SSN_MLRG1@LT-EDI-ACL2022: Multi-Class Classification using BERT models for Detecting Depression Signs from Social Media Text

Karun Anantharaman, S. Rajalakshmi, S. Angel Deborah,
M. Saritha, R. Sakaya Milton
Department of Computer Science and Engineering
Sri Sivasubramaniya Nadar College of Engineering
Chennai 603 110, Tamil Nadu, India
karun19049@cse.ssn.edu.in,
{rajalakshmis, angeldeborahs}@ssn.edu.in,
{sarithamadhesh, miltonrs}@ssn.edu.in

Abstract

DepSign-LT-EDI@ACL-2022 aims to ascertain the signs of depression of a person from their messages and posts on social media wherein people share their feelings and emotions. Given social media postings in English, the system should classify the signs of depression into three labels namely "not depressed", "moderately depressed", and "severely depressed". To achieve this objective, we have adopted a fine-tuned BERT model. This solution from team SSN_MLRG1 achieves 58.5% accuracy on the DepSign-LT-EDI@ACL-2022 test set.

1 Introduction

Depression is a frequently found mental illness that involves sadness and lack of interest in all day-to-day activities. It is vital to detect and treat depression at an early stage to avoid consequences. Treatment involves diagnosis of patient who might have depression, but patient would have have to initiate contact in order to receive this opportunity.

It has been proven by multiple studies that depression is preventable and early stage detection and the most severe effect of this disease can be mitigated by quick treatment. However, openly accessible tools to this end are very few and very rare. The rise of social media as one of humanity's most important public communication platforms presents a potential prospect for early identification and management of mental illness (Priyadharshini et al., 2021; Kumaresan et al., 2021).

People's daily lives are increasingly dominated by social media (Chakravarthi, 2020; Chakravarthi and Muralidaran, 2021). On social media, a lot of multimedia content, mostly brief words and photographs, is constantly exchanged (Chakravarthi et al., 2021, 2020). Information put on the Internet, as opposed to conventional human contact, may be swiftly disseminated by acquaintances and accessed by strangers (Sampath et al., 2022; Ravikiran et al., 2022; Chakravarthi et al., 2022; Bharathi et al., 2022; Priyadharshini et al., 2022). This method allows users to avoid direct interaction with individuals while also increasing their urge to convey their emotions.

This task 4 in Second Workshop on Language Technology for Equality, Diversity, Inclusion (LT-EDI) aims to detect depression from english text (Durairaj et al., 2022). This research article evinces how a BERT Transformer Model can effectively classify social media texts into 3 classes "not depressed", "moderately depressed", and "severely depressed".

The model is trained on social media texts from various sources, labelled as above. The process involves 2 subtasks - PreProcessing and Training. In Subtask-A, the text is cleaned up, and converted to a format more suitable for context and sentiment analysis for depression detection. In Subtask-B a simple transformer BERT classifcation model is trained on the

Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion, pages 296 - 300
May 27, 2022 ©2022 Association for Computational Linguistics

task data, and the performance is evaluated.

2 Background

2.1 Definitions

The section contains descriptions of the models made use of, and related terminology

Transformers - Every output element is related to every input element, and the weightings between them are dynamically determined depending on their relationship. (In NLP, this is referred to as attention.)

BERT - BERT is based on Transformers and stands for Bidirectional Encoder Representations from Transformers. Earlier models could only read input text linearly for a long time, either from right to left or from left to right; they couldn't do both at the same time. In this way, BERT differs from previous models in that it is designed to read in both directions at the same time. Bidirectionality is a feature that was made possible with the introduction of Transformers.

2.2 Related Work

Depression Detection

Models for detecting depression must be extremely precise and quick in order for early intervention to be feasible. (Shen et al., 2017) advocated the extraction of six feature groups, which were then used to train a multi-modal depression dictionary learning model to detect depressed Twitter users. (Burdisso et al., 2019) presented the SS3 text classification system for early depression diagnosis in social media streams that is easy and effective. (Lin et al., 2020) proposed SenseMood, a system that employs a BERT classifier and a CNN to categorise depressed/not-depressed social media messages and photographs. (**?**) asserts that existing depression detection assessments are ineffective at quantifying model delay, and proposes a remedy to this problem.

BERT

In the field of natural language processing, BERT models are widely used. To further understand how such models function, (van Aken et al., 2019) gives A Layer-Wise Analysis of Transformer Representations. (Devlin et al., 2018) demonstrates how pre-trained models may be utilised to interpret natural language. A overview of BERT-based models for text-based emotion recognition may be found in (Acheampong et al., 2021). An early departing modification of BERT for quicker inference is shown in (Xin et al., 2020).

Our earlier research work in contextual emotion and sentiment analysis uses ensemble techniques and Gaussian process models in (Angel Deborah et al., 2019), (Angel Deborah et al., 2021), (Rajalakshmi et al., 2018), (Rajendram et al., 2017b), (Rajendram et al., 2022) and (Rajendram et al., 2017a) forms the base for depression detection. We have used transformer models and its variants to detect offense and humor in text (Sivanaiah et al., 2020), (Sivanaiah et al., 2021) and (Nanda et al., 2021).

2.3 Data

The task data set contains social media texts in English. The data set contains 3 columns, the pid, the social media text in English, and the label as "not depressed", "moderately depressed", and "severely depressed". The test, development and train data sets all have data pertaining to these 3 classes.

The training set has a total of 8891 entries, of which 1971 are labelled "not depressed", 6019 are labelled "moderately depressed", and the remaining 901 are "severely depressed".

The development set has a total of 4496 entries which are split as 1830 "not depressed", 2306 "moderately depressed" and 360 "severely depressed". The test set has 3245 data points.

3 System Overview

The first step in the system flow is preprocessing the data. The aim is to remove any unnecessary elements from the text, and transform the data given into a more uniform form. This involves the following steps:

(i) Extend Contractions - A contraction is an abbreviated version of a word, such as don't, which stands for do not, and aren't, which stands for are not. In order for the model to perform better, we need to broaden this contraction in the text data.

(ii) Lower Case - Because lower case and upper case are interpreted differently by the machine, it is easier for a machine to read the words if the text is in the same case.
(iii) Remove Punctuations - Another text processing approach is punctuation removal. There are 32 punctuation marks that need to be eliminated in total. We may use a regular expression and the string module to replace any punctuation in text with an empty string.
(iv) Remove words and numbers that contain digits - Sometimes words and digits are written together in the text, which is difficult for machines to grasp. As a result, we must exclude terms that are a mix of words and numerals, such as game57 or game5ts7. Because this sort of term is difficult to handle, it's best to remove it or replace it with a NULL string.
(v) Remove Stopwords - Stopwords are the most frequently occurring words in a text that offer no useful information. Stopwords include words like them, they, who, this, and there.
(vi) Stemming and Lemmatization - Stemming is the process of reducing a word to its root stem, such as run, running, runs, and runed, which are all derived from the same word. Words like ing, s, and es, for example, are stemmed to eliminate prefixes and suffixes. The words are stemmed using the NLTK package.
(vii) Remove White Spaces - We need to control this problem since most text data has additional spaces or more than one space is left between the text while completing the preceding preparation processes.
(viii) Data Augmentation - This technique is used to create synthetic data to take care of the imbalance in the dataset.

The next part of the system is the BERT classification model. The pre-trained BERT model from simpletransformer API has been used in this model. The BERT model is fine tuned on the processed data, to give a 3-class classification model capable of effectively classifying new data encountered into various classes. The working of BERT model is shown in Figure 1. It has two phases as pre-training and fine-tuning.

4 Experimental Setup

The data is imported as a pandas dataframe. This dataframe is first passed to a function to expand all contractions, this is done with a precollected dictionary of contractions. Next, the sentence is converted to lower case, and punctuation's are removed using a regex compiled expression. The data at this point is parsed for english stopwords, a list of which are obtained from the Natural Language Tool Kit (NLTK) in python. These stopwords are removed. Stemming and Lemmatization are also done using the python NTLK. White spaces are removed using a regex expression. NLPAUG library is used for data augmentation to balance the data between the 3 classes since the data is imbalanced across the labels.

A BERT model is trained on the above processed data, multiple parameters have been tested using WANDB Sweeps, and the highest scoring configuration has been used. A learning rate of 1e-4 and a batch size of 16 were used.

5 Results

The efficacy of this model has been proven by the results given below:

Metric	Score
Accuracy	0.585
Macro F1-Score	0.412
Macro Recall	0.403
Macro Precision	0.436
Weighted F1-Score	0.576
Weighted Recall	0.585
Weighted Precision	0.572

Table 1: Results

We have obtained an accuracy of 59% and the top rank team has achieved 66% accuracy. Further improvement in the system can be achieved by tweaking the hyper parameters.

6 Conclusion

In summation, our research work presents a BERT model for classification of social media texts into the 3 target classes. The current model does not perform very well on the given data. One reason that can be attributed to this is the complexity of different texts, with various parts involving various sentiments. Future models, will aim to remedy this through splitting the sentences based on their complexity, and using different models for different levels of complexity. The other reason for the low results may be the different manifestations of depression symptoms in different people, this will be remedied by using various other features along with social media texts. In the future, a classifier to segregate texts based on complexity and the number of sentiment expressions may be supplied to further improve the efficiency of the classifier.

References

Francisca Adoma Acheampong, Henry Nunoo-Mensah, and Wenyu Chen. 2021. Transformer models for text-based emotion detection: a review of bert-based approaches. *Artificial Intelligence Review*, 54(8):5789–5829.

S Angel Deborah, TT Mirnalinee, and S Milton Rajendram. 2021. Emotion analysis on text using multiple kernel gaussian... *Neural Processing Letters*, 53(2):1187–1203.

S Angel Deborah, S Rajalakshmi, S Milton Rajendram, and TT Mirnalinee. 2019. Contextual emotion detection in text using ensemble learning. In *Emerging Trends in Computing and Expert Technology. COMET 2019. Lecture Notes on Data Engineering and Communications Technologies, vol 35.*, pages 1179–1186. Springer, Cham.

B Bharathi, Bharathi Raja Chakravarthi, Subalalitha Chinnaudayar Navaneethakrishnan, N Sripriya, Arunaggiri Pandian, and Swetha Valli. 2022. Findings of the shared task on Speech Recognition for Vulnerable Individuals in Tamil. In *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion*. Association for Computational Linguistics.

Sergio G. Burdisso, Marcelo Errecalde, and Manuel Montes y Gómez. 2019. A text classification framework for simple and effective early depression detection over social media streams. *Expert Systems with Applications*, 133:182–197.

Bharathi Raja Chakravarthi. 2020. HopeEDI: A multilingual hope speech detection dataset for equality, diversity, and inclusion. In *Proceedings of the Third Workshop on Computational Modeling of People's Opinions, Personality, and Emotion's in Social Media*, pages 41–53, Barcelona, Spain (Online). Association for Computational Linguistics.

Bharathi Raja Chakravarthi and Vigneshwaran Muralidaran. 2021. Findings of the shared task on hope speech detection for equality, diversity, and inclusion. In *Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 61–72, Kyiv. Association for Computational Linguistics.

Bharathi Raja Chakravarthi, Vigneshwaran Muralidaran, Ruba Priyadharshini, and John Philip McCrae. 2020. Corpus creation for sentiment analysis in code-mixed Tamil-English text. In *Proceedings of the 1st Joint Workshop on Spoken Language Technologies for Under-resourced languages (SLTU) and Collaboration and Computing for Under-Resourced Languages (CCURL)*, pages 202–210, Marseille, France. European Language Resources association.

Bharathi Raja Chakravarthi, Ruba Priyadharshini, Thenmozhi Durairaj, John Phillip McCrae, Paul Buitaleer, Prasanna Kumar Kumaresan, and Rahul Ponnusamy. 2022. Findings of the shared task on Homophobia Transphobia Detection in Social Media Comments. In *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion*. Association for Computational Linguistics.

Bharathi Raja Chakravarthi, Ruba Priyadharshini, Rahul Ponnusamy, Prasanna Kumar Kumaresan, Kayalvizhi Sampath, Durairaj Thenmozhi, Sathiyaraj Thangasamy, Rajendran Nallathambi, and John Phillip McCrae. 2021. Dataset for identification of homophobia and transophobia in multilingual YouTube comments. *arXiv preprint arXiv:2109.00227*.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR*, abs/1810.04805.

Thenmozhi Durairaj, Bharathi Raja Chakravarthi, Jerin Mahibha C, and Kayalvizhi Sampath. 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"*. Association for Computational Linguistics.

Prasanna Kumar Kumaresan, Ratnasingam Sakuntharaj, Sajeetha Thavareesan, Subalalitha Navaneethakrishnan, Anand Kumar Madasamy, Bharathi Raja Chakravarthi, and John P McCrae. 2021. Findings of shared task on offensive language identification in Tamil and Malayalam. In *Forum for Information Retrieval Evaluation*, pages 16–18.

Chenhao Lin, Pengwei Hu, Hui Su, Shaochun Li, Jing Mei, Jie Zhou, and Henry Leung. 2020. *SenseMood: Depression Detection on Social Media*, page 407–411. Association for Computing Machinery, New York, NY, USA.

Ayush Nanda, Abrit Pal Singh, Aviansh Gupta, Rajalakshmi Sivanaiah, Angel Deborah Suseelan, S Milton Rajendram, and Mirnalinee TT. 2021. Techssn at haha@ iberlef 2021: Humor detection and funniness score prediction using deep learning.

Ruba Priyadharshini, Bharathi Raja Chakravarthi, Subalalitha Chinnaudayar Navaneethakrishnan, Thenmozhi Durairaj, Malliga Subramanian, Kogilavani Shanmugavadivel, Siddhanth U Hegde, and Prasanna Kumar Kumaresan. 2022. Findings of the shared task on Abusive Comment Detection in Tamil. In *Proceedings of the Second Workshop on Speech and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Ruba Priyadharshini, Bharathi Raja Chakravarthi, Sajeetha Thavareesan, Dhivya Chinnappa, Durairaj Thenmozhi, and Rahul Ponnusamy. 2021. Overview of the DravidianCodeMix 2021 shared task on sentiment detection in Tamil, Malayalam, and Kannada. In *Forum for Information Retrieval Evaluation*, pages 4–6.

S Rajalakshmi, S Milton Rajendram, TT Mirnalinee, et al. 2018. Ssn mlrg1 at semeval-2018 task 1: Emotion and sentiment intensity detection using rule based feature selection. In *Proceedings of the 12th International Workshop on Semantic Evaluation*, pages 324–328.

S Milton Rajendram, TT Mirnalinee, et al. 2017a. Ssn_mlrg1 at semeval-2017 task 4: sentiment analysis in twitter using multi-kernel gaussian process classifier. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 709–712.

S Milton Rajendram, TT Mirnalinee, et al. 2017b. Ssn_mlrg1 at semeval-2017 task 5: fine-grained sentiment analysis using multiple kernel gaussian process regression model. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 823–826.

S Milton Rajendram, Mirnalinee TT, et al. 2022. Contextual emotion detection on text using gaussian process and tree based classifiers. *Intelligent Data Analysis*, 26(1):119–132.

Manikandan Ravikiran, Bharathi Raja Chakravarthi, Anand Kumar Madasamy, Sangeetha Sivanesan, Ratnavel Rajalakshmi, Sajeetha Thavareesan, Rahul Ponnusamy, and Shankar Mahadevan. 2022. Findings of the shared task on Offensive Span Identification in code-mixed Tamil-English comments. In *Proceedings of the Second Workshop on Speech and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Anbukkarasi Sampath, Thenmozhi Durairaj, Bharathi Raja Chakravarthi, Ruba Priyadharshini, Subalalitha Chinnaudayar Navaneethakrishnan, Kogilavani Shanmugavadivel, Sajeetha Thavareesan, Sathiyaraj Thangasamy, Parameswari Krishnamurthy, Adeep Hande, Sean Benhur, Kishor Kumar Ponnusamy, and Santhiya Pandiyan. 2022. Findings of the shared task on Emotion Analysis in Tamil. In *Proceedings of the Second Workshop on Speech and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Guangyao Shen, Jia Jia, Liqiang Nie, Fuli Feng, Cunjun Zhang, Tianrui Hu, Tat-Seng Chua, and Wenwu Zhu. 2017. Depression detection via harvesting social media: A multimodal dictionary learning solution. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, pages 3838–3844.

Rajalakshmi Sivanaiah, S Milton Rajendram, Mirnalinee Tt, Abrit Pal Singh, Aviansh Gupta, Ayush Nanda, et al. 2021. Techssn at semeval-2021 task 7: Humor and offense detection and classification using colbert embeddings. In *Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021)*, pages 1185–1189.

Rajalakshmi Sivanaiah, Angel Suseelan, S Milton Rajendram, and Mirnalinee Tt. 2020. Techssn at semeval-2020 task 12: Offensive language detection using bert embeddings. In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, pages 2190–2196.

Betty van Aken, Benjamin Winter, Alexander Löser, and Felix A. Gers. 2019. How does bert answer questions? a layer-wise analysis of transformer representations. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, CIKM '19, page 1823–1832, New York, NY, USA. Association for Computing Machinery.

Ji Xin, Raphael Tang, Jaejun Lee, Yaoliang Yu, and Jimmy Lin. 2020. Deebert: Dynamic early exiting for accelerating BERT inference. *CoRR*, abs/2004.12993.