MUCS@Text-LT-EDI@ACL 2022: Detecting Sign of Depression from Social Media Text using Supervised Learning Approach

Asha Hegde[1 a], **Sharal Coelho**[1 b],
Ahmad Elyas Dashti[1 c], **Hosahalli Lakshmaiah Shashirekha**[1 d]
[1]Department of Computer Science, Mangalore University, Mangalore, India
{[a]hegdekasha,[b]sharalmucs,[c]elyas.dashti808,[d]hlsrekha}@gmail.com

Abstract

Social media has seen enormous growth in its users recently and knowingly or unknowingly the behavior of a person will be reflected in the comments she/he posts on social media. Users having the sign of depression may post negative or disturbing content seeking the attention of other users. Hence, social media data can be analysed to check whether the users' have the sign of depression and help them to get through the situation if required. However, as analyzing the increasing amount of social media data manually in laborious and error-prone, automated tools have to be developed for the same. To address the issue of detecting the sign of depression content on social media, in this paper, we - team MUCS, describe an Ensemble of Machine Learning (ML) models and a Transfer Learning (TL) model submitted to "Detecting Signs of Depression from Social Media Text-LT-EDI@ACL 2022" (DepSign-LT-EDI@ACL-2022) shared task at Association for Computational Linguistics (ACL) 2022. Both frequency and text based features are used to train an Ensemble model and Bidirectional Encoder Representations from Transformers (BERT) fine-tuned with raw text is used to train the TL model. Among the two models, the TL model performed better with a macro averaged F-score of 0.479 and placed 18th rank in the shared task. The code to reproduce the proposed models is available in github page[1].

1 Introduction

A person feeling unimportant, useless, or unhappy may be a sign of depression. Depression is one of the most severe mental health conditions which may be unnoticed, undiagnosed and untreated in many cases. People worldwide suffer from depression and the affected person may operate poorly at work, studies, and in the community. Recent research studies have shown that the popularity of social media networks in one's life is increasing day by day. People are using social media to share their thoughts, feelings, emotions and sentiments (Islam et al., 2018). Knowingly or unknowingly the behavior of a person will be reflected in the comments she/he posts on social media (Sampath et al., 2022a; Ravikiran et al., 2022; Chakravarthi et al., 2022; Bharathi et al., 2022; Priyadharshini et al., 2022). Social media users who are usually in depression try to seek the attention and sympathy of others by posting negative and disturbing messages or requesting help. Some have even reached to the extent of going live on social media before taking drastic steps such as suicide (Chakravarthi, 2020; Chakravarthi et al., 2021; Chakravarthi and Muralidaran, 2021). Due to all these issues, understanding mental health on social media has become a popular field of study (Alhuzali et al., 2021).

Studies have indicated that the analysis of the messages posted on social media platforms by the users can help to predict the sign of depression (Chiong et al., 2021) of the users and the early prediction can help the users to get through the situation. Researchers are exploring to analyze social media content to predict the mental health of users in order to lend a helping hand to the needy at the earliest. In this paper, we - team MUCS, describe the models submitted to DepSign-LT-EDI@ACL 2022[2] shared task to detect signs of depression in social media text and classify them into into three categories: "not depressed", "moderately depressed", and "severely depressed". Two models: i) An ensemble of ML classifiers, namely: Random Forest (RF), Multinomial Naive Bayes (MNB), Multi-Layer Perceptron (MLP), and Gradient Boosting (GB) with soft voting ii) TL model with BERT, are proposed to classify the given input into one of the three predefined categories. The rest of the article is structured as follows: A review of relevant work is included in Section 2, and the

[1]https://github.com/hegdekasha/Detecting-sign-of-depression

[2]https://competitions.codalab.org/competitions/36410

Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion, pages 312 - 316

methodology is discussed in Section 3. Experiments, results, and error analysis are described in Section 4 followed by concluding the paper with future work in Section 5.

2 Literature Review

Researchers have experimented various methodologies to build systems capable of detecting the signs of depression in social media content and a few of the relevant ones are described below:

To analyze suicide ideation symptoms on Reddit social media, Tadesse et al. (2020) developed a combined Long Short Term Memory (LSTM) - Convolutional Neural Network (CNN) model based on Word2Vec features and obtained 93.8% accuracy. Haque et al. (2021) implemented the Boruta algorithm in association with RF classifier to predict depression in kids and teenagers aged from 4 to 17. Their proposed model was evaluated on Youth Minds Matter (YMM) dataset and their model predicted the depressed classes with 95% accuracy. To identify the signs of depression in Twitter, K S et al. (2019) used Word2Vec word embeddings to represent the Tweets and train the combination of the LSTM and CNN model and Support Vector Machine (SVM). The LSTM and CNN model combination and SVM model obtained an overall weighted avg F1-scores of 0.97 and 0.85 respectively.

Zygadło et al. (2021) employed Naive Bayes, SVM, and BERT for sentiment and emotion recognition in English and Polish texts. They built CORTEX[3] - a Polish version of the dataset for sentiment and emotion recognition. BERT-based classifier achieved accuracies of over 90% and around 80% for sentiment and emotion classification respectively. Hämäläinen et al. (2021) have created a dataset for detecting depression in Thai blog posts and tested it with four different models: (i) Bidirectional LSTM (BiLSTM) based model using Open-Source Neural Machine Translation (OpenNMT)[4] toolkit (ii) LSTM model with Word2Vec features, (iii) Thai BERT [5] model, and (iv) Multilingual BERT model. Among these models, the Thai BERT model achieved the highest overall accuracy of 77.53%.

Even though several techniques have been developed to detect the sign of depression in social media text, there are no full-fledged models for all datasets. Further, the trend in posting comments on social media changes frequently because of creative users. Hence, this necessitates the need for the development of new models to detect the sign of depression in a social media text.

3 Methodology

The proposed methodology includes two distinct models namely: i) Ensemble of ML classifiers and ii) TL model with BERT, for detecting the sign of depression in social media text. Description of the two models are given below:

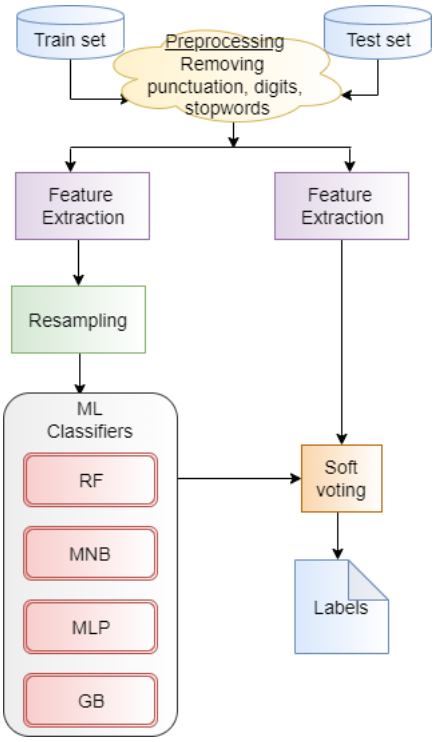

Figure 1: The proposed framework of Ensemble of ML classifiers

3.1 Ensemble of Machine Learning Classifiers

The proposed Ensemble of ML classifiers consists of Pre-processing, Feature Extraction and Model Building steps and the framework of the proposed model is shown in Figure 1. Each of the steps are explained below:

Pre-processing - Dataset is pre-processed to remove punctuation, digits, and stopwords, as they do not contribute to the classification task. The English stopwords list available at Natural Language Tool Kit (NLTK) library[6] is used to remove stop-

[3]https://github.com/azygadlo/CORTEX
[4]https://github.com/OpenNMT/OpenNMT-py
[5]https://github.com/ThAIKeras/bert
[6]https://www.nltk.org

words and Porter Stemmer[7] is used to reduce the words to their stems.

Feature Extraction - As the given dataset is imbalanced, resampling is carried out using randomoversampling[8] technique to bring balance in the dataset. Frequency based features, namely: TF-IDF of character bigrams and trigrams and word unigrams and text based features, namely: count of words and characters followed by the count of adjectives, adverbs, nouns, and pronouns are extracted. These features are combined and used to train the Ensemble of ML classifiers. The number of character bigrams and trigrams extracted amounts to 9,024 and word unigrams amounts to 13,169.

Model Building - ML classifiers are generally ensembled by making use of the strength of one classifier to overcome the weakness of another classifier to improve the results. RF, MLP, MNB, and GB classifiers are ensembled to detect the sign of depression in social media text and soft voting is used to predict the category of the Test set.

The RF algorithm consists of a set of decision trees, each of which is trained with a random subset of features, and the prediction is carried out based on majority voting of all the trees in the forest (Islam et al., 2019). The MLP classifier is widely used in classification as they are simple and easy to implement. It is a feed-forward neural network which consists of three layers, namely: input layer, an output layer, and one or more hidden layers (Lakhotia and Bresson, 2018). The MNB model is a popular ML classifier because of its computing efficiency and relatively good predictive performance (Harjule et al., 2020). GB classifier will benefit the regularization methods that penalize different parts of the algorithm and improve the overall performance by reducing overfitting (Stein et al., 2019).

3.2 Transfer Learning model with BERT

BERT is a popular language representation model used to train TL model for text classification. It is pre-trained on Wikipedia corpus with 2,500 million words of unlabelled text and 800 million words from huggingface Book Corpus. Further, it is a bidirectional model which learns information from both left and right sides of the context.

BERT accepts raw text for fine-tuning the pre-trained embeddings. The model provides positional encoding based BERT tokenizer followed by BERT embeddings which transforms each token into tensors so that the classifiers can be trained using these tensors. The framework of the proposed TL model is shown in Figure 2.

Classes	Train set	Dev set
moderate	6,019	2,306
not depression	1,971	1,830
severe	901	360

Table 1: Class-wise distribution of the dataset

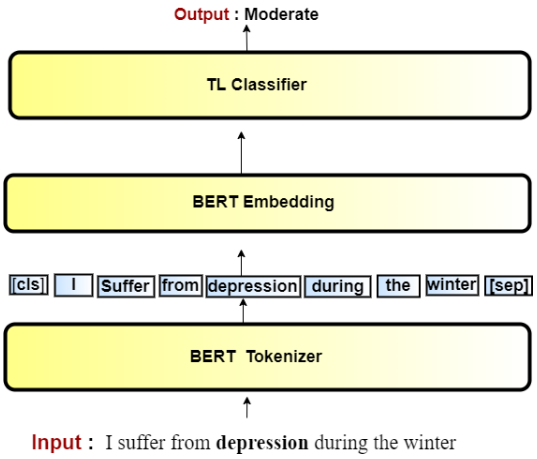

Figure 2: The framework of TL model with BERT

4 Experiments and Results

Several experiments were conducted with different resampling techniques and various combinations of features and classifiers and the models that gave good performance on the Development set are applied for the Test set. The dataset provided by the organisers to detect the sign of depression consists of social media comments in English (Sampath et al., 2022b). Table 1 gives the class-wise distribution of the dataset. For TL model, the pre-trained BERT-base-uncased[9] model is used with ClassificationModel[10] - a transformer based classifier, to predict the labels for the given Test set. Table 2 shows the hyperparameters and the values of the hyperparameters used to implement TL model.

The proposed models were evaluated by the organizers of the shared task based on macro averaged F-score and the results are shown in Table 3. Ensemble of ML classifiers model achieved macro averaged F-scores of 0.573 and 0.419 for Develop-

[7]https://www.nltk.org/_modules/nltk/stem/porter.html
[8]https://imbalanced-learn.org/
[9]https://huggingface.co/docs/transformers/model_doc/bert
[10]https://simpletransformers.ai/docs/classification-models/

Hyperparameters	Value
Layers	12
Hidden size	768
Self attention heads	12
110 M trainable parameters	

Table 2: The values of the hyperparameters used in TL model

Models	Dev set	Test set
Ensemble based model	0.573	0.419
TL based model	0.620	0.479

Table 3: Performance of macro averaged F-score of the proposed models

ment (Dev) set and Test set respectively. Further, the TL model outperformed the other model with macro averaged F-scores of 0.620 and 0.479 for Dev set and Test set respectively. In spite of resampling the data using random over sampling to balance the dataset, the results are still low. This may be because the random over sampling technique duplicates features from the minority classes resulting in overfitting for some models (Yap et al., 2014).

5 Conclusion and Future work

This paper describes the models submitted by our team - MUCS to DepSign-LT-EDI@ACL-2022 shared task to detect signs of depression from social media text in English. The two proposed models are: i) Ensemble of ML classifiers trained with the combination of frequency and text based features and ii) TL model with BERT. Resampling is also explored to handle the data imbalance problem. The TL model outperformed Ensemble model with a macro averaged F-score of 0.479 securing 18th rank in the shared task. Future research will explore different sets of features and feature selection algorithms for detecting sign of depression from social media text.

References

Hassan Alhuzali, Tianlin Zhang, and Sophia Ananiadou. 2021. Predicting Sign of Depression via Using Frozen Pre-trained Models and Random Forest Classifier. In *CLEF (Working Notes)*.

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.

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, 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*.

Raymond Chiong, Gregorius Satia Budhi, Sandeep Dhakal, and Fabian Chiong. 2021. A Textual-based Featuring Approach for Depression Detection using Machine Learning Classifiers and Social Media Texts. volume 135, page 104499. Elsevier.

Mika Hämäläinen, Pattama Patpong, Khalid Alnajjar, Niko Partanen, and Jack Rueter. 2021. Detecting Depression in Thai Blog Posts: a Dataset and a Baseline.

Umme Marzia Haque, Enamul Kabir, and Rasheda Khanam. 2021. Detection of Child Depression using Machine Learning Methods. volume 16, page e0261131. Public Library of Science San Francisco, CA USA.

Priyanka Harjule, Astha Gurjar, Harshita Seth, and Priya Thakur. 2020. Text Classification on Twitter Data. In *2020 3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things (ICETCE)*, pages 160–164.

Md Islam, Muhammad Ashad Kabir, Ashir Ahmed, Abu Raihan M Kamal, Hua Wang, Anwaar Ulhaq, et al. 2018. Depression Detection from Social Network

Data using Machine Learning Techniques. volume 6, pages 1–12. Springer.

Md Zahidul Islam, Jixue Liu, Jiuyong Li, Lin Liu, and Wei Kang. 2019. A Semantics Aware Random Forest for Text Classification. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, pages 1061–1070.

Aswathy K S, Rafeeque C, and Reena Murali. 2019. Deep Learning Approach for the Detection of Depression in Twitter.

Suyash Lakhotia and Xavier Bresson. 2018. An Experimental Comparison of Text Classification Techniques. In *2018 International Conference on Cyberworlds (CW)*, pages 58–65. IEEE.

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.

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. 2022a. 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.

Kayalvizhi Sampath, Thenmozhi Durairaj, Bharathi Raja Chakravarthi, and Jerin Mahibha C. 2022b. 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.

Roger Alan Stein, Patricia A Jaques, and Joao Francisco Valiati. 2019. An Analysis of Hierarchical Text Classification using Word Embeddings. volume 471, pages 216–232. Elsevier.

Michael Mesfin Tadesse, Hongfei Lin, Bo Xu, and Liang Yang. 2020. Detection of Suicide Ideation in Social Media Forums Using Deep Learning. volume 13, page 7. Multidisciplinary Digital Publishing Institute.

Bee Wah Yap, Khatijahhusna Abd Rani, Hezlin Aryani Abd Rahman, Simon Fong, Zuraida Khairudin, and Nik Nik Abdullah. 2014. An Application of Oversampling, Undersampling, Bagging and Boosting in Handling Imbalanced Datasets. In *Proceedings of the first international conference on advanced data and information engineering (DaEng-2013)*, pages 13–22. Springer.

Artur Zygadło, Marek Kozłowski, and Artur Janicki. 2021. Text-Based Emotion Recognition in English and Polish for Therapeutic Chatbot. volume 11, page 10146. Multidisciplinary Digital Publishing Institute.