Abstract

Depression is a common mental illness that involves sadness and lack of interest in all day-to-day activities. It is important to detect depression at an early stage as it is treated at an early stage to avoid consequences. In this paper, we present our system submission of ARGUABLY for DepSign-LT-EDI@ACL-2022. We aim to detect the signs of depression of a person from their social media postings wherein people share their feelings and emotions. The proposed system is an ensembled voting model with fine-tuned BERT, RoBERTa, and XLNet. Given social media postings in English, the submitted system classify the signs of depression into three labels, namely “not depressed,” “moderately depressed,” and “severely depressed.” Our best model is ranked 3\textsuperscript{rd} position with 0.54\% accuracy. We make our codebase accessible here\footnote{https://github.com/Muskaan-Singh/Depression-Detection.git}.

1 Introduction

Depression is a common mental illness that involves sadness and lack of interest in all day-to-day activities\footnote{http://ghdx.healthdata.org/gbd-results-tool?params=gbd-api-2019-permalink/d780dffbe8a381b25e1416884959e88b}. Detecting depression is essential as it has to be observed and treated at an early stage to avoid severe consequences (Evans-Lacko et al., 2018; Losada et al., 2017). Depression implies mental disorder which may cause disability (Organization et al., 2012; Whiteford et al., 2015; Vigo et al., 2016), very few people are able to receive treatment (Wang et al., 2007). It is far more difficult for the people with low socioeconomic status or people living in low economic conditions (Steele et al., 2007; Ormel et al., 2008), even adjusting for disorder severity (Mojtabai and Olfsen, 2010; Andrade et al., 2014). Consequently, there is a need to detect these signs of depression early in time to avoid further repercussions. In this work, we detect the signs of depression, namely in “not depressed,” “moderately depressed,” and “severely depressed” from person’s social media postings where people share their feelings and emotions.

There are dataset available for detecting depression task from social media platform such as Twitter (Leis et al., 2019; Arora and Arora, 2019; Yazdavar et al., 2020; de Jesús Titla-Tlatelpa et al., 2021; Chiong et al., 2021; Safa et al., 2021), Reddit (de Jesús Titla-Tlatelpa et al., 2021; Rissola et al., 2019; Tadesse et al., 2019; Burdisso et al., 2019; Martínez-Castaño et al., 2020), Facebook (Chiong et al., 2021; Wongkoblap et al., 2019; Wu et al., 2020; Yang et al., 2020), Instagram (Mann et al., 2020; Ricard et al., 2018), Weibo (Li et al., 2018; Yu et al., 2021) and NHANES, K-NHANES (Oh et al., 2019). The linguistic feature extraction methods used for detecting depression signs on social media such as Word embedding (Mandelbaum and Shalev, 2016), N-grams (Cavnar et al., 1994), Tokenization (Webster and Kit, 1992), Bag of words (Zhang et al., 2010; Aho and Ullman, 1972), Stemming (Jivani et al., 2011), Emotion analysis (Leis et al., 2019; Shen et al., 2017; Chen et al., 2018), Part-of-Speech (POS) tagging (Chiong et al., 2021; Wu et al., 2020), Behavior features(Wu et al., 2020) and Sentiment polarity (Leis et al., 2019; Rissola et al., 2019).

2 Related Work

There have been several attempts to use machine learning algorithms as SVM (Rissola et al., 2019; Arora and Arora, 2019; Burdisso et al., 2019; Yang et al., 2020), Logistic regression (Rissola et al., 2019; Chen et al., 2018; Tadesse et al., 2019; Yang et al., 2019), Neural networks (Wu et al., 2020; Liu et al., 2019), Random forests (Yang et al., 2020; Chiong et al., 2021), Bayesian statistics (Yang et al., 2020; Chien et al., 2018), Decision trees (Yang et al., 2020).
Table 1: Data distribution for the DepSign-LT-EDI dataset.

<table>
<thead>
<tr>
<th>Label</th>
<th>Train</th>
<th>Dev</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not depressed</td>
<td>3801</td>
<td>1830</td>
<td>5631</td>
</tr>
<tr>
<td>Moderately depressed</td>
<td>8325</td>
<td>2306</td>
<td>10631</td>
</tr>
<tr>
<td>Severely depressed</td>
<td>1261</td>
<td>360</td>
<td>1621</td>
</tr>
</tbody>
</table>

Table 2: Examples for Not depressed, Moderately depressed and severely depressed DepSign EDI dataset.

<table>
<thead>
<tr>
<th>Comment</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>Happy New Years Everyone : We made it another year</td>
<td>not depressed</td>
</tr>
<tr>
<td>Sat in the dark and cried myself going into the new year. Great start to 2020 :</td>
<td>moderately depressed</td>
</tr>
<tr>
<td>Words can’t describe how bad I feel right now : I just want to fall asleep forever.</td>
<td>severely depressed</td>
</tr>
</tbody>
</table>

4 Methodology

Firstly, we pre-process the social media tweets with the basic NLT library (Loper and Bird, 2002) for stop words removal, emojis removal, and punctuation removal. Secondly, we extract the features by tokenizing all the sentences and mapping those tokens with the word IDs. For every sentence in the dataset, we follow a series of steps (i) tokenize the sentences (ii) prepend the [CLS] token to the start (iii) append the [SEP] token to the end (iv) map the token to their IDs (v) pad or truncate the sentences to max length (vi) mapping of attention masks for [PAD] tokens. We padded and truncated the max_length=30. The generated sequence sentences are passed for encoding with its attention mask (simply differentiating padding from non-padding). Finally, we predicted the labels using ensembles voting model for BERT (Devlin et al., 2018), XLNET (Liu et al., 2019) and RoBERTa model (Liu et al., 2019). BERT is Bi-directional Encoder Representation from Transformers (BERT), involving pre-training Bi-directional transformers for language understanding from an unlabelled text by jointly conditioning left t-right context for all layers. Fine-tuning of a pre-trained BERT model can be easily done with just one additional output layer for developing a state-of-art model for a wide range of NLP tasks without substantial task-specific architecture modifications. Robustly Optimized BERT approach has emphasized data being used for pre-training and the number of passes for training. The BERT model is optimized with dynamic masking, more extended training with big batches over more data, removing the next prediction objective, and dynamically changing masking patterns for training data. The model achieved
Figure 1: We predicted the labels using fine-tuned BERT, XLNET, and RoBERTa models, respectively, then we applied an ensemble voting classifier. Each model gives a label to the sentence, highest vote is chosen as the final label.

state-of-art results on GLUE, RACE, and SQuAD without multi-task finetuning for GLUE or additional data for SQuAD. ERNIE 2.0 is another continual pre-training framework that efficiently supports customized training tasks in multi-task learning incrementally. The pre-trained model is fine-tuned to adapt to various language understanding tasks. The framework has demonstrated significant improvement over BERT and XLNET on approximately 16 tasks, including GLUE. We take each label to the sentence and number of labels with the highest vote if chosen as the final label in ensemble voting (Dimitriadou et al., 2001).

4.0.1 Experimental Setup
We use V1 100 GPU with 53GB RAM alongside 8 CPU cores for the experimental setup. We divide the entire dataset into a 90:10 training and validation split of 8 batches, with a learning rate (1e-5) and Adam optimizer (Kingma and Ba, 2014) with epsilon (1e-8). We feed a seed_val of 42. For calculating the training loss over all the batches, we use gradient descents (Andrychowicz et al., 2016) with clipping the norm to 1.0 to avoid exploding gradient problem.

5 Results
We evaluate our model quantitatively and qualitatively for the DepSign-LT-EDI dataset. The classification report for our proposed model with average and best submission among all the teams is reported in Table 3. The proposed model has shown progressive results with the 3rd position on the leaderboard https://competitions.codalab.org/competitions/36393#learn_the_details-result. Analysing our quantitative results, 0.53, 0.57, 0.54 are the reported precision, Recall, and F1-score, which is relatively 0.06, 0.07, 0.06 more than the average and 0.05, 0.02, 0.04 less for best-performing submission, respectively. Qualitative analysis of the predicted labels by the proposed methodology can be seen in Table 4. The first, third, and fifth comments were not depressed, moderately depressed, and severely depressed. They are correctly classified instances indicating our model has efficiently identified the phrases with a negative sentiment, such as "depressed," "anxious," "I
Table 3: Classification system’s performance measured in terms of macro averaged Precision, macro averaged Recall and macro averaged F-Score across all the classes. Sklearn classification report was utilized to generate the reports by all the submission teams.

<table>
<thead>
<tr>
<th></th>
<th>Accuracy</th>
<th>Recall</th>
<th>Precision</th>
<th>Weighted F1-score</th>
<th>Macro F1-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average of all teams</td>
<td>0.5988</td>
<td>0.5058</td>
<td>0.4782</td>
<td>0.6012</td>
<td>0.4821</td>
</tr>
<tr>
<td>Best of all teams</td>
<td>0.6709</td>
<td>0.5912</td>
<td>0.586</td>
<td>0.666</td>
<td>0.583</td>
</tr>
<tr>
<td>Our submission</td>
<td>0.6253</td>
<td>0.572</td>
<td>0.5303</td>
<td>0.6333</td>
<td>0.5467</td>
</tr>
</tbody>
</table>

Text data | Label
---|---
Sometimes people can be either too oblivious or choose not to care and they may not intend to harm us but it does hurt. [removed] | not depressed
TMS : My doctor wants me to do TMS for my depression. Has anyone done TMS or is doing it? I was just want to know it is worth it. | not depressed
Depressed : I have nothing to look forward to, I wake up feeling so down and depressed , anxious about everything. I look at myself in the mirror and i feel and look so ugly . I shouldn’t be allowed out in public being so disgusting looking... | moderate
Uncertain : I would like to die, but I’m scared of the repercussions. More specifically, I have to attend a birthday party and a gathering to say goodbye to a friend who will be moving in the next few days and I don’t want to ruin their celebrations | moderate
my whole life has fallen apart : everyone hates me. all my friends hate me. my monster hates me and my dads too busy for me. i don’t talk to my family. the only person i have is my boyfriend who will probably leave me soon because of how i am. i eat lunch in the bathroom. no one in my classes talks to me. i got my boyfriend and his friend accidentally suspended for an incident they jokingly started that ended in me almost getting beat up (they meant no harm). i cried all day and i had to leave school early. i can’t eat. my head is pounding. there’s no hope. there’s no point in living and no one cares. everyone just hates me. and i’m not a bad or mean person i don’t think, but now that’s all i am to everyone. i want to end it, but if i fail i get readmitted to the psych ward and i promised myself if i ever went back there, i would kill myself. i don’t know what to do anymore. | severe
Antidepressants : Do antidepressants help if your not depressed? I started taking them to get through a rough patch and they have helped me - does this mean I technically have depression because I read online that antidepressants don’t help if your not depressed? | severe

Table 4: Qualitative Results for not depression, moderate, severe

would kill myself,“ and so on. Since the first comment barely had any negative phrases, the model classified it as not depressed. However, in the case of the second instance, the comment is labeled as not depressed when in reality, it is a case of severe depression. The probable reason for this misclassification is that the model cannot identify medical terms like "TMS," and overall, the second comment barely has any negative words or expressions.

The fourth instance is labeled as moderate; however, the person claims that they want to die; this indicates that this comment is instead a case of severe depression. The probable reason for this misclassification is that the model focuses more on the phrases like "party," "celebration," "die." rather than the entire sentences. Since this statement has a mix of positive and negative phrases, the model assumes it to be a moderate case. Lastly, the sixth instance is classified as severe; it seems like the case of mild depression.

6 Conclusion

In this paper we present our system paper submission for DepSign-LT-EDI@ACL-2022. We aim to detect the signs of depression of a person from their social media postings wherein people share their feelings and emotions. The proposed system is an ensembled voting model with fine-tuned BERT, RoBERTa, and XLNet. Given social media postings in English, the submitted system classify the signs of depression into three labels, namely “not depressed,” “moderately depressed,” and “severely depressed.” Our best model is ranked 3rd position with 0.54% accuracy. The system performs quite well to recognize the comments for depression comments; In the future, we intend to work on a multi-task learning framework to handle all kinds of depression or illness and even the severity of depression. We also aim to detect multilingual depression speech in the code-mixing scenarios.

Acknowledgements

This work was supported by the European Union’s Horizon 2020 research and innovation program under grant agreement No. 833635 (project ROX-ANNE: Real-time network, text, and speaker analytics for combating organized crime, 2019-2022).
References


Akkapon Wongkoblap, Miguel A Vadillo, and Vasa Curcin. 2019. Predicting social network users with depression from simulated temporal data. In IEEE EUROCON 2019-18th International Conference on Smart Technologies, pages 1–6. IEEE.


