TechSSN4@LT-EDI-RANLP2023: Depression Sign Detection in Social Media Postings using DistilBERT Model

Krupa Elizabeth Thannickal, Sanmati P, Rajalakshmi S, Angel Deborah S
Department of Computer Science and Engineering,
Sri Sivasubramaniya Nadar College of Engineering, Chennai - 603110, Tamil Nadu, India
krupa19054@cse.ssn.edu.in, sanmati19098@cse.ssn.edu.in,
rajalakshmis@ssn.edu.in, angeldeborahs@ssn.edu.in

Abstract

As world population increases, more people are living to the age when depression or Major Depressive Disorder (MDD) commonly occurs. Consequently, the number of those who suffer from such disorders is rising. There is a pressing need for faster and reliable diagnosis methods. This paper proposes a method to analyse text input from social media posts of subjects to determine the severity class of depression. We have used the DistilBERT transformer to process these texts and classify the posts across three severity labels - 'not depression', 'moderate' and 'severe'. The results showed the macro F1-score of 0.437 when the model was trained for 5 epochs with a comparative performance across the labels.The team acquired 6^{th} rank while the top team scored macro F1-score as 0.470. We hope that this system will support further research into the early identification of depression in individuals to promote effective medical research and related treatments.

1 Introduction

The role of deep learning (DL) is growing in importance when it comes to automating the diagnosis and treatment of diseases, particularly in the field of mental health. Mental health disorders are a major cause of disability globally, as reported by the World Health Organization (WHO).

Depression, or Major Depressive Disorder (MDD), is denoted by symptoms such as low mood, loss of interest in activities, and negative effects on thoughts, behavior, motivation, and emotions, which can even lead to suicide. The causes of depression can be attributed to various factors, including biological, social, and psychological influences. According to WHO's 2022 report (Freeman, 2022), approximately 970 million people worldwide are living with mental disorders, of which 28.9% are depressive disorders.

While there are different treatment options available including psychological therapies like behavioral activation and problem-solving therapy, self-care plays a significant role in addressing depression. Accurate identification of mental health conditions is vital for effective care, especially considering the shortage of psychiatrists in many areas. Depression, as a prevalent mental health disorder, requires early identification and a comprehensive approach combining various treatment options to tackle its significant impact on individuals and society.

The following paper proposes a method to analyse text input from social media posts of subjects to determine the severity class of depression. Our work covers the research question of how the application of a developed typology for social media texts, which aims to detect depression severity, contribute to effectively identifying subtle signs of depressive disorders in tweets. We have used the DistilBERT transformer to process these texts and classify the posts across three severity labels - 'not depression', 'moderate' and 'severe'.

2 Related Work

Studies have been carried out to detect the presence or absence of depression in individuals by analysing any one or a combination of text, audio and video inputs.

Islam et al. (2018) present their efforts to analyze depression based on Facebook data obtained from an online public source utilizing machine learning (ML) techniques. Their findings resulted in a substantially enhanced accuracy and reduced classification error rates, demonstrating that Decision Tree (DT) models outperform other simpler ML methods.

(Sadeque et al., 2018) attempt to overcome the shortcomings in accurately assessing model latency during depression detection. They have identified concerns regarding the widely used ERDE metric and put forth an alternative measure called latency-weighted F1, which effectively addresses these is-

LT-EDI 2023 – Third Workshop on Language Technology for Equality, Diversity and Inclusion, pages 239–243, Varna, Bulgaria, Sep 7, 2023.
https://doi.org/10.26615/978-954-452-084-7_036

sues. Subsequently, this evaluation approach is used to assess multiple models as part of the eRisk 2017 shared task on depression detection. Their results showed more effective distinctions captured between systems by using this metric.

Another study merges posts of users from two platforms, Twitter and Facebook, in order to detect the level of depression (Asad et al., 2019). This research employs ML techniques to classify data using Support Vector Machine (SVM) and Naïve Bayes algorithms and potentially identify depression.

A survey-based study carried out by Zafar and Chitnis (2020) indicated the increasing interest in data-mining and analysis of information from social networking sites for recognizing depression in users. Yasaswini et al. (2021) use tweets from Twitter to examine users' expressions and gain insights into their emotional states. Their use of the DistilBERT model for a binary classification of depressed or non-depressed subjects resulted in an enhanced accuracy.

With respect to DL models, a comparison study was carried out by Senn et al. (2022). They considered three variants of BERT and four ensemble models of these variants in classifying depression using transcripts of responses to 12 clinical interview questions. Their findings reveal that the utilization of ensembles leads to improved mean F1 scores.

As part of DepSign-LT-EDI@ACL-2022, Janatdoust et al. (2022) present a predictive ensemble model that leverages the fine-tuned contextualized word embeddings from DistilBERT, BERT, RoBERTa and ALBERT base models. Their findings demonstrated a performance surpassing the baseline models across all evaluated metrics, achieving an impressive 61% accuracy and F1 score of 54%.

A predictive model from text has also been developed using Long-Short Term Memory (LSTM) and Recurrent Neural Network (RNN) models (Amanat et al., 2022). The RNN is trained on text-based data to recognize depression using semantics and written content. With a 99.0% accuracy rate, this framework performed better than frequency-based DL models for textual detection, and has a lower false positive rate.

Zavorina and Makarov (2021) use a transformer encoder model for their research on voice-base depression detection. In order to address the limited size of the available dataset, the researchers extracted low-level features from audio recordings and applied augmentation techniques. By leveraging these approaches, their network achieved a recognition accuracy of 73.51% on the E-DAIC database. Anantharaman et al. (2022) uses BERT model for detecting depression from text while Esackimuthu et al. (2022) uses AlBERT model for depression detection.

3 Proposed System

Our depression detection system involves data preprocessing, encoding, model building and predicting the unseen test samples. Various experiments were conducted and it was concluded that the DistilBERT model outperformed other models in terms of accuracy, leading to its selection for text processing. The model is implemented using Python's torch library based on the transformer architecture. It takes the encoded input text and utilizes a pretrained "distilbert-base-uncased" model to generate contextualized embeddings for each token. In order to mitigate the risk of overfitting, a dropout layer is included, employing a dropout rate of 0.1. Additionally, a linear classification layer is employed to map the hidden state size of the DistilBERT model to the desired number of output classes. As a result, logits are obtained, representing raw scores that indicate the likelihood of the input belonging to each output class.

Our system comprises of the DistilBERT model trained using the train and dev sets provided, over 5 epochs. The trained model was subsequently evaluated on the test dataset, yielding an accuracy of 47.9%.

4 Dataset and Methodologies Used

4.1 Dataset Used

The dataset used is provided as part of the DepSign-LT-EDI@RANLP-2023 challenge (Kayalvizhi et al., 2022; Evans-Lacko et al., 2017; Losada et al., 2017). It comprises labelled training and development sets of 7201 and 3245 texts respectively. Testing is carried out over a set of 499 social media posts. The distribution of training and development set data across the three severity labels is seen in Table 1.

Dataset balancing has not been carried out as can be seen in Table 1 where the number of samples for the Severe class is much lesser than those for the Not Depression and Moderate classes. Since

Classes	Train	Dev	Test
Not Depression	2755	848	135
Moderate	3678	2169	275
Severe	768	228	89
Total samples	7201	3245	499

Table 1: Class distribution over train, dev and test sets

the dataset focuses on social media texts, this may limit its generalizability to other forms of communication or contexts. It also may not capture the impact of external factors, such as cultural differences, language variations, or contextual elements, which can influence the interpretation of social media texts and the detection of depression.

4.2 DistilBERT

DistilBERT is a transformer-based text model used in tasks involving Natural Language Processing (NLP) like translation and text classification (Sanh et al., 2019). Although based on the BERT network, it relies on the idea of knowledge distillation during the pre-training phase itself to reduce the size of the model and make it faster. The DistilBERT architecture comprises several transformer layers that are each fed a sequence of contextualized embeddings generated by encoding the input text. The transformer layers learn to grasp connections among the tokens within the sequence, resulting in the production of more significant representations of the input text.

4.3 Implementation Modules

The system comprises of two modules corresponding to the stages of training and testing. The training module is where the DistilBERT model is defined. The model takes the train and development set as input and performs training for 5 epochs. On completion of training, the model is evaluated with the development dataset and the parameters are tweaked. The final model is saved for later use.

The testing module loads the previously trained model and then feeds the test set as input to it. The model is used to classify the test inputs as one of the three depression severity classes.

The results of testing are evaluated against the expected outcomes in terms of accuracy for the whole set and class-wise. The confusion matrix is also plotted in order to better analyse the performance of the model for individual classes and check for scope of improvements.

5 Results and Discussion

The performance of DistilBert model trained on depression dataset is evaluated on test dataset to predict the output depression class labels. The overall performance of the system was determined across several metrics including accuracy and weighted and macro averaged values of precision, recall and F1-Score and is tabulated in Table 2. The system achieved the macro F1 score of 0.437 and secured 6^{th} rank while the first rank team achieved 0.470 macro F1 score.

Metric	Value
Accuracy	0.479
Macro Precision	0.501
Macro Recall	0.436
Macro F1-Score	0.437
Weighted Precision	0.509
Weighted Recall	0.479
Weighted F1-Score	0.475

Table 2: Performance metrics for test dataset

Accuracy measures the effectiveness of classification models by determining the proportion of accurate predictions out of the total predictions made, represented as a percentage while F1 score gives the blend of precision and recall. The class-wise accuracy and F1-scores of the model are depicted in Table 3. We can infer that the classes 'not depression' and 'moderate' have a better accuracy compared to the 'severe' class. Similarly, these classes have a higher F1-score as well, compared to the 'severe' class label.

Classes	Accuracy	F1-Score
Not Depression	54.81	0.54
Moderate	52.36	0.44
Severe	23.59	0.34

Table 3: Individual class accuracy for test data

The confusion matrix was also visualized over the results for a better class-wise comparison as seen in Table 4. This matrix also presents the scope of improvement of model performance for individual classes.

Labels	**Not Depression**	**Moderate**	**Severe**
Not Depression	74	59	2
Moderate	118	144	13
Severe	12	56	21

Table 4: Confusion matrix for test dataset

It is noticed from Table 3 that the accuracy for severe class is half of the other classes. It is inferred that the reason for this low score is because the total number of samples in severe class is very less for learning when compared to the moderate and non depression classes. Also from the confusion matrix in Table 4, the severe cases are classified as moderate, since most of the severe class sentences more or less give the meaning of moderate depression class. There is a thin line dividing the moderate and severe class.

It is also noticed that 59 non-depression cases are considered as moderate and 2 as severe, which leads to a major problem. These non-depression classes need to be given more importance as we are mainly concentrating on the depression classes alone. We have to investigate more on the methods to differentiate these classes in a fine-grained manner in future.

6 Conclusion and Future Work

Our DistilBERT model was trained by combining both the train and dev sets of data, but other approaches could consider these separately and improve the model through a validation phase post training. Another technique that can improve training is running the model for a larger number of epochs, allowing for more iterations and potentially better learning. Prior to training and testing, it is also crucial to ensure the cleanliness of the text data, as incorrect encoding and decoding can lead to the loss of important information.

While this study focuses on detecting depression through social media postings it can be extended for analysis of longer texts that could also be speech transcripts. However, analyzing longer sequences of text poses a challenge as prevalent models in NLP typically have limited input lengths.

While mental health awareness has increased, there is still a stigma associated with depression. The implications for society with the availability of a rapid and reliable depression detection system are immense. Early identification of such conditions can result in enhanced treatment results and a higher standard of living for individuals impacted by them.

References

Amna Amanat, Muhammad Rizwan, Abdul Rehman Javed, Maha Abdelhaq, Raed Alsaqour, Sharnil Pandya, and Mueen Uddin. 2022. Deep learning for depression detection from textual data. *Electronics*, 11(5):676.

Karun Anantharaman, S Rajalakshmi, S Angel Deborah, M Saritha, and R Sakaya Milton. 2022. Ssn_mlrg1@ lt-edi-acl2022: Multi-class classification using bert models for detecting depression signs from social media text. *LTEDI 2022*.

Nafiz Al Asad, Md. Appel Mahmud Pranto, Sadia Afreen, and Md. Maynul Islam. 2019. Depression detection by analyzing social media posts of user. In *2019 IEEE International Conference on Signal Processing, Information, Communication Systems (SPICSCON)*, pages 13–17.

Sarika Esackimuthu, H Shruthi, Rajalakshmi Sivanaiah, S Angel Deborah, R Sakaya Milton, and TT Mirnalinee. 2022. Ssn_mlrg3@ lt-edi-acl2022-depression detection system from social media text using transformer models. *LTEDI 2022*, page 196.

Sara Evans-Lacko, Sergio Aguilar-Gaxiola, Ali Al-Hamzawi, Jordi Alonso, Corina Benjet, Ronny Bruffaerts, W. Chiu, Silvia Florescu, Giovanni de Girolamo, Oye Gureje, Josep Maria Haro, Y. He, Chenchung Hu, Elie Karam, Norito Kawakami, Sing Lee, Crick Lund, Viviane Kovess, Daphna Levinson, and G. Thornicroft. 2017. Socio-economic variations in the mental health treatment gap for people with anxiety, mood, and substance use disorders: Results from the who world mental health (wmh) surveys. *Psychological Medicine*, 48:1–12.

Melvyn Freeman. 2022. The world mental health report: transforming mental health for all. *World psychiatry: official journal of the World Psychiatric Association (WPA)*, 21(3):391–392.

Md Rafiqul Islam, Muhammad Ashad Kabir, Ashir Ahmed, Abu Raihan M. Kamal, Hua Wang, and Anwaar Ulhaq. 2018. Depression detection from social network data using machine learning techniques. *Health information science and systems*, 6(1).

Morteza Janatdoust, Fatemeh Ehsani-Besheli, and Hossein Zeinali. 2022. KADO@LT-EDI-ACL2022: BERT-based ensembles for detecting signs of depression from social media text. In *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 265–269, Dublin, Ireland. Association for Computational Linguistics.

S Kayalvizhi, 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. Association for Computational Linguistics.

David E. Losada, Fabio Crestani, and Javier Parapar. 2017. erisk 2017: Clef lab on early risk prediction on the internet: Experimental foundations. In *Experimental IR Meets Multilinguality, Multimodality, and Interaction*, pages 346–360, Cham. Springer International Publishing.

Farig Sadeque, Dongfang Xu, and Steven Bethard. 2018. Measuring the latency of depression detection in social media. page 495–503, New York, NY, USA. Association for Computing Machinery.

Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2019. Distilbert, a distilled version of BERT: smaller, faster, cheaper and lighter. *CoRR*, abs/1910.01108.

Saskia Senn, ML Tlachac, Ricardo Flores, and Elke Rundensteiner. 2022. Ensembles of bert for depression classification. In *2022 44th Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, pages 4691–4694.

U. Yasaswini, Y. Sasidhar, P. Siva Sai, P. Eswar, and V. Swathi. 2021. Detecting depression in tweets using distilbert. *SSRN Electronic Journal*.

Aqsa Zafar and Sanjay Chitnis. 2020. Survey of depression detection using social networking sites via data mining. In *2020 10th International Conference on Cloud Computing, Data Science Engineering (Confluence)*, pages 88–93.

Evgeniya Zavorina and Ilya Makarov. 2021. Depression detection by person's voice. In *Analysis of Images, Social Networks and Texts: 10th International Conference, AIST 2021, Tbilisi, Georgia, December 16–18, 2021, Revised Selected Papers*, page 250–262.