INFOTEC-MCDI at MentalRiskES: Gambling Risk Early Prediction in Social Media Users Through Bag of Word and BERT Ensembles

Alberto Herrera[1], Eric S. Tellez[1,2,*]

[1]*INFOTEC Centro Público de Investigación en Tecnologías de la Información y Comunicación, 112 Circuito Tecnopolo Sur, Parque Industrial Tecnopolo 2, Aguascalientes, 20326, México.*

[2]*Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI), 1582 Insurgentes Sur 1582, Crédito Constructor, Ciudad de México, 03940 México*

Abstract

Excess gambling can contribute to mental health disorders, including financial instability, interpersonal conflicts, increased risk of anxiety and depression, and impaired occupational or academic functioning. This report presents a computational model to early detect possible gambling disorders, particularly tackles the MentalRiskES challenge Task 2 running at IberLEF 2025 forum. The task asks to identify four gambling disorders in Twitch and Telegram text posts in Spanish. Our approach comprises a bag-of-words model and a BERT-based model achieving competitive performance with respect to quality and speed.

Keywords

gambling disorders, early mental risk identification, author profiling, text classification models

1. Introduction

Gambling issues could produce significant mental health disorders; they serve as a catalyst for addiction, initiating a domino effect of devastating consequences that ripple through individuals, families, and entire communities. These repercussions manifest themselves in the form of profound financial instability, the unraveling of personal relationships, and an increasing risk of mental health disorders, including anxiety and depression. The impacts extend further to the erosion of professional or academic performance and, in the most harrowing cases, lead to homelessness or even suicidal ideation. The insidious reach of gambling, particularly with the explosive growth of online platforms, has only magnified these hazards, underscoring the urgent need for a thorough and nuanced understanding of its multifaceted impact to inform robust prevention and intervention strategies.

This manuscript presents our solution to the MentalRiskES task 2 challenge at IberLEF 2025 [1]. This task consists of identifying a subject's possible gaming addiction based on text messages written by Spanish speakers on Telegram and Twitch, particularly, with an early

IberLEF 2025, September 2025, Zaragoza, Spain

*Corresponding author.

angelus2112@gmail.com (A. Herrera); eric.tellez@infotec.mx (E. S. Tellez)

https://sadit.github.io/ (E. S. Tellez)

0009-0001-0154-6131 (A. Herrera); 0000-0001-5804-9868 (E. S. Tellez)

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

published 2025-12-10

identification approach. The challenge also requires determining the degree level going binary, that is, low and high.

The categories of gambling activities identified by the challenge are:

Betting: Placing wagers on the outcomes of sporting events to obtain potential monetary gains.

Online gaming: Participation in traditional chance-based games like roulette, blackjack, and slots over the Internet.

Trading and crypto: Engaging in high-risk investments, especially in cryptocurrencies, with uncertain financial outcomes.

Lootboxes: A gambling-like mechanic in video games where players purchase virtual containers with randomized contents.

The MentalRiskES challenge is defined as an *early identification* text classification problem, that is, predictions are made with different numbers of messages, controlled by the organizers, and released incrementally in stages called rounds. The model then should refine the predictions as it access more complete profiles, but it is important to give a precise prediction using the least possible messages per profile, since the underlying problem should be attended as soon as possible to reduce the undesired effects. The complete rules and evaluation methodology are described in [2].

1.1. Text classification

In a supervised learning algorithm, like a classification model, it receives a dataset of examples (X, y), where $X = x_1, x_2, \ldots, x_n$ and $y = y_1, y_2, \ldots, y_n$. Each x_i is a vector belonging to $\mathbb{R}^\delta$, with δ as its dimension. In contrast, each y_i denotes a category representing a legitimate class. In this context, supervised learning-based profiling approaches require the conversion of input messages into real-valued vectors, where y is typically generated by human annotators who label each example.

A text classification pipeline consists of a model that transforms text inputs into high-dimensional vectors that can be used as input to a classification model to learn to predict the label. Some approaches like the bag-of-words approaches compute the vocabulary of a corpus and use it to map each text to very-high-dimensional vectors where each component corresponds with some word in the vocabulary. The precise approaches can be quite sophisticated [3, 4], yet the general pipelines for training and predictions are as follows:

training stage: Preprocess text messages, tokenize, compute vocabulary, compute a weight to each word in the vocabulary, compute a high-dimensional vector for each document, learn a classification model using the computed vectors and their associated labels.

prediction stage: Each message to be classified needs to be pre-processed, tokenized to create a high-dimensional vector using the vocabulary and associated weights, then the vector should be given to the classification model to obtain the prediction.

Conversely, deep learning methods require identical input and output formats, but these models are trained to carry out all essential processes within a single framework, facilitated by a suitable architecture. These models perform exceptionally well and benefit from the extensive knowledge corpus on which they were pretrained.

Author profiling

The author profiling problem is a forensic text classification problem that predicts specific characteristics of an author through the analysis of his/her documents. Typical characteristics of AP consist of the authors' writing style, such as *filler* words and phrases, as well as the message's direct content and intended meaning. These traits can be a possible determination of his/her mental health, gender, age, personality, native language, among others [5, 6]. Examples of this are the 2024 and 2023 MentalRiskEs competitions, where the tasks included determining psychological problems such as anxiety, depression, suicidal thoughts [7], eating disorders, or depression [8]. We can find that the profiling problem is not limited solely to identifying patterns from text; for example, in the sixth edition of the author profiling task at PAN 2018, this problem associated with gender identification was addressed from a multimodal approach, using a combination of text and images [9].

Our contribution

This article describes our approach to MentalRisk2025 Task 2, which involves early detection of gambling addiction among Spanish-speaking users on Telegram and Twitch. For the gambling-type subtask, we implement a bag-of-words method, MicroTC, while the RoBERTuito fine-tuned model is employed for the intensity subtask; both models used to solve one aspect of task 2, together, in each of the runs. We describe both of our approaches and outline our methodology for formulating our solutions.

Roadmap

The rest of the paper is organized as follows. Section 2 presents a brief review of work on the author profile and identification of mental risk. Section 3 describes our approach to solve the problem and the system we have implemented. Section 4 details our experimental methodology and lists our results. Finally, conclusions are given in Section 5.

2. Related work

[4] introduce the μTC framework,[1] a text classifier based on bag-of-words representations that is optimized in a large graph of possible configurations through combinatorial optimization. This method has shown competitive results in various tasks, especially in author profiling [10].

More recently, [11] introduced the Transformer architecture, a deep learning framework developed for Natural Language Processing (NLP), which has shown exceptional efficacy in a wide range of tasks. Transformers use attention mechanisms to highlight important information

[1] https://github.com/INGEOTEC/microtc

and understand long-range word dependencies within a sentence. Although this approach can considerably enhance performance over other methods, it tends to be computationally demanding. Therefore, people are typically limited to finetune a list of pretrained models to solve tasks. The Bidirectional Encoder Representations from Transformers (BERT) is an encoder-based transformer designed for pretraining models for various NLP tasks. BERT operates in two phases: pretraining for language understanding and finetuning for specific tasks ([12]). RoBERTuito is a language model based on the RoBERTa variant of BERT [13], trained on over 500 million Spanish tweets [14]. Its extensive training corpus and pretraining decisions make it a strong model for text classification.

In previous editions of MentalRiskEs [8, 7] we found that participants have mainly used pre-trained transformers to solve the tasks, among which we can find BETO, RoBERTa or RoBERTuito which was used in this work, however, we can notice that, particularly in the 2023 edition, it is expressed that not all participants used state-of-the-art models, belonging to deep learning, but rather opted for more traditional models such as Random Forest, Naïve Bayes or Support Vector Machines. Another situation that stands out from that edition is that there were contestants who used intensive feature selection methods that managed to become the closest to end-to-end solutions, within a set of classical algorithms. However, one of the conclusions we can draw from this edition is that even if the problems fall within the field of natural language processing, and even more so if they fall within similar areas, such as psychology, the way of expressing oneself, including the words used, can cause a more traditional model to perform better than a model compared to other models.

Author profiling. We can find that there have been different approaches to solve the profiling problem, for example, for the Author Profiling Task at PAN 2018, for age and gender identification, the model that obtained the best score consisted of a support vector machine and a combination of stylistic and second order features, while for gender identification the best model was one based on logistic regression with a mix of stylometric, lexical and n-gram features, on the other hand, the best model for age and gender profiling was one based on a support vector machine trained from stylometric, n-grams and second order features [15].

3. System description

We approached the challenge as a user profiling problem, messages were gathered around a particular user, defining its behavior; the challenge asks for early detection capabilities that were achieved by feeding each profile incrementally as messages were retrieved. Task 2 asks for prediction of the type of gambling (four classes listed in §1) and the level of risk of the individual, i.e., *low* and *high*; we created a solution employing two NLP techniques, a bag of words to predict types of gambling and a RoBERTuito-based model to identify risk levels.

The official training set comes from the *PRECOM-SM Corpus: Gambling in Spanish Social Media* prepared by [16]. This corpus comprises 350 individuals each associated with a gambling disorder. Each subject has multiple messages, and we compile each individual's profile by merging all their messages. Due to the limited size of the official evaluation dataset, we address this by dividing the official training set into an internal training and a test partition, allocating

Table 1
Distribution of profiles per gambling type in our internal train and test partition.

	Total	Labels			
		Betting	Online gaming	Trading and crypto	Lootboxes
train	262	68	79	96	19
test	88	17	25	39	7
sum	350	85	105	135	26

75% for training and the remaining 25% for testing. Table 1 provides the final distribution details of our internal partition.

As mentioned above, we select to probe our MicroTC framework and RoBERTuito to solve the task. Even when transformer-based models are unbeatable in many NLP tasks, the educated guess is that the core of our framework has proved to be very competitive in author profiling tasks [10], yet early identification was never tested. We selected RoBERTuito as the neural model because it has demonstrated flexibility and power [14]. We fine-tuned the model using the training set to predict the type of addiction and took an additional heuristic to predict the risk level.

The MicroTC framework produces a high-dimensional vector space from a broad vocabulary by utilizing different preprocessing functions and tokenizers. Hyperparameters for MicroTC include num_option='group' to consolidate all numbers into a single token, usr_option='group' for grouping user mentions, url_option='group' to merge all URLs into one token, and emo_option='group' to replace emoticons with a single token. We also remove duplicate characters, punctuation, and diacritics. After forming the vector space, LinearSVC [17] was trained with default hyperparameters. Both the MicroTC and RoBERTuito models underwent training using K-Fold Cross-Validation with $K = 10$. Figure 1 shows the prediction process, further detailed in the subsequent paragraphs.

Predicting the risk of having a gambling disorder. We use the RoBERTuito model for this problem. First, incoming messages are tokenized by the pysentimento library into character strings. We use prediction *logit* values for each category (*Betting*, *Online gaming*, *Trading and crypto*, *Lootboxes*). If there are at least two positive predictions, a high risk of betting disorder is indicated as 1, otherwise, a low risk is indicated as 0.

Predicting gambling disorder type by subject. We use MicroTC model predictions directly.

Tables 2 and 3 present the results obtained during the experimentation phase and across our data partitions. Table 2 shows results similar to those reported by [7] in task 1; This task consisted of classifying the illness that the subjects could have, from three options, namely depression, anxiety, or none of them. As we can see, thanks to a comparison between the MicroTc and RoBERTuito models, the former turns out to have better performance in most of the metrics used in experimentation. On the other hand, Table 3 shows better results compared to [7] in task 3, which addresses the problem of identifying the possibility that the subject has

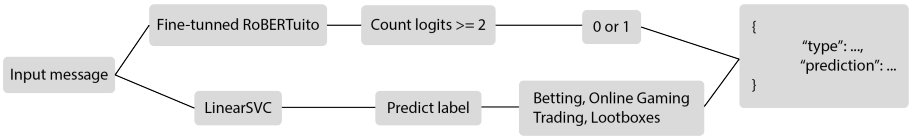

Figure 1: This diagram represents how the data is passed from its capture, through each of the models, until obtaining the corresponding predictions where associated with the key type we find the risk that the user who wrote the message has or does not have a betting disorder, taking as value 0 or 1 and, on the other hand, the key predictions indicates the type of disorder that said user could have, taking as value any of *Betting*, *Online gaming*, *Trading*, and *Lootboxes*.

Table 2
Prediction performance of the *addiction kind* (*Betting*, *Online gaming*, *Trading* or *Lootboxes*) on the test dataset using MicroTC in experimentation.

Model	Accuracy	Macro average			Micro average		
		P	R	F1	P	R	F1
MicroTC	0.613	0.671	0.668	0.652	0.654	0.654	0.654
RoBERTuito	0.571	0.614	0.601	0.589	0.569	0.569	0.569

Table 3
Prediction performance of the *risk level* (*low* or *high*) on the test dataset using RoBERTuito in experimentation.

Accuracy	Macro average			Micro average		
	P	R	F1	P	R	F1
0.862	0.843	0.813	0.824	0.848	0.848	0.848

suicidal thoughts or ideas. Since the problems mentioned in a previous edition of MentalRiskES are within the same field and are approached from a similar perspective, this being a multiclass classification problem and a binary classification problem, we decided to take the results obtained in these problems as a reference point in the selection of the models that make up our proposal.

4. Results

Our experiments were conducted in Google Colab[2] utilizing an L4 GPU system an Intel (R) Xeon (R) CPU @ 2.20GHz with two cores. We employed the Huggingface framework with a Pytorch back-end.

Table 4 illustrates the performance of our MicroTC model. Our analysis indicates that the model attained an accuracy rate of 0.594, alongside other performance metrics that enable us to assess class balance and the model's generalization capabilities in a multiclass scenario. It is evident that the model demonstrates relatively balanced outcomes in different classes. In contrast, we also report metrics related to the delay in identifying positive instances [18], our

[2]Google Colab site https://colab.research.google.com/.

Table 4
Prediction performance of the *addiction kind* on the test dataset using MicroTC.

Run	Accuracy	Macro avg.			Early detection		Latency		
		P	R	$F1$	ERDE5	ERDE30	latencyTP	speed	latency
1,2,3	0.594	0.605	0.599	0.589	0.381	0.343	2	0.990	0.540

findings revealing low values: 0.381 for ERD5 and 0.343 for ERD30. The results show that predicting the type of gaming disorder solely from text messages poses difficulties. An accuracy of 0.594, surpassing the chance level of 0.25 for four equally likely classes, suggests the model has captured certain discriminatory patterns in the texts. The small gap between macro and micro averaged metrics indicates a data imbalance naturally present in the database.

Performance limitations could be attributed to the significant linguistic ambiguity on informal platforms like Twitch and Telegram, where language often appears colloquial, abbreviated, or full of slang. Additionally, shared vocabulary across platforms and various potential gambling disorders may also play a role.

Table 5
Prediction performance of the *risk level*, i.e., modeled as binary classification with Low and High classes.

Run	Accuracy_c	Macro			Micro		
		P_c	R_c	$F1_c$	P_c	R_c	$F1_c$
1,2,3	0.838	0.900	0.756	0.721	0.838	0.838	0.838

Table 5 displays the results of our RoBERTuito model for assessing the risk of developing a gambling disorder from the studied messages, employing binary classification: 0 (low risk), 1 (high risk). The metrics indicate robust generalization to new data, with strong performance in classifying risk levels.

The results obtained in this case reflect the significantly robust performance of the model in predicting the risk of gambling disorder. The overall accuracy of .838 indicates the model's high ability to correctly assign the risk level to most examples. The macro accuracy of 0.900 is remarkably high, implying that, on average, the model's positive predictions (particularly those classified as *high risk*) are largely correct. This could be important from an intervention perspective. A high-accuracy model minimizes false positives, which prevents the user from being incorrectly alerted to low-risk users.

However, the closeness between the microaveraged metrics and the overall accuracy suggests that the classes are not excessively unbalanced or that the model has been able to learn useful patterns even in the presence of some imbalance. This is reinforced by the high microaveraged F1, i.e., 0.838, which balances precision and recall at the overall level, consolidating the model's overall effectiveness. For the purpose of this competition, our implementation placed 20th for Task 2.

5. Conclusions

This paper describes our solution for the early detection of gambling problems and the prediction of mental risk levels, using Spanish-language text messages collected from Telegram and Twitch, as part of the MentalRiskEs@IberLEF2025 challenge. Our approach involves a MicroTC model, based on a bag-of-words methodology, with LinearSVC serving as the classifier for the early identification of gambling disorder types. In addition, a fine-tuned RoBERTuito model is used to predict mental risk levels.

According to the official results, our solution demonstrated strong performance in detecting both gambling disorders and mental risk levels, particularly excelling in the former and achieving the top rankings in various metrics. When it comes to assessing the risk level of developing a gambling problem, the solution also boasted low latency and high speed, making it a competitive option to address this problem. While putting these models in practical situations requires additional investigation, our method offers alternative routes to mainstream strategies that predominantly use transformer-based models, demonstrating its competitive quality and relatively low computational demands.

Declaration on Generative AI

During the preparation of this work, the author(s) used Google translate in order to: translate from spanish to english. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

References

[1] J. Á. González-Barba, L. Chiruzzo, S. M. Jiménez-Zafra, Overview of IberLEF 2025: Natural Language Processing Challenges for Spanish and other Iberian Languages, in: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2025), co-located with the 41st Conference of the Spanish Society for Natural Language Processing (SEPLN 2025), CEUR-WS. org, 2025.

[2] A. M. Mármol-Romero, P. Álvarez-Ojeda, A. Moreno-Muñoz, F. M. P. del Arco, M. D. Molina-González, M.-T. Martín-Valdivia, L. A. Ureña-López, A. Montejo-Ráez, Overview of mentalriskes at iberlef 2025: Early detection of mental disorders risk in spanish, Procesamiento del Lenguaje Natural 75 (2025).

[3] C. Manning, H. Schutze, Foundations of statistical natural language processing, MIT press, 1999.

[4] E. S. Tellez, D. Moctezuma, S. Miranda-Jiménez, M. Graff, An automated text categorization framework based on hyperparameter optimization, Knowledge-Based Systems 149 (2018) 110–123.

[5] F. Rangel, P. Rosso, M. Potthast, B. Stein, Overview of the 5th author profiling task at pan 2017: Gender and language variety identification in twitter, Working notes papers of the CLEF (2017) 1613–0073.

[6] M. Potthast, J. Kiesel, K. Reinartz, J. Bevendorff, B. Stein, A stylometric inquiry into hyperpartisan and fake news, in: I. Gurevych, Y. Miyao (Eds.), Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Melbourne, Australia, 2018, pp. 231–240. URL: https://aclanthology.org/P18-1022/. doi:10.18653/v1/P18-1022.

[7] 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 2024: Early detection of mental disorders risk in spanish, Procesamiento del Lenguaje Natural (2024). URL: https://www.who.int/news/item/17-06-2021-. doi:10.26342/2024-73-33.

[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 (2023) 329–350. doi:10.26342/2023-71-26.

[9] F. Rangel, P. Rosso, M. Montes-Y-Gómez, M. Potthast, B. Stein, Overview of the 6th Author Profiling Task at PAN 2018: Multimodal Gender Identification in Twitter, Technical Report, 2018. URL: http://pan.webis.de.

[10] M. Graff, D. Moctezuma, E. S. Téllez, Bag-of-word approach is not dead: A performance analysis on a myriad of text classification challenges, Natural Language Processing Journal 11 (2025) 100154. URL: https://www.sciencedirect.com/science/article/pii/S2949719125000305. doi:https://doi.org/10.1016/j.nlp.2025.100154.

[11] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need (2017). URL: http://arxiv.org/abs/1706.03762.

[12] F. A. Acheampong, H. Nunoo-Mensah, W. Chen, Transformer models for text-based emotion detection: a review of bert-based approaches, Artificial Intelligence Review 54 (2021) 5789–5829. doi:10.1007/s10462-021-09958-2.

[13] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, Roberta: A robustly optimized bert pretraining approach, 2019. URL: https://arxiv.org/abs/1907.11692. arXiv:1907.11692.

[14] J. M. Pérez, D. A. Furman, L. Alonso Alemany, F. M. Luque, RoBERTuito: a pre-trained language model for social media text in Spanish, in: Proceedings of the Thirteenth Language Resources and Evaluation Conference, European Language Resources Association, Marseille, France, 2022, pp. 7235–7243. URL: https://aclanthology.org/2022.lrec-1.785.

[15] F. Rangel, P. Rosso, B. Verhoeven, W. Daelemans, M. Potthast, B. Stein, Overview of the 4th Author Profiling Task at PAN 2016: Cross-Genre Evaluations, Technical Report, 2016. URL: http://pan.webis.de.

[16] P. Álvarez-Ojeda, M. V. Cantero-Romero, A. Semikozova, A. Montejo-Ráez, The precomsm corpus: Gambling in spanish social media, in: Proceedings of the 31st International Conference on Computational Linguistics, 2025, pp. 17–28.

[17] Scikit-learn, Linearsvc, 2025. URL: https://scikit-learn.org/stable/modules/generated/sklearn.svm.LinearSVC.html.

[18] D. E. Losada, F. Crestani, A test collection for research on depression and language use, Lecture Notes in Computer Science (2016) 28–39. doi:10.1007/978-3-319-44564-9_3.