



Use Of RoBERTa Model To Analyze Depressive Texts

¹Shreya Rani, ²Viraj Patil, ³Fatema Katawala, ⁴Aaron DSouza, ⁵Ratnamala Paswan

^{1,2,3,4}Students at Pune Institute of Computer Technology, Pune

⁵Assistant Professor, Department of Computer Engineering, Pune Institute of Computer Technology, Pune

Abstract : Mental health is the state of well-being of an individual. It includes emotional, social, and psychological conditions. It affects an individual's everyday life, how one feels, thinks, handles stress, makes choices, and acts as we cope with life. Our mental health is crucial at every stage of life, from childhood through adulthood and aging. Unfortunately, not much importance or support is given to our mental health.

Mental illness is one of the leading problems worldwide. According to a study, more than 792 million suffered from any mental health disorder in 2017. More than one in ten people face such issues globally. Depression and anxiety disorder are the major contributors. And it's getting worse in the pandemic. A large number of depressed people commit suicide globally. Most of the time, people find it difficult to express their thoughts and feelings and sometimes take it to social media. So we created our custom dataset from Twitter and Reddit that is available publicly: [ShreyaR/DepressionDetection](#). And with the help of Natural Language Processing (NLP), we can identify depressive phrases using fine-tuned RoBERTa model and classify the severity of depression based on a questionnaire. It can be helpful to friends and family to determine the mental health of their loved ones and provide support and prevent worse situations like suicides. The fine-tuned model can be found on HuggingFace: [finetuned-roberta-depression](#).

IndexTerms - Natural Language Processing, Transformer, RoBERTa, Depression, Twitter, Reddit.

I. INTRODUCTION

Depression is a common illness and a concern for everyone, this illness is increasing day by day and is observed in people of all age groups, and it further leads to suicide. Various helplines and wellness centers have been working to create awareness for lowering the rate of suicide and for creating awareness for people with depression to connect with professionals and get the necessary treatment. With the rise in popularity of social media, people tend to text freely about their worries, state of mind, and everyday happening on social media than talk about it to people. They also share their journey fighting depression, taking steps to lead a happy life. This helps people with similar conditions to get motivation and inspiration to deal with it and fight it. Because of social media, where people deal with depression as a community we have textual data that can be useful in the early detection of depression. Depressive texts can be identified and detected using sophisticated natural language processing techniques. Technology is evolving and helping people deal with problems. With the new natural language processing methods and tools, depression can be detected in an early stage and prevent any self-harm. The idea is to monitor posts over social media and assess the state of the person posting them. This paper includes a fine-tuned RoBERTa model trained on a custom dataset consisting of tweets and subreddits to identify depressive phrases and classify the severity of depression based on a questionnaire - Burns Depression Checklist.

II. LITERATURE REVIEW

- A. USDB at eRisk 2020: Deep learning models to measure the Severity of the Signs of Depression using Reddit Posts [1]
 - The main objective is to measure the severity of the signs of depression by evaluating history of posts of an individual.
 - Experiments were carried on CNN and BiLSTM separately and results obtained were similar.
 - It was believed that a combination of CNN model with BiLSTM could improve the process of feature extraction.
- B. Early Detection of Depression: Social Network Analysis and Random Forest Techniques [2]
 - After an analysis on the dataset, subjects' behaviour is characterized on the basis of their writing such as time gap, textual spreading and time span.
 - Two approaches were introduced, one of them used one random forest classifier with two threshold functions, and the other one used two independent random forest classifiers to detect depressed and non depressed individuals respectively.

- Features included textual, writing, and semantic similarities in both the approaches.

C. Detection and Classification of mental illnesses on social media using RoBERTa [3]

- Analyzing user data on social media to detect and classify mental illness.
- It was tested on LSTM, BERT and RoBERTa.
- Transformer based models such as BERT and RoBERTa performed better than LSTM, and RoBERTa achieved the highest accuracy.
- Automation of detection and classification process in the public health system.

D. Depression Risk Prediction for Chinese Microblogs via Deep-Learning Methods: Content Analysis [4]

- Investigation of BERT, RoBERTa and XLNET for depression risk prediction.
- Depression risks were assessed at 4 levels - 0 (no inclination), 1 (mild), 2 (moderate) and 3 (severe).
- Weibo, a Chinese microblog was used to collect dataset.
- BERT performed the best with a microaveraged F1 score of 0.856.

E. Depression Detection from Social Media Texts [5]

- Early depression detection from textual messages of VKontakte users, a Russian social network.
- Proposal of psycho-linguistic and stylistic markers of depression in a text.
- Beck Depression Inventory is a questionnaire to calculate depression score on a scale of 0–63.
- Experiments were carried out using RF and SVM.

Due to increasing use of a social media in this era and also as effect of COVID-19 people are expressing themselves through posts on different social medias like Reddit, Twitter etc. It is creating a huge opportunity for researchers to gather data and perform research using different NLP techniques and algorithms. Like [2] and [3] used Reddit data to detect and classify metal illness like depression, PTSD, anxiety etc. There use of dataset like eRisk 2020 which is used in [1], data from Weibo used in [4] and also CLEF eRisk 2017 dataset used in [5].

There are variety of NLP techniques and latest algorithms based on neural networks which are highly efficient than classical techniques. Like transformers based models RoBERTa, BERT and LSTM neural network model used for Classification of five prominent kinds of mental illnesses, depression, anxiety, bipolar disorder, ADHD and PTSD in [3]. And also [4] came up with the approach of using BERT, RoBERTa, XLNET for Depression risk prediction and it's severity using data from Chinese microblogs, Weibo. Also the NLP techniques like random forest and SVM are used by researchers, in [2] random forest model is used to detect depressed subjects and non-depressed subjects based on features defined from textual, semantic, and writing similarities. [5] uses SVM, CNN for automatic detection of depression signs from textual messages of Russian social network VKontakte user.

So, here we are using a fine-tuned RoBERTa model, as it is one of the latest transformer which gives high performance to solve this difficult problem of depression detection.

III. METHODOLOGY

A. Data Collection

Due to the unavailability of datasets required for depression detection, a custom dataset was created by scrapping Twitter with the help of Twint using keywords such as depression. We extracted posts from subreddits like r/Anxiety, r/Depression, r/SuicideWatch using Reddit API. For non depressive tweets, Sentiment140 dataset [6] was used. We also added a questionnaire, Burns Depression Checklist, which consists of 15 questions to check the severity of depression.

Dataset is available publicly: [ShreyaR/DepressionDetection](#)

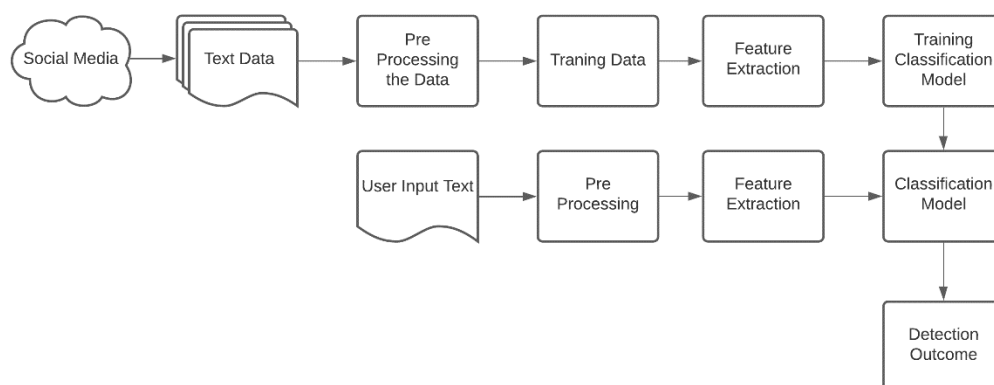


Fig. 1: Data Flow Diagram

B. Model

- i. RoBERTa[7]: RoBERTa stands for Robustly Optimized BERT Pre-training Approach. It has a similar architecture as compared to BERT [8].

RoBERTa model was trained on five datasets:

- BookCorpus
- English Wikipedia
- CC-News
- OpenWebText
- Stories

Features of RoBERTa:

- More training data (160GB)
- Uses dynamic masking pattern rather than static masking pattern
- Training is done on longer sequences
- Replacing NSP objective with full sentences without NSP

- ii. DistilBERT[9]: It is a small, fast, cheap and light transformer model. It has a similar architecture as compared to BERT. To reduce the size of a BERT model [8] by 40%, knowledge distillation is performed during the pretraining phase.

DistilBERT model was trained on datasets:

- BookCorpus
- English Wikipedia

C. Training Model

Both the pretrained models, RoBERTa and DistilBERT were fine-tuned using our custom dataset as mentioned above using Trainer API and then the model was pushed on HuggingFace. For training a model we split the dataset into training dataset and testing dataset and then perform normalization. 70% and 30% of data is used for training and testing respectively.

The fine-tuned RoBERTa model can be found on HuggingFace: [finetuned-roberta-depression](#)

Training hyperparameters are as follows:

- learning_rate: 5e-05
- train_batch_size: 8
- eval_batch_size: 8
- seed: 42
- optimizer: Adam with betas=(0.9,0.999) and epsilon=1e-08
- lr_scheduler_type: linear
- num_epochs: 3.0

The fine-tuned DistilBERT model can be found on HuggingFace: [finetuned-distil-bert-depression](#)

Training hyperparameters are as follows:

- learning_rate: 5e-05
- train_batch_size: 8
- eval_batch_size: 8
- seed: 42
- optimizer: Adam with betas=(0.9,0.999) and epsilon=1e-08
- lr_scheduler_type: linear
- num_epochs: 3.0

IV. RESULTS

The fine-tuned RoBERTa model was tested and evaluation results achieved were as follows:

- Loss: 0.1385
- Accuracy: 0.9745

Training Loss	Epoch	Validation Loss	Accuracy
0.0238	1.0	0.1385	0.9745
0.0333	2.0	0.1385	0.9745
0.0263	3.0	0.1385	0.9745

Table 1: Training Results for RoBERTa Model

The fine-tuned DistilBERT model was tested and evaluation results achieved were as follows:

- Loss: 0.1695
- Accuracy: 0.9445

Training Loss	Epoch	Validation Loss	Accuracy
0.0243	1.0	0.2303	0.9205
0.0341	2.0	0.1541	0.933
0.0244	3.0	0.1495	0.9445

Table 2: Training Results for DistilBERT Model

V. CONCLUSION

Our main motivation and purpose for this paper is to detect depression which is increasing in this pandemic situation. To do so we are using a different number of posts from users across different social media platforms as a dataset. We fine-tuned RoBERTa and DistilBERT models for this purpose. RoBERTa model performed better with an accuracy of 97.45%.

VI. ACKNOWLEDGMENT

The authors would like to thank Prof. R. S. Paswan for her help and support in making this work possible.

VII. RESOURCES

Website: <https://sense-it-depression-detection.herokuapp.com/>
 Model: <https://huggingface.co/ShreyaR/finetuned-roberta-depression>
 Dataset: <https://huggingface.co/datasets/ShreyaR/DepressionDetection>

REFERENCES

- [1] Amina Madani, Fatima Boumahdi, Anfel Boukenaoui, Kritli Mohamed Chaouki, and Hentabli Hamza. Usdb at erisk 2020: Deep learning models to measure the severity of the signs of depression using reddit posts. 09 2020.
- [2] Fidel Cacheda, Diego Fernandez, Francisco J Novoa, and Victor Carneiro. Early detection of depression: Social network analysis and random forest techniques. *J Med Internet Res*, 21(6), Jun 2019.
- [3] Ankit Murarka, Balaji Radhakrishnan, and Sushma Ravichandran. Detection and classification of mental illnesses on social media using roberta, 2020.
- [4] Tao Li Wanting Li Yejie Zhou Jie Zheng Qingcai Chen Jun Yan Xiaofeng Wang, Shuai Chen and Buzhou Tang. Depression risk prediction for chinese microblogs via deep-learning methods: Content analysis, 2020.
- [5] Evgenia Kuminskaya Ivan Smirnov Maxim Stankevich, Andrey Latyshev and Oleg Grigoriev. Depression detection from social media texts, 2018.
- [6] Alec Go, Richa Bhayani, and Lei Huang. Twitter sentiment classification using distant supervision. *CS224N project report, Stanford*, 1(12):2009, 2009.
- [7] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach, 2019.
- [8] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding, 2018.
- [9] Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter, 2019