TechWhiz@LT-EDI : Transformer Models to Detect Levels of Depression from Social Media Text

Madhumitha M, C.Jerin Mahibha
Meenakshi Sundararajan Engineering College, Chennai
madhumithamurthy2002@gmail.com
jerinmahibha@gmail.com

Durairaj Thenmozhi
Sri Sivasubramaniya Nadar College of Engineering, Chennai
theni_d@ssn.edu.in

Abstract

Depression is a mental fitness disorder characterised by persistent reactions of unhappiness, voiding, and a deficit of interest in activities. It can influence differing facets of one's life, containing their hopes, sympathy, and nature. Depression can stem from a sort of determinant, in the way that ancestral willingness, life occurrences, and social circumstances In recent years, the influence of social media on mental fitness has become an increasing concern. Excessive use of social media and the negative facets that guide it can exacerbate or cause impressions of distress. Nonstop exposure to cautiously curated lives, social comparison, cyberbullying, and the pressure to meet unreal standards can impact an individual's pride, social connections, and overall well-being. We participated in the shared task at DepSign-LT-EDI@RANLP 2023 and have proposed a model that identifies the levels of depression in social media text using the data set shared for the task. Different transformer models, like ALBERT and RoBERTa, are used by the proposed model for implementing the task. The macro F1 scores obtained by the ALBERT model and the RoBERTa model are 0.258 and 0.143, respectively.

1 Introduction

Social media has transformed the way we combine, correspond, and share facts. While it has led to numerous benefits, there is an increasing concern regarding its negative effect on mental health, specifically when it comes to depression (Jones et al., 2022). The loyal uncovering of carefully curated lives, social comparison, and cyberbullying are just instances of how social media can cause feelings of depression. One of the negative effects of social media is the phenomenon of social comparison. Platforms like Instagram and Facebook frequently present idealized interpretations of crowd's lives, stressing their realizations, travels, and happy moments. This never ending risk to seemingly perfect lives can lead human beings to compare themselves negatively, developing feelings of failure, envy, and depression (Winstone et al., 2023). Cyberbullying is another important concern on social media podiums. The obscurity and distance given by these platforms can encourage human beings to undertake harmful behaviors, like spreading rumors, making cruel comments, or posting offensive content (Roy et al., 2022). Such experiences can lead to increased social isolation, reduced pride, and depression. To address the negative impact of social media on mental health, it is owned by adopting healthy habits and practices. Firstly, limiting social media use can help humble exposure and counter excessive comparison or rumination. Engaging in offline activities, spending time accompanying loved ones, and the following amusement can determine a much-needed break from the in-essence globe. Cultivating a healthful online atmosphere is important. This includes being aware of the content we consume and share, encouraging positiveness and support, and vigorously combating cyberbullying. Building a forceful support network online and offline can supply more emotional support and neutralize the negative effects of social media. While social media has transformed communication, it is crucially expected to be aware of its potential negative effects on mental health, particularly depression. Depression is considered as one of the most severe mental health diseases, as it often leads to suicide. Hence identifying and summarizing existing evidence concerning depression from data provided by users on social media has become important (Salas-Zárate et al., 2022). The shared task on Detecting Signs of Depression from Social Media Text (Sampath et al., 2023) was a part of RANLP 2023 which is based on English comments.

The task of detecting signs of depression from Social Media Text is a multi-class classification problem, in which the model has to predict the la-

LT-EDI 2023 – Third Workshop on Language Technology for Equality, Diversity and Inclusion, pages 198–203, Varna, Bulgaria, Sep 7, 2023.
https://doi.org/10.26615/978-954-452-084-7_030

bel associated with the text as severe, moderate, or not depression. For example, the text ”I didn’t deserve all of this: I have been suffering from depression for 5 years following personal traumas, I am surviving, in certain situations, I put on an infinite sadness, the panic disorder truncates all my attempts at recourse” represents severe depression. The text, “Any advice? : So... I don’t know where to start and even if I should post this here is moderate and insecurities, fuck em. : I constantly feel like anyone I talk to at all, or act like myself around is just trying to get me to shut up.” represents a not depressed case

2 Related Works

A gold standard data set had been developed by Kayalvizhi and Thenmozhi (2022) to classify the text based on the levels of depression. An empirical analysis using different traditional machine learning algorithms had been presented. The problem of data imbalance had been overcome using data augmentation. The model with Word2Vec victories and Random Forest classifier on augmented data had outperformed the other models.

Salas-Zárate et al. (2022) summarized different works on detecting depression from social media posts. It had been identified that Twitter was the most studied social media for depression sign detection, and Word embedding was the most prominent linguistic feature extraction method. Support vector machine (SVM) was the most used machine-learning algorithm.

Long-Short Term Memory (LSTM) model with two hidden layers and large bias together with Recurrent Neural Network (RNN) with two dense layers had been used by Amanat et al. (2022) to predict depression from text and had provided better results.

Different transformer models like DistilBERT, RoBERTa and ALBERT had been used by Sivamanikandan et al. (2022) to classify social media posts based on the severity of depression associated with them.

The detection of mental illness including depression had been implemented as a multi-class classification problem by Ameer et al. (2022). The use of traditional machine learning, deep learning, and transfer learning-based methods had been explored and the pre-trained RoBERTa transfer learning model resulted in better outcomes.

The strengths of the sequence model and Transformer model had been consolidated by Zhang et al. (2022). The model used a robustly optimized BERT approach to map words into word embedding space and a bidirectional Long Short-Term Memory model to capture the long-distance contextual semantics.

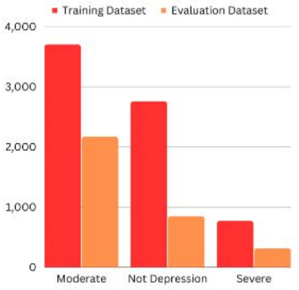

Figure 1: Data Distribution

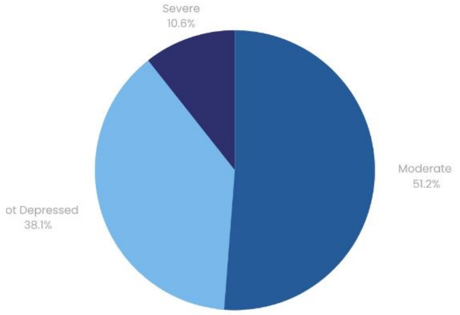

Figure 2: Training Dataset Statistics

3 Dataset Description

The data set that is used to implement the depression detection was the training, evaluation and test dataset that was provided by the organizers of the shared task. Each instance of the training dataset had a label specifying whether the text is moderate, severe, or not depression. The previous version of the task (S et al., 2022) had used a dataset in which the social media text were classified as one of the same three categories.

The data distribution of the training and development dataset for the task is shown in Table 1. The training dataset of the Task had 7201 instances of which 3700 instances were under the moderate category and 2755 instances were under the not depression category and 768 instances were under the severe category. The development dataset of the same task had 2169, 848, and 316 instances under the moderate, not depression, and severe categories respectively. This is also represented by Figure 1. This shows the unbalanced nature of the data set.

Category	Training dataset	Evaluation dataset
Moderate	3700	2169
Not depression	2755	848
Severe	768	316

Table 1: Dataset Statics

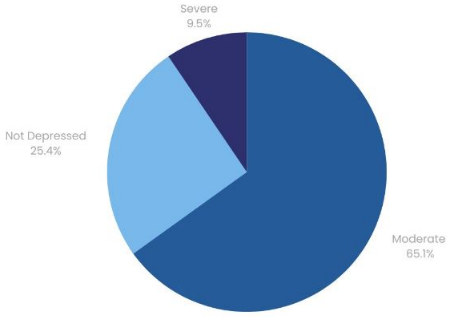

Figure 3: Evaluation Dataset Statistics

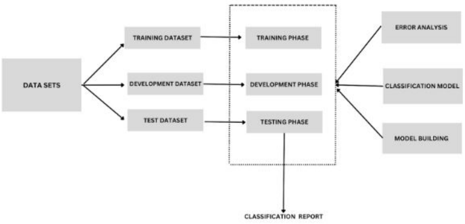

Figure 4: Proposed Architecture

The test data had 499 instances for which the predictions had to be done using the proposed model. The data distribution of different classes of data in the training data is represented in Figure 2 and the development dataset is represented in Figure 3.

4 System Description

Initially, the three datasets provided by the task organizers, namely the training dataset, development dataset, and testing dataset were collected. The training dataset is preprocessed where unnecessary digits, characters, and white spaces are removed using tokenization and it is followed by an encoding process. Then the model is created. In this system, two pre-trained transformer models were used namely ALBERT and RoBERTa. The preprocessed dataset along with the model created is used for the training phase. Each model is then evaluated using the development dataset. The ALBERT model that provided the highest accuracy is taken as the final run for submission and was used to find the predictions for the testing dataset.

The proposed architecture is represented in Figure 4. The removal of unnecessary information is taken care of by the preprocessing phase. All three datasets namely the training, evaluation, and test dataset are preprocessed. This is followed by the process of model building, where trained models namely ALBERT and RoBERTa were used. In the training phase, the pre-trained models are trained using the preprocessed training dataset. The evaluation of the trained model is carried out in the evaluation phase using the evaluation dataset which makes use of the accuracy as the parameter of evaluation. Fine-tuning of hyper-parameters is performed to improve the accuracy of the proposed system. The labels for the text in the dataset are predicted during the testing phase. Contextual embeddings are generated and are used during the training of the model.

4.1 ALBERT

ALBERT (Lan et al., 2019) is a powerful transformer-based language model introduced by Lan et al. as a more efficient and scalable alternative to BERT. ALBERT follows a similar pre-training approach as BERT but introduces parameter-sharing techniques to reduce the model's size and computational requirements. ALBERT has the ability to achieve impressive performance while significantly reducing the number of parameters. By employing parameter sharing and factorization techniques, ALBERT achieves parameter reduction of up to 89%, making it more lightweight and computationally efficient compared to BERT.

The ”Albert-base-v2” model consists of 12 transformer layers, 768 hidden units, and 12 attention heads. This model retains the expressive power and linguistic understanding of larger models like BERT while being more efficient to train and deploy. ALBERT's reduced parameter size not only makes it more computation friendly but also enables faster training and inference times.

4.2 RoBERTa

RoBERTa (Liu et al., 2019) is a transformer model pre-trained on a large corpus of English data and is

Model	F1-Score	Accuracy
ALBERT	0.258	0.421
ROBERTA	0.143	0.263

Table 2: Performance Score

based on the BERT model and modifies key hyperparameters and training is implemented with larger mini-batches and learning rates. RoBERTa is a Robust BERT method that has been trained on a far extra large data set and for a whole lot of large quantities of iterations with a bigger batch length of 8k.

The "RoBERTa–base" model was also used for the task which is a pre-trained model on the English language using a masked language modeling (MLM) objective. This model is case-sensitive and it comprises 12 layers, 768-hidden layers, 12-heads, and 125M parameters.

5 Results

The metrics that were considered for the evaluation of the task was the macro-F1 score and Accuracy. The F1 score is an overall measure of a model's accuracy that merges precision and recall. An extreme F1 score represents that the classification has happened accompanying the reduced number of false positives and low false negatives. The values of the performance metrics particularly the F1 score and accuracy acquired for various models are shown in Table 2. It could be found that the ALBERT model outperformed the other model. The tasks were evaluated based on the macro F1 score acquired by the proposed model. The proposed model resulted in a macro F1 score of 0.258 based on which the task was evaluated. The accuracy acquired was 0.421 and have obtained the 29th rank on the leaderboard. The values for different metrics associated with the ALBERT and ROBERTA models are represented by Figure 6 and Figure 5 respectively.

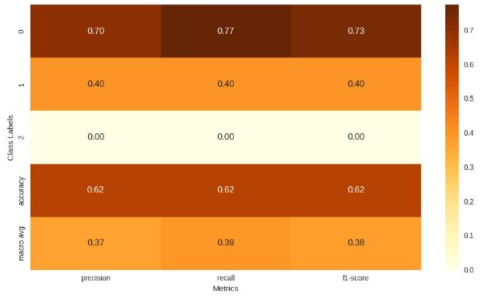

Figure 5: Classification Report - RoBERTa Model

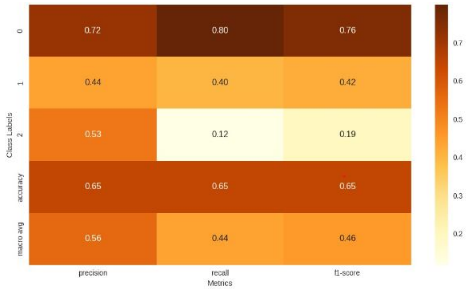

Figure 6: Classification Report - ALBERT Model

6 Error Analysis

The F1 score obtained for the Task using the proposed ALBERT model shows that more false positive and false negative classification has occurred. One reason for this could be considered as the data imbalance nature of the dataset. Considering the number of instances for the class labeled severe is higher, and the F1 score, precision, and recall associated with this class are high when compared to the class, not depression. This represents that the number of misclassifications increases when the number of instances for training is lower, which is associated with data imbalance. Data augmentation could be considered to improve the model's performance. Examples of texts that are misclassified are shown in Table 3. Considering the first text of the table, it has a specific depression marker "kill myself" and is classified as moderate instead of the correct label of severe. The second text of the table, does not have any specific depression marker and is classified as severe instead of the correct label of moderate. The table also shows texts that are sarcastic which are misclassified. All the example texts show that depression markers and sarcasm play a major role in the process of classification and identifying whether the text is associated with depression.

7 Conclusions

Depression detection has become an important area of research as it is interlinked with different application areas. Having this in mind RANLP 2023 had come up with the task of depression detection where the text is classified into moderate, severe, and not depression. The exploration of detecting signs of depression from social media text using the ALBERT and RoBERTa models at LT-EDI@RANLP 2023 demonstrates the significance of leveraging advanced natural language process-

S.No.	Text	Predicted Label	Actual Label
1	I hate that people don't understand that i don't want to kill myself, I just don't want to be alive anymore	Moderate	Severe
2	But here I am, 24 years old man and doing exactly that	Severe	Moderate
3	I'm trapped inside. Does anyone else get that feeling? My memories from the past few years are shoddy at best. I think I'm losing it	Severe	Moderate

Table 3: Examples of Wrong Predictions

ing techniques for mental health analysis. Through the utilization of these models, we were able to classify social media text into three categories: severe, moderate, and not depression. This multi-class classification approach provides a more comprehensive understanding of individuals' mental states and allows for targeted interventions and support. The results obtained from the ALBERT and RoBERTa models contribute to the growing body of research in depression detection from social media. The successful application of these models highlights their effectiveness in capturing subtle linguistic cues and contextual information that indicate depressive symptoms.

Future enhancement to this work can be associated with handling contextual information which can help in effectively detecting depression. The usage of hybrid approaches where different deep learning models are combined can also facilitate the efficient detection of depression from the text.

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.

Iqra Ameer, Muhammad Arif, Grigori Sidorov, Helena Gòmez-Adorno, and Alexander Gelbukh. 2022. Mental illness classification on social media texts using deep learning and transfer learning. *arXiv preprint arXiv:2207.01012*.

Amelia Jones, Megan Hook, Purnaja Podduturi, Haley McKeen, Emily Beitzell, and Miriam Liss. 2022. Mindfulness as a mediator in the relationship between social media engagement and depression in young adults. *Personality and individual differences*, 185:111284.

S Kayalvizhi and D Thenmozhi. 2022. Data set creation and empirical analysis for detecting signs of depression from social media postings. *arXiv preprint arXiv:2202.03047*.

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*.

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*.

Pradeep Kumar Roy, Snehaan Bhawal, and Chinnaudayar Navaneethakrishnan Subalalitha. 2022. Hate speech and offensive language detection in dravidian languages using deep ensemble framework. *Computer Speech & Language*, 75:101386.

Kayalvizhi S, 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.

Rafael Salas-Zárate, Giner Alor-Hernández, María del Pilar Salas-Zárate, Mario Andrés Paredes-Valverde, Maritza Bustos-López, and José Luis Sánchez-Cervantes. 2022. Detecting depression signs on social media: a systematic literature review. In *Healthcare*, volume 10, page 291. MDPI.

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.

S Sivamanikandan, V Santhosh, N Sanjaykumar, Thenmozhi Durairaj, et al. 2022. scubemsec@ lt-edi-acl2022: detection of depression using transformer models. In *Proceedings of the second workshop on language technology for equality, diversity and inclusion*, pages 212–217.

Lizzy Winstone, Becky Mars, Claire MA Haworth, and Judi Kidger. 2023. Types of social media use and digital stress in early adolescence. *The Journal of Early Adolescence*, 43(3):294–319.

Yazhou Zhang, Yu He, Lu Rong, and Yijie Ding. 2022. A hybrid model for depression detection with transformer and bi-directional long short-term memory. In *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 2727–2734. IEEE.