IJS@LT-EDI: Ensemble Approaches to Detect Signs of Depression from Social Media Texts

Jaya Caporusso[1,2], **Thi Hong Hanh Tran**[1,2,3] and **Senja Pollak**[2]
[1] Jožef Stefan International Postgraduate School, Jamova cesta 39, 1000 Ljubljana, Slovenia
[2] Jožef Stefan Institute, Jamova cesta 39, 1000 Ljubljana, Slovenia
[3] University of La Rochelle, L3i, La Rochelle, France
jaya.caporusso96@gmail.com
thi.tran@univ-lr.fr

Abstract

This paper presents our ensembling solutions for detecting signs of depression in social media text, as part of the Shared Task at LT-EDI@RANLP 2023. By leveraging social media posts in English, the task involves the development of a system to accurately classify them as presenting signs of depression of one of three levels: "severe", "moderate", and "not depressed". We verify the hypothesis that combining contextual information from a language model with local domain-specific features can improve the classifier's performance. We do so by evaluating: (1) two global classifiers (support vector machine and logistic regression); (2) contextual information from language models; and (3) the ensembling results. The best results were not achieved by any of the ensembling approaches, but by employing the RoBERTa language model.

1 Introduction[1]

In the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5), depressive disorders (which include major depressive disorder) are defined by "*the presence of sad, empty, or irritable mood, accompanied by somatic and cognitive changes that significantly affect the individual's capacity to function*" (American Psychiatric Association et al., 2013). Importantly, depression differs from regular feelings and mood changes, as it has an important impact on the lives of those who suffer from it. An extreme example of this is the high comorbidity between major depressive disorder and suicidal behavior (Orsolini et al., 2020), which is the cause of death of more than 700,000 people every year (World Health Organization, 2023). Depression is worryingly common: according to the World Health Organization, around 5 percent of adults suffer from depression globally (Villarroel and Terlizzi, 2020), and a survey conducted in 2019 in the US found that 18.5 percent of the participants had experienced some sort of depressive symptoms only in the two weeks preceding the study. The situation worsened due to the Covid-19 world pandemic: only in the first year, did the presence of anxiety or depression increase by 25 percent globally (World Health Organization, 2022). Another prominent aspect of nowadays is the use of social networks: in 2022, social media users were 4.59 billion (Statista, 2023). The relationship between the use of social networks and mental health has been studied from various points of view, one of them being the correlation between using social networks and suffering from depression (Park et al., 2015; Baker and Algorta, 2016). However, social networks can also represent a place where people suffering from depression can express and share how they feel and sometimes seek help. On date 15^{th} of June 2023, the subreddit "*depression, because nobody should be alone in a dark place: Peer support for anyone struggling with a depressive disorder*"[2], created on the 1^{st} of January 2009, counted 962,000 members. Therefore, social media posts provide precious data for the investigation and the (early) detection of depression (Leiva and Freire, 2017; Trotzek et al., 2018; Chiong et al., 2021; Liu and Shi, 2022; Ortega-Mendoza et al., 2022; Poświata and Perełkiewicz, 2022; Tavchioski et al., 2022b,a; Wang et al., 2022; Tavchioski et al., 2023).

DepSign-LT-EDI@RANLP-2023 (Detecting Signs of Depression from Social Media Text) (Sampath et al., 2023) is a shared task hosted by the Language Technology for Equality, Diversity, Inclusion workshop[3]. Its aim is to detect signs

[1] Trigger warning: the paper contains mentions of suicidal behavior and ideation.

[2] www.reddit.com/r/depression/

[3] sites.google.com/view/lt-edi-2023/

LT-EDI 2023 – Third Workshop on Language Technology for Equality, Diversity and Inclusion, pages 172–178, Varna, Bulgaria, Sep 7, 2023.
https://doi.org/10.26615/978-954-452-084-7_026

of depression from social media posts. More specifically, the goal is to produce a system that, given social media posts in English, classifies them as presenting signs of depression belonging to one of the following classes: “severe”, “moderate”, or “not depression”.

Our contributions are threefold. We evaluate (1) the standalone contextual information from language models; (2) the machine learning-based classifier with global information; and (3) the ensembling of the two best classifiers. Our work brings valuable insights for detecting signs of depression.

The paper is organized as follows. Section 2 provides an overview of related works. Section 3 discusses the dataset used in this research, while Section 4 explains the process of developing our solution. Section 5 shows the evaluation of the experiments, while the error analysis is presented in Section 6. Finally, Section 7 concludes the paper with our ideas for future work.

2 Related work

Aspects such as one’s personality, emotional state, ideology, and mental health are shown to be reflected in one’s language–not only in the semantics but also in the syntax used (Chung and Pennebaker, 2007; Pennebaker, 2011). Having depression as the focus opens two main paths: the analysis of which kind of language is used by individuals suffering from depression (approach a), and the detection of depression through language analysis (approach b). Furthermore, any of the two can contribute to the other (the first to the second (approach c) and the second to the first (approach d)). The first path showed, for example, that people suffering from depression use first-person singular pronouns more frequently (in spoken language: (Bucci and Freedman, 1981); in written text: (Ortega-Mendoza et al., 2022). Furthermore, as one would expect, the sentiment is more negative in individuals suffering from depression’s language (Babu and Kanaga, 2022). These and other aspects have been employed as features to detect depression from text data (approach c) (Trotzek et al., 2018; Babu and Kanaga, 2022; Ortega-Mendoza et al., 2022; Kolenik et al., 2023). Depression detection is becoming a trending topic in shared tasks. An example is the Depression and PTSD on Twitter task organized in the context of the Workshop on Computational Linguistics and Clinical Psychology[4] in 2015 (Coppersmith et al., 2015). The Early Risk Detection on the Internet workshop (eRisk) has been hosting shared tasks about depression, along with shared tasks about other mental-health-related issues, since 2017[5]. The Workshop on Language Technology for Equality, Diversity, and Inclusion (LT-EDI)[6], organized by the Association for Computational Linguistics[7] since 2021, has been including the shared task “Detecting Signs of Depression from Social Media Text” since 2022 (Sampath et al., 2022).

In the occasion of the 2022 edition (LT-EDI-ACL2022) (Sampath et al., 2022), the first and second place went, respectively, to (Poświata and Perełkiewicz, 2022) and (Wang et al., 2022). The latter employed VADER and sentence embeddings from pre-trained models to generate sentiment scores. Then, the authors adopted three different methods: a) gradient boosting models, b) pre-trained models, and c) contrastive pre-trained models. They finally ensembled the three approaches, obtaining the classifier which ranked second in the competition. The winning approach (Poświata and Perełkiewicz, 2022) included three main parts. a) In the first, transformer-based language models were fine-tuned on the train set. b) In the second, a corpus based on Reddit posts on mental health, depression, and suicide was created. A transformer-based language model was pre-trained on the corpus, resulting in DepRoBERTa (Poświata and Perełkiewicz, 2022). DepRoBERTa was then fine-tuned on the train set. c) In the third step, the best between the developed models were ensembled, obtaining the best-performing classifier of last year’s DepSign-LT-EDI shared task.

3 Dataset

We used the dataset provided by the organizers of the shared task. The dataset contains 7,201 instances for training, 3,245 instances for validation, and 499 for evaluation. Each sample is composed of three columns: *PID*, *Text*, and *Label*.

Table 1 shows the label distribution of the training, development, and test set with the overall samples per set. What is worth noting is that the dataset is imbalanced with the under-representation of the severe class, as shown in Figure 1. The sample instances (one for class) are shown in Table 2.

[4] clpsych.org

[5] erisk.irlab.org

[6] sites.google.com/view/lt-edi-2023/

[7] www.aclweb.org/portal/

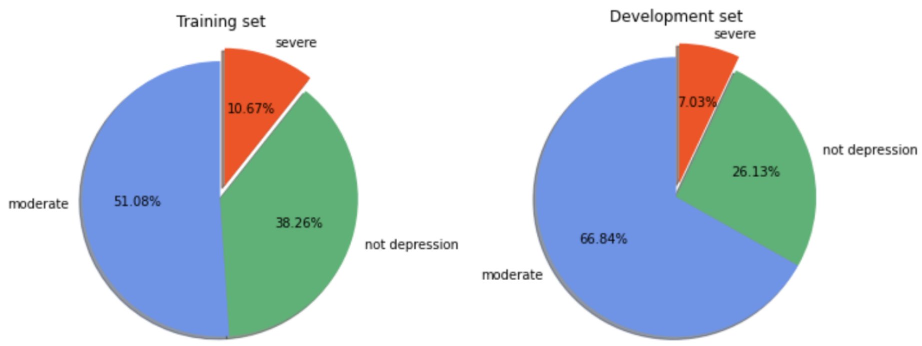

Figure 1: The distribution of sequence length for each class in the training and development sets.

Labels	Train	Dev.	Test
Not depression	2,755	848	135
Moderate	3,678	2,169	275
Severe	768	228	89
Total	7,201	3,245	499

Table 1: Data distribution by class.

PID	Text data	Label
train_pid_1	My life gets worse every year: That's what it feels like anyway....	moderate
train_pid_2	Words can't describe how bad I feel right now: I just want to fall asleep forever.	severe
train_pid_3	Is anybody else hoping the Coronavirus shuts everybody down?	not depressed

Table 2: Sample instances.

4 Methods

Our proposed solution involves four main steps (global classifiers, contextual classifiers, ensembling, and post-processing), as presented in the following subsections. By global classifiers, we mean classifiers that take non-contextual features into account, i.e., features that apply to the entire document, without considering the relationships between words or phrases. Contrarily, contextual classifiers do take such relationships into account.

4.1 Global classifiers

4.1.1 Features

Based on interdisciplinary knowledge of depression (Ratcliffe, 2014), we analyzed the distribution of certain features across the three classes of text. In particular, we considered: (1) the use of first-person singular pronouns; (2) the use of first-person pronouns (both singular and plural); (3) the use of third-person pronouns (both singular and plural); (4) the use of time-related terms; and (5) the sentiment analysis scores. To select the time-related terms, we started with a set of words identified by us, such as *time*, *now*, and *today*. We then expanded the list by finding synonyms and similar words in WordNet[8], which we further filtered. We obtained the following list: *today*, *now*, *soon*, *tomorrow*, *ago*, *yesterday*, *time*, *month*, *day*, *year*, *late*, *present*, *past*, *future*, *nowadays*, *instant*, *minute*, *second*, *early*, *young*, *old*, *recent*, *nowadays*, *hereafter*, *moment*. In Table 3, we display the statistics of all the features across the three classes. The fifth feature was analyzed by using Valence Aware Dictionary and sEntiment Reasoner (VADER) (Hutto and Gilbert, 2014). All the features proved to be statistically different across groups, besides the presence of third-person pronouns (in red). Although the presence of first-person pronouns (both singular and plural) was significantly different across groups, it was less significantly different than the sole presence of first-person-singular pronouns. Therefore, we did not further take it into account in our study (and we highlighted it in yellow in Table 3). The features that we further considered are highlighted in green in Table 3.

4.1.2 Models

We trained two global classifiers: support vector machine (SVM) and logistic regression (LR). In doing so, we used three attributes that we found to be different across groups as features: the presence of first-person-singular pronouns, the presence of time-related terms, and the sentiment scores. In particular, this was carried out by adapting part of the code developed by Koloski et al. (2021)[9].

[8] wordnet.princeton.edu
[9] bkolosk1/c19_rep

Features	"Not depression" (mean)	"Moderate" (mean)	"Severe" (mean)	Difference across groups
First-person singular pronouns (e.g., *I* and *my*)	0.0163	0.0275	0.0391	**F-Statistic**: 6.7717 **p-value**: 0.0011 **SIGNIFICANT**
First-person pronouns (e.g., *I* and *we*)	0.0189	0.0294	0.0417	**F-Statistic**: 6.0417 **p-value**: 0.0024 **SIGNIFICANT**
Third-person pronouns (e.g., *they* and *she*)	0.0374	0.0522	0.0573	**F-Statistic**: 2.0025 **p-value**: 0.1351 **NOT SIGNIFICANT**
Time-related terms (e.g., *now* and *soon*)	1.7877	2.9712	4.4102	**F-Statistic**: 138.6717 **p-value**: 8.0705e-32 **SIGNIFICANT**
Sentiment analysis	-0.0800	-0.2938	-0.4149	**F-Statistic**: 103.1463 **p-value**: 6.8241e-45 **SIGNIFICANT**

Table 3: Feature analysis results.

4.2 Contextual classifiers

Fine-tuning pre-trained language models has proved to be a successful approach to a wide range of downstream NLP tasks, especially since it does not require the effort of training a model from scratch. In our solution, we fine-tuned several commonly used English pre-trained language models in both monolingual and multilingual settings. We did so by following a standard procedure, which involves training a pre-trained language model with a classification head on top (i.e., a linear layer on top of the pooled output).

Parameters	Value
Max sequence length	512
Number of training epochs	5
Training batch size	16
Evaluation batch size	32
Learning rate	5e-5
Use early stopping	True
Early stopping patience	3
Manual seed	42

Table 4: Hyper-parameters configuration.

Specifically, we employed the following models: (1) Monolingual models: RoBERTa[10] (Liu et al., 2019), ALBERT[11] (Lan et al., 2019), BERT[12] (Devlin et al., 2018), XLNET[13] (Yang et al., 2019), DistilBERT[14] (Sanh et al., 2019); (2) Multilingual models: mBERT[15], mDistilBERT[16].

[10] roberta-base
[11] albert-base-v2
[12] bert-base-cased
[13] xlnet-base-cased
[14] distilbert-base-uncased
[15] bert-base-multilingual-cased
[16] distilbert-base-multilingual-cased

We utilized multilingual models due to their ability to: capture broader linguistic patterns, facilitate the cross-lingual transfer of knowledge, enhance contextual understanding with model robustness, allow for future adaptability to other languages, and address the issue of data scarcity. Additionally, we implemented distilled versions to reduce computational time.

We used the SimpleTransformers framework[17] (Rajapakse, 2019) to capture the level of depression per sentence while fine-tuning. All experiments were run on a single GPU Tesla V100 presenting the same standard hyperparameter configuration shown in Table 4 for better comparison.

4.3 Ensembling approach

In the last step, we combined the best models obtained in the previous stages by using ensemble averaging. This method involves averaging the predictions (expressed as probability) from a group of models. In particular, we utilized (1) the two global, (2) the two best contextual, and (3) the best global and the best contextual classifiers.

4.4 Post-processing steps

Unlike the training and development sets, the test set includes several non-English characters. Thus, a post-processing step was applied to filter out them.

4.5 Evaluation metrics

To evaluate the performance of our classifiers, we employed the following metrics: macro-averaged F1-score across all the classes; and Precision, Recall, and F1-score (F1) for each individual class.

[17] simpletransformers

Features	Model	Avg. Macro F1	Not Depression			Moderate			Severe		
			Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
Contextual	BERT	0.425	0.378	0.622	0.471	0.583	0.524	0.552	0.500	0.169	0.252
	XLNET	0.406	0.354	0.593	0.443	0.565	0.505	0.534	0.519	0.157	0.241
	ALBERT	0.383	0.386	0.630	0.479	0.548	0.524	0.535	0.438	0.079	0.133
	RoBERTa	0.447	0.389	0.622	0.479	0.592	0.560	0.576	0.696	0.180	0.286
	DistilBERT	0.418	0.368	0.607	0.458	0.566	0.513	0.538	0.556	0.169	0.259
	mBERT	0.418	0.332	0.563	0.418	0.566	0.498	0.530	0.643	0.202	0.308
	mDistilBERT	0.436	0.371	0.563	0.447	0.587	0.564	0.575	0.567	0.191	0.286
Global	LR	0.327	0.348	0.519	0.417	0.542	0.585	0.563	0.000	0.000	0.000
	SVM	0.331	0.352	0.570	0.435	0.554	0.564	0.559	0.000	0.000	0.000
Ensembling	LR + SVM	0.330	0.355	0.526	0.424	0.544	0.589	0.565	0.000	0.000	0.000
	RoBERTa + mDistilBERT	0.428	0.397	0.556	0.463	0.577	0.615	0.595	0.706	0.135	0.226
	RoBERTa + SVM	0.430	0.391	0.622	0.480	0.581	0.560	0.570	0.684	0.146	0.241

Table 5: Performance comparison.

5 Results

Table 5 shows the results obtained when applying our classifiers to the test set. Among the fine-tuned Transformer-based language models, the RoBERTa model presents the best Average Macro F1-score (0.447). Meanwhile, in machine learning-based approaches, SVM achieves the best Average Macro F1-score score (0.331). However, machine learning methods suffer from the lack of meaningful information to capture the "Severe" level of depression.

While testing with average ensembling, the combination of a language model with a machine learning classifier proves to perform better than combining two models of the same type. However, standalone RoBERTa still surpasses the performance of this combination. More machine learning-based features can be explored in the future.

6 Error Analysis

While previous studies (Poświata and Perełkiewicz, 2022; Wang et al., 2022) showed the final ensemble to be the best-performing approach, this is not the case in our study. In fact, RoBERTa shows an overall higher performance than any other contextual classifier, any global classifier, and any ensemble. One factor contributing to the fact that the best-performing ensemble, RoBERTa + SVM, has such a low performance, is that the class imbalance was not addressed properly when training the global classifiers. Besides, we noticed inconsistency among the training, development, and test sets. While the training and development sets contain only Latin alphabet characters, the test set contains special non-Latin characters (i.e., Chinese characters) as shown in Table 6. For example, the letter "*’t*" was decoded into 鈥檛 in the test set, which covers 24% of all the testing samples. Further data preprocessing steps would be required to solve this inconsistency.

Non-latin patterns	Ratio
鈥檛	0.240
鈥檓	0.214
鈥檓	0.214
鈥檚	0.196
鈥	0.092
鈥渟	0.008
...	...

Table 6: Examples of non-Latin characters in the test set given ratio is the frequency that the pattern appears in a sentence above all the given sentences.

7 Conclusion

In this paper, we presented a framework to detect signs of depression from social media text, as proposed by the LT-EDI shared task. Our procedure involved 4-steps: (1) the extraction of global information from language models, (2) the extraction of local information from SVM and LG classifiers, (3) average ensembling, and (4) post-processing. The results demonstrate the effectiveness of our framework. Our BERT-based model ranked 7^{th}/31 while our RoBERTa model could have achieved the 2^{nd}/31 in the LT-EDI competition. However, our ensemble approaches showed lower performance than our RoBERTa classifier.

In our future work, we intend to apply upsampling to imbalanced datasets like the one provided for DepSign-LT-EDI@RANLP-2023 and improve our feature engineering. Furthermore, depression detection could be followed by the development of interventional systems to support change in the context of mental health (Kolenik and Gams, 2021; Kolenik, 2022; Kolenik et al., 2023).

Acknowledgements

We acknowledge the financial support from the Slovenian Research Agency for research core funding for the programme Knowledge Technologies (No. P2-0103) and from the project CANDAS (Computer-assisted multilingual news discourse analysis with contextual embeddings, No. J6-2581). We wish to thank our colleagues at the Jožef Stefan Institute, Tine Kolenik, and the anonymous reviewers for their insightful comments.

References

D American Psychiatric Association, American Psychiatric Association, et al. 2013. *Diagnostic and statistical manual of mental disorders: DSM-5*, volume 5. American psychiatric association Washington, DC.

Nirmal Varghese Babu and E Grace Mary Kanaga. 2022. Sentiment analysis in social media data for depression detection using artificial intelligence: a review. *SN Computer Science*, 3:1–20.

David A Baker and Guillermo Perez Algorta. 2016. The relationship between online social networking and depression: A systematic review of quantitative studies. *Cyberpsychology, Behavior, and Social Networking*, 19(11):638–648.

Wilma Bucci and Norbert Freedman. 1981. The language of depression. *Bulletin of the Menninger Clinic*, 45(4):334.

Raymond Chiong, Gregorius Satia Budhi, Sandeep Dhakal, and Fabian Chiong. 2021. A textual-based featuring approach for depression detection using machine learning classifiers and social media texts. *Computers in Biology and Medicine*, 135:104499.

Cindy Chung and James W Pennebaker. 2007. The psychological functions of function words. *Social communication*, 1:343–359.

Glen Coppersmith, Mark Dredze, Craig Harman, Kristy Hollingshead, and Margaret Mitchell. 2015. Clpsych 2015 shared task: Depression and ptsd on twitter. In *Proceedings of the 2nd workshop on computational linguistics and clinical psychology: from linguistic signal to clinical reality*, pages 31–39.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Clayton Hutto and Eric Gilbert. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media*, volume 8, pages 216–225.

Tine Kolenik. 2022. Methods in digital mental health: smartphone-based assessment and intervention for stress, anxiety, and depression. In *Integrating Artificial Intelligence and IoT for Advanced Health Informatics: AI in the Healthcare Sector*, pages 105–128. Springer.

Tine Kolenik and Matjaž Gams. 2021. Intelligent cognitive assistants for attitude and behavior change support in mental health: state-of-the-art technical review. *Electronics*, 10(11):1250.

Tine Kolenik, Günter Schiepek, and Matjaž Gams. 2023. Computational psychotherapy system for mental health prediction and behavior change with a conversational agent. Manuscript submitted for publication.

Boshko Koloski, Timen Stepišnik-Perdih, Senja Pollak, and Blaž Škrlj. 2021. Identification of covid-19 related fake news via neural stacking. In *Combating Online Hostile Posts in Regional Languages during Emergency Situation: First International Workshop, CONSTRAINT 2021, Collocated with AAAI 2021, Virtual Event, February 8, 2021, Revised Selected Papers 1*, pages 177–188. Springer.

Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, and Radu Soricut. 2019. Albert: A lite bert for self-supervised learning of language representations. *arXiv preprint arXiv:1909.11942*.

Victor Leiva and Ana Freire. 2017. Towards suicide prevention: early detection of depression on social media. In *Internet Science: 4th International Conference, INSCI 2017, Thessaloniki, Greece, November 22-24, 2017, Proceedings 4*, pages 428–436. Springer.

Jingfang Liu and Mengshi Shi. 2022. A hybrid feature selection and ensemble approach to identify depressed users in online social media. *Frontiers in Psychology*, 12:6571.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Laura Orsolini, Roberto Latini, Maurizio Pompili, Gianluca Serafini, Umberto Volpe, Federica Vellante, Michele Fornaro, Alessandro Valchera, Carmine Tomasetti, Silvia Fraticelli, et al. 2020. Understanding the complex of suicide in depression: from research to clinics. *Psychiatry investigation*, 17(3):207.

Rosa María Ortega-Mendoza, Delia Irazú Hernández-Farías, Manuel Montes-y Gómez, and Luis Villaseñor-Pineda. 2022. Revealing traces of depression through personal statements analysis in social media. *Artificial Intelligence in Medicine*, 123:102202.

Sungkyu Park, Inyeop Kim, Sang Won Lee, Jaehyun Yoo, Bumseok Jeong, and Meeyoung Cha. 2015. Manifestation of depression and loneliness on social networks: a case study of young adults on facebook. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*, pages 557–570.

James W Pennebaker. 2011. The secret life of pronouns. *New Scientist*, 211(2828):42–45.

Rafał Poświata and Michał Perełkiewicz. 2022. Opi@ lt-edi-acl2022: Detecting signs of depression from social media text using roberta pre-trained language models. In *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 276–282.

TC Rajapakse. 2019. Simple transformers. *URL: https://simpletransformers. ai/[accessed 2022-08-25]*.

Matthew Ratcliffe. 2014. *Experiences of depression: A study in phenomenology*. OUP Oxford.

Kayalvizhi Sampath, Thenmozhi Durairaj, Bharathi Raja Chakravarthi, and Jerin Mahibha C. 2022. Findings of the shared task on detecting signs of depression from social media. In *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 331–338, Dublin, Ireland. Association for Computational Linguistics.

Kayalvizhi Sampath, Thenmozhi Durairaj, Bharathi Raja Chakravarthi, Jerin Mahibha C, Kogilavani Shanmugavadivel, and Pratik Anil Rahood. 2023. Overview of the second shared task on detecting signs of depression from social media text. In *Proceedings of the Third Workshop on Language Technology for Equality, Diversity and Inclusion*, Varna, Bulgaria. Recent Advances in Natural Language Processing.

Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2019. Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*.

Statista. 2023. Number of social media users worldwide from 2017 to 2027. `https://www.statista.com/statistics/278414/number-of-worldwide-social-network/-users/`. Accessed on 2023-06-22.

Ilija Tavchioski, Boshko Koloski, Blaž Škrlj, and Senja Pollak. 2022a. E8-ijs@ lt-edi-acl2022-bert, automl and knowledge-graph backed detection of depression. In *Proceedings of the second workshop on language technology for equality, diversity and inclusion*, pages 251–257.

Ilija Tavchioski, Marko Robnik-Šikonja, and Senja Pollak. 2023. Detection of depression on social networks using transformers and ensembles. *arXiv preprint arXiv:2305.05325*.

Ilija Tavchioski, Blaž Škrlj, Senja Pollak, and Boshko Koloski. 2022b. Early detection of depression with linear models using hand-crafted and contextual features. *Working Notes of CLEF*, pages 5–8.

Marcel Trotzek, Sven Koitka, and Christoph M Friedrich. 2018. Utilizing neural networks and linguistic metadata for early detection of depression indications in text sequences. *IEEE Transactions on Knowledge and Data Engineering*, 32(3):588–601.

Maria A Villarroel and Emily P Terlizzi. 2020. *Symptoms of depression among adults: United States, 2019*. US Department of Health and Human Services, Centers for Disease Control and ....

Wei-Yao Wang, Yu-Chien Tang, Wei-Wei Du, and Wen-Chih Peng. 2022. Nycu_twd@ lt-edi-acl2022: Ensemble models with vader and contrastive learning for detecting signs of depression from social media. In *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 136–139.

World Health Organization. 2022. Covid-19 pandemic triggers 25% increase in prevalence of anxiety and depression worldwide. `https://www.who.int/news/item/02-03-2022-covid-19-pandemic-\triggers-25-increase-in-\prevalence-of-anxiety-and-\depression-worldwide`. Accessed on 2023-06-22.

World Health Organization. 2023. Depressive disorder (depression). `https://www.who.int/news-room/fact-sheets/detail/depression`. Accessed on 2023-06-22.

Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Russ R Salakhutdinov, and Quoc V Le. 2019. Xlnet: Generalized autoregressive pretraining for language understanding. *Advances in neural information processing systems*, 32.