** Este es el link. Muchas gracias saludos no he podido conectarme por otras cuestiones pero muchas gracias.	`Non-Depression`

Spanish LM pre-trained on texts annotated at emotion level, in particular on the EmoEvent dataset. The second system (SINAI-SELA-r1) followed an adaptive fine-tuning approach, and it was also pre-trained on the Depression User Dataset [5] in order to incorporate the domain knowledge into the LM.

The first submission (SINAI-SELA-r0) reached the first position in the preemptive detection of depression, which shows that the claim of our participation holds. The system adapted to the depression domain (SINA-SELA-r1) also was ranked as the top-5 best systems, which is an additional evidence of the importance of incorporation knowledge from the domain.

We also conducted an error analysis which allow to arise the following insights: (1) the detection of depression has to take into account the person and number of the subject of verbs in order to differ among those users that speak about personal problems (first person) and those ones that speak about the problems of other people (second and third person); (2) we need to consider the maintaining of the depression signals on time for differ of users that some time write messages with negative emotions; and (3) to surpass the technical limitation of length size processing constraint of LM.

We will continue working in this problem by studying how to incorporate morphological information in the classification of messages with depression to surpass the errors related to people that speak about other people.

7. Acknowledgements

We thank to Angela Leis and the rest of the authors of [5] by sharing with us their three datasets of tweets with signals of depression. Likewise, this work has been supported by the R&D&I grants PID2021-122263OB-C21, PID2020-116118GA-I00 and TED2021-130145B-I00 funded by MCIN/ AEI/10.13039/501100011033.

References

[1] A. Wongkoblap, M. A. Vadillo, V. Curcin, Researching mental health disorders in the era of social media: systematic review, Journal of medical Internet research 19 (2017) e228.

[2] D. Vigo, G. Thornicroft, R. Atun, Estimating the true global burden of mental illness, The Lancet Psychiatry 3 (2016) 171–178. doi:https://doi.org/10.1016/S2215-0366(15)00505-2.

[3] W. H. Organization, Depressive disorder (depression). key facts., https://www.who.int/news-room/fact-sheets/detail/depression, 2023. [Accessed 2023-06-08].

[4] J. A. Naslund, K. A. Aschbrenner, G. J. McHugo, J. Unützer, L. A. Marsch, S. J. Bartels, Exploring opportunities to support mental health care using social media: A survey of social media users with mental illness, Early intervention in psychiatry 13 (2019) 405–413.

[5] A. Leis, F. Ronzano, M. A. Mayer, L. I. Furlong, F. Sanz, Detecting signs of depression in tweets in spanish: behavioral and linguistic analysis, Journal of medical Internet research 21 (2019) e14199.

[6] M. De Choudhury, S. Counts, E. Horvitz, Social media as a measurement tool of depression in populations, in: Proceedings of the 5th Annual ACM Web Science Conference, WebSci

'13, Association for Computing Machinery, New York, NY, USA, 2013, p. 47–56. doi:10.1145/2464464.2464480.
[7] G. Coppersmith, M. Dredze, C. Harman, Quantifying mental health signals in Twitter, in: Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality, Association for Computational Linguistics, Baltimore, Maryland, USA, 2014, pp. 51–60. URL: https://aclanthology.org/W14-3207. doi:10.3115/v1/W14-3207.
[8] A. M. Mármol-Romero, A. Moreno-Muñoz, F. M. Plaza-del-Arco, M. D. Molina-González, M. T. Martín-Valdivia, L. A. Ureña-López, A. Montejo-Ráez, Overview of MentalriskES at IberLEF 2023: Early Detection of Mental Disorders Risk in Spanish, Procesamiento del Lenguaje Natural 71 (2023).
[9] F. Torres, What is depression?, 2020. https://www.psychiatry.org/patients-families/depression/what-is-depression [Accessed: 2023/07/26].
[10] I. A. van de Leemput, M. Wichers, A. O. Cramer, D. Borsboom, F. Tuerlinckx, P. Kuppens, E. H. van Nes, W. Viechtbauer, E. J. Giltay, S. H. Aggen, et al., Critical slowing down as early warning for the onset and termination of depression, Proceedings of the National Academy of Sciences 111 (2014) 87–92.
[11] D. Owen, D. Antypas, A. Hassoulas, A. F. Pardiñas, L. Espinosa-Anke, J. C. Collados, et al., Enabling early health care intervention by detecting depression in users of web-based forums using language models: Longitudinal analysis and evaluation, JMIR AI 2 (2023) e41205.
[12] S. G. Burdisso, M. L. Errecalde, M. Montes y Gómez, Using text classification to estimate the depression level of reddit users, Journal of Computer Science & Technology 21 (2021).
[13] A. P. Association, What is depression?, https://www.psychiatry.org/patients-families/depression/what-is-depression, 2020. [Accessed 2023-06-09].
[14] L. Ren, H. Lin, B. Xu, S. Zhang, L. Yang, S. Sun, Depression detection on reddit with an emotion-based attention network: algorithm development and validation, JMIR Medical Informatics 9 (2021) e28754.
[15] J. M. Pérez, J. C. Giudici, F. Luque, pysentimiento: A python toolkit for sentiment analysis and socialnlp tasks, 2021. arXiv:2106.09462.
[16] F. M. Plaza del Arco, C. Strapparava, L. A. Urena Lopez, M. Martin, EmoEvent: A multilingual emotion corpus based on different events, in: Proceedings of the Twelfth Language Resources and Evaluation Conference, European Language Resources Association, Marseille, France, 2020, pp. 1492–1498. URL: https://aclanthology.org/2020.lrec-1.186.
[17] E. Martínez-Cámara, M. T. Martín-Valdivia, L. A. Urena-López, A. Montejo-Ráez, Sentiment analysis in twitter, Natural language engineering 20 (2014) 1–28.
[18] J. Cañete, G. Chaperon, R. Fuentes, J.-H. Ho, H. Kang, J. Pérez, Spanish pre-trained bert model and evaluation data, in: PML4DC at ICLR 2020, 2020.
[19] M. García-Vega, M. Díaz-Galiano, M. García-Cumbreras, F. Plaza del Arco, A. Montejo-Ráez, S. M. Jiménez-Zafra, E. Martínez Cámara, C. A. Aguilar, M. Sobrevilla Cabezudo, L. Chiruzzo, D. Moctezuma, Overview of tass 2020: Introducing emotion detection, in: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2020) Co-Located with 36th Conference of the Spanish Society for Natural Language Processing (SEPLN 2020), Málaga, Spain, 2020, pp. 163–170.

[20] Narrativa, Roberta model fine-tuned for spanish summarisation, 2023. https://huggingface.co/Narrativa/bsc_roberta2roberta_shared-spanish-finetuned-mlsum-summarization [Accessed: 2023/07/26].

[21] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, J. Davison, S. Shleifer, P. von Platen, C. Ma, Y. Jernite, J. Plu, C. Xu, T. Le Scao, S. Gugger, M. Drame, Q. Lhoest, A. Rush, Transformers: State-of-the-art natural language processing, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, Association for Computational Linguistics, Online, 2020, pp. 38–45. doi:10.18653/v1/2020.emnlp-demos.6.

[22] A. M. Dai, Q. V. Le, Semi-supervised sequence learning, in: C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, R. Garnett (Eds.), Advances in Neural Information Processing Systems, volume 28, Curran Associates, Inc., 2015.

[23] R. Schwartz, J. Dodge, N. A. Smith, O. Etzioni, Green ai, Communications of the ACM 63 (2020) 54–63.