

NLP-UNED-2 at eRisk 2023: Detecting Pathological Gambling in Social Media through Dataset Relabeling and Neural Networks

Notebook for the eRisk Lab at CLEF 2023

Hermenegildo Fabregat^{1,3}, Andres Duque^{1,2,*}, Lourdes Araujo^{1,2} and Juan Martinez-Romo^{1,2}

¹NLP & IR Group, Dpto. Lenguajes y Sistemas Informáticos, Universidad Nacional de Educación a Distancia (UNED), Juan del Rosal 16, Madrid 28040, Spain

²IMIENS: Instituto Mixto de Investigación, Escuela Nacional de Sanidad, Monforte de Lemos 5, Madrid 28019, Spain

³Avature Machine Learning, Marqués de Valdeiglesias, 3, Madrid 28004, Spain

Abstract

This paper describes our participation in Task 2 (Early Detection of Signs of Pathological Gambling) from the CLEF 2023 eRisk Workshop, addressed to detecting early signs of pathological gambling in messages written by Social Media users. Since the original dataset is annotated at user level, we perform a relabeling process based on Approximate Nearest Neighbors (ANN) on vectorial representations of the messages, in order to produce a dataset annotated at message level. Then, different neural network architectures are tested using the re-labeled training dataset in order to develop models for classifying test instances. Our system obtains the second best performance in the decision-based evaluation, and is one of the best performing techniques in the ranking-based evaluation. Hence, the combination of the re-labeling technique with neural architectures leads to an accurate detection of signs of pathological gambling.

Keywords

Pathological gambling detection, Approximate Nearest Neighbors, Relabeling, Neural Networks.

1. Introduction

Research on potential health risks through social media analysis has emerged as a captivating research domain in the last few years. Within this field of study, the scientific community has undertaken various initiatives, including the eRisk workshop, which has been a recurring event in the Conference Labs of the Evaluation Forum (CLEF) since 2017. This workshop serves as a collaborative platform for the development of methodologies and practical approaches aimed at the early detection of diverse health risks, such as eating disorders, self-harm, and depression.

CLEF 2023: Conference and Labs of the Evaluation Forum, September 18–21, 2023, Thessaloniki, Greece

*Corresponding author.

✉ gildo.fabregat@lsi.uned.es (H. Fabregat); aduque@lsi.uned.es (A. Duque); lurdes@lsi.uned.es (L. Araujo); juaner@lsi.uned.es (J. Martinez-Romo)

🆔 0000-0001-9820-2150 (H. Fabregat); 0000-0002-0619-8615 (A. Duque); 0000-0002-7657-4794 (L. Araujo); 0000-0002-6905-7051 (J. Martinez-Romo)



© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

By analysing the textual content of social media posts and messages, valuable insights can be gained to identify individuals at risk.

In this paper we present a system for tackling Task 2 of the eRisk 2023 Workshop: Early Detection of Signs of Pathological Gambling [1], which is a continuation from Task 1 of the eRisk 2021 Workshop [2] and Task 1 of the eRisk 2022 Workshop [3]. Our system is also an improved version of the one that we presented as the “UNED-NLP” team in the 2022 edition of the competition [4].

The proposed approach first transforms each message in the dataset into a vector-based representation through sentence embeddings. Then, we use an Approximate Nearest Neighbor (ANN) technique for relabeling the original dataset, annotated at user level, and generating a new dataset annotated at message level. Finally, we propose different techniques for employing this relabeled dataset in the final classification of the test instances: as a baseline, we propose an ANN-based technique similar to the one employed in the relabeling of the dataset. We also propose two techniques for using the relabeled training dataset as input of a Recurrent Neural Network (RNN). Finally, the neural models trained in the previous step are also employed for generating alternative versions of the relabeled dataset and testing whether their use allows us to improve the obtained results.

The rest of the paper is structured as follows: Section 2 offers a summary of related works and systems participating in previous competitions. A brief description of the proposed task, the available dataset and the metrics employed in the evaluation is presented in Section 3. Our system proposal is described in Section 4 and the obtained results in Section 5. Finally, some conclusions and future lines of work are depicted in Section 6.

2. Related Work

Gambling disorder (GD) refers to a persistent and recurrent gambling behavior that causes significant distress [5]. In the United States, GD is estimated to affect approximately 0.5% of the adult population, with comparable or potentially higher rates observed in other nations. However, individuals with GD often go untreated and frequently remain unrecognized. Moreover, GD frequently co-occurs with other psychiatric disorders, with notable prevalence rates reported for mood disorders, anxiety disorders, attention deficit disorders, and substance use disorders among individuals with GD [6]. Furthermore, GD is often accompanied by a higher incidence of unemployment, economic hardships, divorce, and compromised health. Notably, GD holds close associations with other addictive disorders and stands as the first non-substance addictive behavior to be officially acknowledged [7].

The advent of social networks has presented a valuable source of information for studying and identifying individuals with gambling problems at an early stage. In alignment with this perspective, the eRisk competition incorporated the issue of pathological gambling into its agenda for the first time in 2021 [2]. No training data was provided to the participants in that first edition, hence many of the systems used external resources for training their systems, such as Reddit posts crawled by themselves [8, 9, 10, 11]. Transformer-based architectures [12] were selected for classification by some of the systems [8, 10]. Other architectures such as LSTM networks were also used by other participants [11], while an Embedding Topic Model (ETM)

was used by [9] for modeling users and similarity measures with other external resources such as gambling testimonials and questionnaires were then employed for determining those users more likely to be positive. The best performing system was presented by the UNSL team [13], which tested different representation techniques (Bag of words, Doc2Vec) and classification methods (LSTMs, SVMs) on a dataset generated from Reddit posts.

In the 2022 edition the participants had access to labeled training data. The UNSL team improved its method with new policies within their classification techniques [14], also obtaining interesting results on the task. Other teams such as SINAI [15] employed Transformer-based methods for obtaining the representation of the messages in the dataset, and applied regression techniques on different features (volumetry or lexical diversity, among others). Deep learning models were also employed by teams like NLPGroup-IISERB [16], BLUE [17] or BioNLP-UniBuc [18]. Other teams selected different classification models such as SVMs or XGBoost, after extracting Glove features from user posts [19]. Our proposal based on dataset relabeling from user-level annotations to message-level annotations through Approximate Nearest Neighbors [4] obtained the best results in the competition, and hence the research presented in this paper is an extension of that particular work.

3. Task 2: Early Detection of Signs of Pathological Gambling

Task 2 of eRisk 2023 is denoted “Early detection of signs of pathological gambling” [1]. This is the third edition of the task, which was first introduced in the CLEF 2021 eRisk Workshop [2] and had a second edition in the CLEF 2022 eRisk Workshop [3]. In this task, participating systems are asked to determine whether an individual can be classified as a pathological gambler (positive users) or a non-pathological gambler (negative users) based on the user’s Social Media messages. Systems must sequentially analyze chronological posts for each user for detecting early traces of pathological gambling.

3.1. Dataset

The dataset used in the task is composed of a set of XML documents, each of them containing chronologically ordered Social Media posts belonging to a particular user. Each document is annotated as “1” (positive) if the user is labeled as a pathological gambler, and “0” (negative) otherwise.

Table 1 shows the main statistics of the dataset. Column “Pathological gamblers” indicates those users marked as positive in the dataset, while “Control” users represent negative users.

3.2. Metrics

System evaluation is twofold:

- Decision-based evaluation: This first type of evaluation aims to analyze the performance of the participating systems in terms of standard measures such as Precision, Recall and F-Measure. However, other metrics are also introduced in this evaluation that take into account the delay incurred by a system before it detects a true positive. Two of

Table 1

Statistics of the dataset employed in Task 2 of the CLEF eRisk Workshop: Early Detection of Signs of Pathological Gambling.

	Pathological Gamblers	Control
Num. subjects	103	2071
Num. submissions (post & comments)	33,719	1,069,152
Avg num. of submissions per subject	327.33	516.25
Svg num. of days from first to last submission	≈ 675	≈ 878
Avg num. of words per submission	28.9	20.47

these metrics, denoted $ERDE$ and $ERDE_o$ consider the number or the percentage of messages that have to be processed before emitting an alert of positive user. In order to overcome the low interpretability of these latter metrics, a latency-weighted F-Score is also introduced by multiplying the standard F-Measure by a penalty factor based on the median delay of true positive detection.

- **Ranking-based evaluation:** The second type of evaluation is a complementary approach that requires the systems to provide a score indicating the risk of pathological gambling of a user every time a new message is analyzed. Users are then ranked using this score and standard ranking metrics such as $P@k$ or $NDCG@k$ can be applied, with the parameter k being the number of analyzed messages before evaluating the ranking.

More information about the complete set of metrics employed in the evaluation can be found in previous overviews of eRisk competitions [2, 3, 20].

4. Proposed Model

In this section we define the configuration of the different techniques that have been tested in this research, all of them based on the idea of relabeling the original dataset for generating a new training dataset annotated at message level.

4.1. Baseline

The baseline system for our participation in the task is the same system that obtained the best results in the 2022 edition of the task [4]. In this system, we use Universal Sentence Encoder [21] to encode each user’s messages, which are transformed into 512-dimensional vectors. Then, an Approximate Nearest Neighbor technique is employed for modeling the training dataset. In particular, and considering the results obtained in the 2022 task, we employed the Hierarchical Navigable Small World (HNSW) graphs method, implemented in the Non-Metric Space Library [22]. The main idea behind this method is to build a proximity graph in which each datum corresponds to a node and the edges connecting some nodes define the neighborhood relationship. In this representation, a neighbor’s neighbor is likely to also be a neighbor. Search can be efficiently performed by iteratively extending neighbors of neighbors in a best-first based search strategy.

Once the ANN search index is built, we perform the relabeling process on the training dataset: in the original corpus, each user is labeled as positive if at least a positive message can be found within his/her posts, and negative otherwise. Hence, we first consider all messages of a positive user to be positive, and all messages of a negative user to be negative. From the training set, we consider only those messages that contain title information. This indicates that the message represents the opening of a Reddit thread. Through this initial filtering, we intend to give preference to those discussions originally initiated by the subject user (e.g. calls for help or topic-related questions). We iteratively process each message from each positive user of the training set, and re-annotate its class according to the similarity with the nearest neighbors of the considered message, through the following process: given a message M from a user U we consider U_M to be positive if the k nearest neighbors retrieved include at least j positive messages. We explored different values of these parameters k and j in order to guarantee the convergence of the algorithm on a non-zero set of positive training instances after applying the relabeling process. In this step, the best values for these parameters were $k = 10$, $j = 8$. After the whole training dataset has been relabeled, this method is repeated until convergence is reached, this is, until there are no changes in the training set labels.

For the final classification step, a different set of values for parameters k and j was determined after a tuning evaluation stage. In this step, $k = 18$ and $j = 18$, this is, the 18 nearest neighbors of a test message are retrieved from the training dataset, and all of them must be positive in order to classify this test message as positive. Regarding the ranking-based evaluation, we also use a scoring function for calculating the risk of pathological gambling of a user given a message, which is the mean distance from U_M to the nearest recovered neighbors: $(1 - \frac{1}{j} \sum_{x=1}^j \text{cosine}(U_M, M_x))$, where M_x is each of the retrieved nearest messages. As the scoring function is only used in the final classification step for the ranking-based evaluation, the values of the mentioned parameters are $k = 18$, $j = 18$. In this case, the scoring function is calculated only for those messages classified as positive, this is, those messages whose 18 nearest neighbors are all positive. Otherwise, the scoring function will return zero. The risk of pathological gambling of a particular user, each time a new message is analysed, is the maximum scoring obtained by any of the (positive) messages of that user, from those messages processed up to that point. We denote this configuration as “**Baseline**” in our experiments.

4.2. Neural Models for Classification

One of the main objectives in our research for this task is to determine whether the use of neural models in our pipeline has a significant impact on the overall results. For this purpose, we have developed two architectures for performing the final classification of the test messages. The first architecture, denoted “**RNN_Base**” is a simple network with a RNN layer composed of 64 neurons with ReLU activation function, and a final layer with a single neuron with sigmoid activation function. We use the relabeled training dataset, annotated at message level, as input for training our model. In inference time, we define a decision threshold for the sigmoid function of 0.9, this is, only messages classified as positive with a 90% confidence level are considered to be positive. The scoring function in this case is the direct output of the sigmoid-activated neuron in the last layer of the network, which always lies in the range from 0 to 1.

The second neural architecture, named “**RNN_Sim**” in our experiments, only presents

modifications in the input part of the neural network, while the number and type of layers and neurons remain the same. In this architecture, we intend to keep studying those approaches based on similarity, but avoiding the huge search space derived from the use of ANN techniques. For this purpose, we define a similarity matrix by taking a number N of random positive messages from the re-labeled dataset and calculating the cosine distance of each input message in the network with all those positive messages in the matrix. Then, we build a distance vector with all the cosine similarities. Hence, the dimension of this vector will also be N . This distance vector is the final input to the network. In order to explore all the positive messages within the training dataset, the similarity matrix is updated in each epoch of the training phase, with a new subset of N random positive messages. Finally, the last subset used for the training phase is then employed as the similarity matrix for performing the final classification of test instances. After some tests on the development dataset, we have defined $N = 50$ as the size of the similarity matrix.

Finally, some adjustments have been included in both neural architectures for optimization: the number of epochs is 20, with early stopping policies (after 4 epochs with no validation loss improvement) and a dropout layer after the RNN layer (dropout rate = 0.25).

4.3. Neural Models for Relabeling

Once that the neural models have been developed, we are interested in testing whether they can also be used for improving the relabeling process described in Section 4.1. We develop two additional configurations of our system for this purpose: the first configuration, named **“RNN_Relabel_Base”**, uses the trained **“RNN_Sim”** model described in Section 4.2 for relabeling the original training dataset. Once the training dataset has been relabeled, the final classification step is performed as described in Section 4.1, this is, through the use of the NMSLIB library for extracting the nearest neighbors for each test message and applying the aforementioned tagging and scoring functions.

Finally, the last configuration of the proposed system, denoted **“RNN_Relabel_Refined”** is a combination of all techniques explored in this work. In this configuration, we consider a training message to be positive if both the original **“Baseline”** relabeling method and the **“RNN_Relabel_Base”** relabeling method annotated it as positive. Hence, only those messages considered as positive by the ANN-based and the neural-based relabeling techniques will be maintained as positive in this case. The final classification step is also performed as in the **“RNN_Relabel_Base”** configuration.

5. Results and Discussion

In this section the main results obtained by the proposed configurations of our system are shown and compared to other participants in the task. Table 2 shows the summarized results for the decision-based evaluation. The five configurations of our systems are depicted first, while only the best run from each of the participating systems is included, ordered by the latency-weighted F1 measure.

Our best performing configuration, **“RNN_Sim”**, is able to achieve the second best result in terms of latency-weighted F1, five points behind from the best performing system, **EliRF-UPV**.

Table 2

Test results: Results of the decision-based evaluation for task T2. Bold indicates the best result for each considered metric.

	Prec	Rec	F1	ERDE5	ERDE50	latencyTP	speed	latency-weighted F1
Baseline	0.945	0.844	0.892	0.038	0.019	18.0	0.933	0.833
RNN_Base	0.764	0.883	0.819	0.033	0.010	13.0	0.953	0.781
RNN_Sim	0.896	0.922	0.909	0.030	0.014	10.0	0.964	0.877 (2)
RNN_Relabel_Base	0.957	0.883	0.919	0.034	0.016	13.0	0.953	0.876
RNN_Relabel_Refined	0.947	0.883	0.914	0.034	0.016	12.0	0.957	0.875
ELiRF-UPV R0	1.000	0.883	0.938	0.026	0.010	4.0	0.988	0.927 (1)
XABI_EHU R0	0.846	0.961	0.900	0.030	0.012	8.0	0.973	0.875 (3)
UNSL R1	0.790	0.806	0.798	0.048	0.014	13.0	0.953	0.761 (4)
BioNLP-IISERB R0	0.933	0.680	0.787	0.038	0.037	62.0	0.766	0.603 (5)
UMUTeam R1	1.000	0.338	0.559	0.047	0.043	94.5	0.651	0.364 (6)
SINAI R3	0.126	1.000	0.224	0.029	0.020	2.0	0.996	0.223 (7)
NUS-eRisk R3	0.067	0.874	0.125	0.065	0.042	17.0	0.938	0.117 (8)
RELAI R1	0.058	0.971	0.109	0.048	0.039	1.0	1.000	0.109 (9)
NLP-UNED R0	0.057	0.903	0.108	0.052	0.052	1.0	1.000	0.108 (10)
OBSER-MENH R0	0.048	1.000	0.092	0.064	0.049	3.0	0.992	0.092 (11)

Our system offers good values for all the proposed metrics, although the latency and speed values are slightly high (this is, the system needs to process a relatively high number of messages before determining that a user is positive), which is detrimental when it comes to the final latency-weighted F1 value.

Regarding our different configurations, the techniques that perform relabeling through the use of the neural architectures obtain latency-weighted values very similar to the “RNN_Sim” configuration, although their performance is better in terms of precision and worse in terms of recall. This, along with the results of the “Baseline” configuration, indicates that ANN-based classification has a higher impact on precision, while neural-based classification improves the recall of the system.

Table 3 offers the main results of the ranking-based evaluation, in terms of Precision@10, NDCG@10 and NDCG@100 when the system has already processed 1, 100, 500 and 1,000 writings. As before, we show in the table the results of our five configurations, together with the best result of the systems participating in the competition.

As shown in the Table, “RNN_Sim” is our best performing configuration, obtaining perfect scores for P@10 and NDCG@10 after 1, 100, 500 and 1000 writings, and high values of NDCG@100 in all the cases. The main differences between the different configurations can be seen after processing 1 message: those techniques that perform RNN-based classification (“RNN_Base” and “RNN_Sim”) achieve better values of the NDCG@100 metric. In these cases, the final score of the user comes directly from the activation value of the neuron in the last layer of the network, and hence a score (risk of being a pathological gambler) is given to the user even if the network decides that the processed message is negative, while the same case would offer a score of 0 in the ANN-based classification. Therefore, assigning some score even to negative users allows the system to generate a better ranking based on the risk of being a pathological gambler.

Regarding the comparison between our best configuration and the best runs of the participat-

Table 3

Test results: Results of the ranking-based evaluation for task T2. Bold indicates the best result for each considered metric.

	1 writing			100 writings			500 writings			1000 writings		
	P@10	NDCG@10	NDCG@100	P@10	NDCG@10	NDCG@100	P@10	NDCG@10	NDCG@100	P@10	NDCG@10	NDCG@100
Baseline	1.00	1.00	0.32	1.00	1.00	0.83	1.00	1.00	0.90	1.00	1.00	0.90
RNN_Base	1.00	1.00	0.57	1.00	1.00	0.89	1.00	1.00	0.89	1.00	1.00	0.87
RNN_Sim	1.00	1.00	0.59	1.00	1.00	0.92	1.00	1.00	0.95	1.00	1.00	0.93
RNN_Relabel_Base	1.00	1.00	0.40	1.00	1.00	0.86	1.00	1.00	0.93	1.00	1.00	0.94
RNN_Relabel_Refined	1.00	1.00	0.40	1.00	1.00	0.85	1.00	1.00	0.92	1.00	1.00	0.93
XABI_EHU R0	1.00	1.00	0.57	1.00	1.00	0.50	0.80	0.88	0.41	0.90	0.94	0.41
NUS-eRisk R3	0.00	0.00	0.06	0.00	0.00	0.04	0.10	0.19	0.08	0.00	0.00	0.02
OBSER-MENH R0	1.00	1.00	0.64	1.00	1.00	0.55	1.00	1.00	0.48	1.00	1.00	0.50
ELiRF-UPV R0	1.00	1.00	0.59	1.00	1.00	0.91	1.00	1.00	0.95	1.00	1.00	0.94
BioNLP-IISERB R0	0.40	0.60	0.14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RELAI R0	0.30	0.25	0.08	0.00	0.00	0.02	0.00	0.00	0.06	0.00	0.00	0.00
NLP-UNED R0	0.00	0.00	0.03	0.00	0.00	0.02	0.00	0.00	0.02	0.00	0.00	0.02
SINAI R1	1.00	1.00	0.73	1.00	1.00	0.90	1.00	1.00	0.85	0.00	0.00	0.00
UNSL R1	1.00	1.00	0.57	1.00	1.00	0.78	1.00	1.00	0.67	1.00	1.00	0.70
UMUTeam R0	1.00	1.00	0.63	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00

ing systems, our system achieves the best results for 10 out of the 12 proposed metrics, which also indicates the robustness of this particular scoring function.

Finally, Table 4 shows the execution times required by the systems for processing the test set. Systems that processed the test set only partially have not been included in the table.

Table 4

Test results: Comparison of the execution times required by those systems that processed the whole test set.

Team	#runs	#user writings processed	lapse of time (from 1st to last response)
NLP-UNED-2	5	2004	7 days 04:24:33.158
XABI_EHU	5	2004	4 days 23:52:41.454
NUS-eRisk	5	2004	9 days 14:39:26.347
OBSER-MENH	5	2004	6 days 03:56:44.247
ELiRF-UPV	1	2004	1 day 13:03:54.419
UNSL	3	2004	1 day 02:17:36.417
UMUTeam	5	2004	14 days 00:29:30.434

Although we are aware that the “Baseline” configuration is quite fast in processing the test dataset, obtaining the best execution times in the 2022 edition of the competition [4], we can observe in the table that our system is not among the fastest participating teams. This is probably due to the use of neural-based inference in the “RNN_Base” and “RNN_Sim” configurations, which are more likely to involve a much longer execution time than searches on the indexes generated by ANN algorithms. It would be useful to separate the execution times of each run submitted by the systems in order to better visualize the trade-off between time consumption

and overall results of each system.

6. Conclusions and Future Work

This paper describes our participation in Task 2 of eRisk 2023: Early detection of signs of pathological gambling. We have made further developments on the system introduced in the 2022 edition which obtained the best overall results. In particular, in this research we explore the use of neural architectures and their combination with Approximate Nearest Neighbor techniques used for performing dataset relabeling, from user level annotations to message level annotations.

Two different neural architectures have been proposed: a simple RNN architecture receiving vectorial representations of the input messages obtained through Universal Sentence Encoders, and a second similar architecture which uses as input the similarity of each input message with a reference similarity matrix containing positive messages from the training dataset. These architectures have been tested for both performing the final message classification, and also for testing whether they can be also useful for relabeling the training dataset (original or already relabeled). We have obtained the second best results in the decision-based evaluation and the best overall results in the ranking-based evaluation. Although the relabeling process already proposed in the past edition of the competition is probably the main reason for achieving such good results, the introduction of neural architectures within our pipeline allows us to obtain some improvements on the final results.

As future lines of work for this research, we consider the use of Transformer-based neural architectures [12] such as BERT [23], and a deeper exploration of both the parameters in the ANN algorithms and the hyperparameters of the employed neural networks. Pre-trained models built with in-domain information could also be useful for better represent the knowledge that we try to model for detecting the risk of pathological gambling. Finally, the application of the proposed system to similar tasks, such as detecting the risk of anorexia, depression or self-harm behaviours is also being currently considered.

Acknowledgments

This work has been partially supported by the Spanish Ministry of Science and Innovation within the DOTT-HEALTH Project (MCI/AEI/FEDER, UE) under Grant PID2019-106942RB-C32 and OBSER-MENH Project (MCIN/AEI/10.13039/501100011033 and NextGenerationEU/PRTR) under Grant TED2021-130398B-C21 as well as project RAICES (IMIENS 2022).

References

- [1] J. Parapar, P. Martín Rodilla, D. E. Losada, F. Crestani, Overview of erisk 2023: Early risk prediction on the internet., *Experimental IR Meets Multilinguality, Multimodality, and Interaction*. 14th International Conference of the CLEF Association, CLEF 2023, Thessaloniki, Greece (2023).

- [2] J. Parapar, P. Martín-Rodilla, D. E. Losada, F. Crestani, Overview of erisk at CLEF 2021: Early risk prediction on the internet (extended overview), Proceedings of the Working Notes of CLEF 2021 - Conference and Labs of the Evaluation Forum, Bucharest, Romania, 2021 2936 (2021) 864–887. URL: <http://ceur-ws.org/Vol-2936/paper-72.pdf>.
- [3] J. Parapar, P. Martín Rodilla, D. E. Losada, F. Crestani, Overview of erisk 2022: Early risk prediction on the internet., Experimental IR Meets Multilinguality, Multimodality, and Interaction. 13th International Conference of the CLEF Association, CLEF 2022, Bologna, Italy (2022).
- [4] H. Fabregat, A. Duque, L. Araujo, J. Martínez-Romo, UNED-NLP at erisk 2022: Analyzing gambling disorders in social media using approximate nearest neighbors, in: G. Faggioli, N. Ferro, A. Hanbury, M. Potthast (Eds.), Proceedings of the Working Notes of CLEF 2022 - Conference and Labs of the Evaluation Forum, Bologna, Italy, September 5th - to - 8th, 2022, volume 3180 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2022, pp. 894–904. URL: <https://ceur-ws.org/Vol-3180/paper-71.pdf>.
- [5] M. N. Potenza, I. M. Balodis, J. Derevensky, J. E. Grant, N. M. Petry, A. Verdejo-Garcia, S. W. Yip, Gambling disorder, *Nature reviews Disease primers* 5 (2019) 1–21.
- [6] M. N. Potenza, T. R. Kosten, B. J. Rounsaville, Pathological gambling, *Jama* 286 (2001) 141–144.
- [7] C. J. Rash, J. Weinstock, R. Van Patten, A review of gambling disorder and substance use disorders, *Substance abuse and rehabilitation* 7 (2016) 3.
- [8] A. Basile, M. Chinea-Rios, A.-S. Uban, T. Müller, L. Rössler, S. Yenikent, B. Chulví, P. Rosso, M. Franco-Salvador, Upv-symanto at erisk 2021: Mental health author profiling for early risk prediction on the internet, Working Notes of CLEF (2021).
- [9] D. Maupomé, M. D. Armstrong, F. Rancourt, T. Soulas, M.-J. Meurs, Early detection of signs of pathological gambling, self-harm and depression through topic extraction and neural networks, Proceedings of the Working Notes of CLEF (2021).
- [10] A.-M. Bucur, A. Cosma, L. P. Dinu, Early risk detection of pathological gambling, self-harm and depression using bert, Working Notes of CLEF (2021).
- [11] R. P. Lopes, Cedri at erisk 2021: A naive approach to early detection of psychological disorders in social media, in: CEUR Workshop Proceedings, CEUR Workshop Proceedings, 2021, pp. 981–991.
- [12] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, *CoRR abs/1706.03762* (2017). URL: <http://arxiv.org/abs/1706.03762>. arXiv:1706.03762.
- [13] J. M. Loyola, S. Burdisso, H. Thompson, L. Cagnina, M. Errecalde, Unsl at erisk 2021: A comparison of three early alert policies for early risk detection, in: Working Notes of CLEF 2021-Conference and Labs of the Evaluation Forum, Bucarest, Romania, 2021.
- [14] J. M. Loyola, H. Thompson, S. Burdisso, M. Errecalde, UNSL at erisk 2022: Decision policies with history for early classification, in: G. Faggioli, N. Ferro, A. Hanbury, M. Potthast (Eds.), Proceedings of the Working Notes of CLEF 2022 - Conference and Labs of the Evaluation Forum, Bologna, Italy, September 5th - to - 8th, 2022, volume 3180 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2022, pp. 947–960. URL: <https://ceur-ws.org/Vol-3180/paper-75.pdf>.
- [15] A. M. Mármol-Romero, S. M. J. Zafra, F. M. P. del Arco, M. D. Molina-González, M. T. M.

- Valdivia, A. Montejo-Ráez, SINAI at erisk@clef 2022: Approaching early detection of gambling and eating disorders with natural language processing, in: G. Faggioli, N. Ferro, A. Hanbury, M. Potthast (Eds.), Proceedings of the Working Notes of CLEF 2022 - Conference and Labs of the Evaluation Forum, Bologna, Italy, September 5th - to - 8th, 2022, volume 3180 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2022, pp. 961–971. URL: <https://ceur-ws.org/Vol-3180/paper-76.pdf>.
- [16] H. Srivastava, L. N. S, S. S, T. Basu, Nlp-iiserb@erisk2022: Exploring the potential of bag of words, document embeddings and transformer based framework for early prediction of eating disorder, depression and pathological gambling over social media, in: G. Faggioli, N. Ferro, A. Hanbury, M. Potthast (Eds.), Proceedings of the Working Notes of CLEF 2022 - Conference and Labs of the Evaluation Forum, Bologna, Italy, September 5th - to - 8th, 2022, volume 3180 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2022, pp. 972–986. URL: <https://ceur-ws.org/Vol-3180/paper-77.pdf>.
- [17] A. Bucur, A. Cosma, L. P. Dinu, P. Rosso, An end-to-end set transformer for user-level classification of depression and gambling disorder, in: G. Faggioli, N. Ferro, A. Hanbury, M. Potthast (Eds.), Proceedings of the Working Notes of CLEF 2022 - Conference and Labs of the Evaluation Forum, Bologna, Italy, September 5th - to - 8th, 2022, volume 3180 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2022, pp. 851–863. URL: <https://ceur-ws.org/Vol-3180/paper-67.pdf>.
- [18] T. Dumitrascu, CLEF erisk 2022: Detecting early signs of pathological gambling using ML and DL models with dataset chunking, in: G. Faggioli, N. Ferro, A. Hanbury, M. Potthast (Eds.), Proceedings of the Working Notes of CLEF 2022 - Conference and Labs of the Evaluation Forum, Bologna, Italy, September 5th - to - 8th, 2022, volume 3180 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2022, pp. 883–893. URL: <https://ceur-ws.org/Vol-3180/paper-70.pdf>.
- [19] S. Stalder, E. Zankov, ZHAW at erisk 2022: Predicting signs of pathological gambling - glove for snowy days, in: G. Faggioli, N. Ferro, A. Hanbury, M. Potthast (Eds.), Proceedings of the Working Notes of CLEF 2022 - Conference and Labs of the Evaluation Forum, Bologna, Italy, September 5th - to - 8th, 2022, volume 3180 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2022, pp. 987–994. URL: <https://ceur-ws.org/Vol-3180/paper-78.pdf>.
- [20] D. E. Losada, F. Crestani, J. Parapar, Overview of erisk at CLEF 2020: Early risk prediction on the internet (extended overview), Working Notes of CLEF 2020 - Conference and Labs of the Evaluation Forum, Thessaloniki, Greece, 2020 2696 (2020). URL: http://ceur-ws.org/Vol-2696/paper_253.pdf.
- [21] D. Cer, Y. Yang, S. Kong, N. Hua, N. Limtiaco, R. S. John, N. Constant, M. Guajardo-Cespedes, S. Yuan, C. Tar, Y. Sung, B. Strope, R. Kurzweil, Universal sentence encoder, CoRR abs/1803.11175 (2018). URL: <http://arxiv.org/abs/1803.11175>. arXiv:1803.11175.
- [22] Y. A. Malkov, D. A. Yashunin, Efficient and robust approximate nearest neighbor search using hierarchical navigable small world graphs, CoRR abs/1603.09320 (2016). URL: <http://arxiv.org/abs/1603.09320>. arXiv:1603.09320.
- [23] J. Devlin, M. Chang, K. Lee, K. Toutanova, BERT: pre-training of deep bidirectional transformers for language understanding, in: J. Burstein, C. Doran, T. Solorio (Eds.), Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Min-

neapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers), Association for Computational Linguistics, 2019, pp. 4171–4186. URL: <https://doi.org/10.18653/v1/n19-1423>. doi:10.18653/v1/n19-1423.