TechSSN1@LT-EDI: Depression Detection and Classification using BERT Model for Social Media Texts

Venkatasai Ojus Yenumulapalli, Vijai Aravindh R, Rajalakshmi S, Angel Deborah S
Department of Computer Science and Engineering,
Sri Sivasubramaniya Nadar College of Engineering, Chennai - 603110, Tamil Nadu, India
venkatasai2110272@ssn.edu.in, vijaiaravindh2110281@ssn.edu.in,
rajalakshmis@ssn.edu.in, angeldeborahs@ssn.edu.in

Abstract

Depression is a severe mental health disorder characterized by persistent feelings of sadness and anxiety, a decline in cognitive functioning resulting in drastic changes in a human's psychological and physical well-being. However, depression is curable completely when treated at a suitable time and treatment resulting in the rejuvenation of an individual. The objective of this paper is to devise a technique for detecting signs of depression from English social media comments as well as classifying them based on their intensity into severe, moderate, and not depressed categories. The paper illustrates three approaches that are developed when working toward the problem. Of these approaches, the BERT model proved to be the most suitable model with an F1 macro score of 0.407, which gave us the 11^{th} rank overall.

1 Introduction

Depression has emerged as a significant mental health issue in recent years, affecting millions of individuals worldwide. Simultaneously, the use of social media platforms has skyrocketed, becoming an integral part of people's daily lives. Beck and Alford (2009) talks about the clinical causes of depression and the various treatments for it. Social media platforms provide individuals with an outlet to express their emotions and share personal experiences. Hammen (2005) examines the relationship between depression and stress over time including effects of childhood and lifetime stress exposure.

People often turn to these platforms as a means of seeking support, and validation, or simply as an avenue to connect with others who may be going through similar struggles. De Choudhury et al. (2013) conducted an analysis on various factors such as social engagement, emotion, language, and linguistic styles, as well as mentions of antidepressant medication. The purpose of this analysis was to develop a statistical classifier that can estimate the likelihood of experiencing depression.

The primary objective of the DepSign-LTEDI task (Sampath et al., 2023) is to identify indications of depression in individuals based on their social media posts, and analyze English-language social media content and classify the signs of depression into three categories, moderate, severe, and not depressed.

2 Related Work

Kayalvizhi and Thenmozhi (2022) developed a gold standard dataset in order to detect the various levels of depression namely moderate, severe, and not depressed using social media texts. Data augmentation techniques were applied to overcome data imbalance. Word2Vec vectorizer and Random Forest classifier model were used which provided better accuracy.

Salas-Zárate et al. (2022) employed Twitter as the primary data repository for depression sign detection. Word embeddings were used as the well-known technique for linguistic extraction. The machine-learning approach utilised was the support vector machine (SVM), and cross-validation (CV) was used to evaluate the outcomes.

Wang et al. (2022) principally used three approaches to meet the objective. The first method makes use of sentence embeddings for which VADER scores are generated following which they are passed into the gradient boosting model along with SMOTE augmentation techniques to combat data imbalance. The second technique centred on optimising pre-trained models using a multi-layer perceptron. The final technique used the multi-layer perceptron and VAD embeddings to classify the symptoms of depression.

Bucur and Dinu (2020) focused on the detection of the early onset of depression through the anal-

LT-EDI 2023 – Third Workshop on Language Technology for Equality, Diversity and Inclusion,
pages 149–154, Varna, Bulgaria, Sep 7, 2023.
https://doi.org/10.26615/978-954-452-084-7_022

ysis of social media posts with special attention on Reddit. Topic modeling embeddings were extracted using a Latent Semantic Indexing Model and then provided as input to the neural network design. The neural network had two outputs - classifying whether the individual was depressed or not and estimating the confidence of the individual.

Trifan et al. (2020) used Reddit posts between January and October of 2016 as the source of the data. The data was pre-processed by being converted to lowercase and tokenized after any extraneous characters were eliminated. Multinomial Naive Bayes, Support Vector Machine in conjunction with Stochastic Gradient Descent, and Passive Aggressive classifiers were the final three classifiers used.

Rajalakshmi et al. (2018) developed a method to detect intensity levels of emotions with the help of rule-based feature selection using tweets as the primary data source. The input feature vectors were generated using one-hot encoding and rule-based feature selection. The model for the detection of emotional intensity classification was Multilayer Perceptron. The models for the subtasks of sentiment intensity regression and emotion intensity regression was constructed using Support Vector Regression. Anantharaman et al. (2022) and Esackimuthu et al. (2022) used transformer models to detection the depression from social media tweets and achieved F1 score of 0.412 and 0.473 respectively.

3 Dataset

The dataset consists of the posting id, text data, and the corresponding label (S et al., 2022). The dataset provides three labels representing the various degrees of depression namely severe, moderate, and no depression. Table 1 illustrates the distribution of the dataset.

Label	Train	Dev	Test
Not Depressed	2755	848	135
Moderate	3678	2169	275
Severe	768	228	89

Table 1: Distribution of Data

4 Depression Detection System

We have employed three different models in the three test runs , in this respective order:

BERT: We employed the 'bert-base-uncased' pre-trained model, known as BERT (Bidirectional Encoder Representations from Transformers), as the foundation for this depression analysis task. To tailor BERT to the specific given depression dataset, fine-tuning of the BERT model is done through training using the labeled train data. Additionally, to ensure compatibility with the model, a label encoder is utilized to convert the target labels into numerical values. Table 2 gives the evaluation metrics of BERT model for the development dataset.

Word2Vec and SVC: In this approach, Word2Vec, a technique that generates word embeddings - distributed numerical representations of word features is used for text-vector representation. These embeddings capture the contextual meaning of words within vocabulary and enable the model to identify semantic relationships among words that share similar meanings. To leverage the power of these word vectors, a classification model Support Vector Classifier (SVC) is employed to train on and predict using these word embeddings. Table 3 gives the evaluation metrics for the development dataset.

TFIDF-LinearSVC: LinearSVC is a machine learning algorithm implemented in Python's scikit-learn library, based on the Support Vector Machine (SVM) algorithm. LinearSVC aims to find a linear decision boundary that maximizes the margin between different classes. The model learns a set of weights for each feature and combines them linearly to make predictions. To vectorize the given texts, TfidfVectorizer is used, which converts text into numerical feature vectors. Subsequently, LinearSVC model is fitted to these vectors. Table 4 gives the evaluation metrics of this model for the development dataset.

Metrics	Score
Accuracy	0.67
Macro Precision	0.57
Macro Recall	0.54
Macro F1-score	0.55
Weighted precision	0.65
Weighted recall	0.67
Weighted F1-score	0.66

Table 2: Evaluation metrics of BERT

Parameters	Score
Accuracy	0.66
Macro Precision	0.58
Macro Recall	0.44
Macro F1-score	0.47
Weighted precision	0.64
Weighted recall	0.66
Weighted F1-score	0.64

Table 3: Evaluation metrics of Word2Vec-SVC

Parameters	Score
Accuracy	0.60
Macro Precision	0.51
Macro Recall	0.49
Macro F1-score	0.50
Weighted precision	0.63
Weighted recall	0.60
Weighted F1-score	0.61

Table 4: Evaluation Metrics of TFIDF-LinearSVC

5 Methodology

The codes utilized in this research paper are available within the GitHub repository GitHub Repository. This repository contains a collection of Jupyter Notebook files that correspond to the methodologies and techniques detailed in the paper.

The methodology employed in this paper consists of three approaches. The first method makes use of the Bidirectional Encoder Representations from Transformers (BERT) "bert-base-uncased" pre-trained model. The BERT model was tailored to the dataset provided. The dataset was tokenized, and converted into the input features with the assistance of the BERT tokenizer. No pre-processing steps were undertaken in this model. Following this, the input features and labels were converted to tensors, and the TensorDataset was created. The model was trained for three epochs using the optimizer AdamW, which has a learning rate of 2e-5. Finally, cross-validation was performed with the help of the development data for fine tuning the trained model and tested on the test data. Figure 1 shows the work flow of the system.

The second approach involved the usage of Word2Vec and Support Vector Classification (SVC). Word embedding is a method in which words are converted into an arithmetic representation termed a vector and each word in a sentence can be represented through this vector. These vectors capture the semantic relationships between words and can be used to assess the similarity and dissimilarity between words. The Word2Vec API trained on Google News was used to generate the word embeddings for the given data. The pre-processing step removed stop words and punctuation, lemmatized the remaining tokens, and returned the average word vector representation using spaCy's word embeddings. If no valid tokens remain, it returns a default zero vector. The Word2Vec makes use of two architectures namely bag-of-words and skip-gram. However, the word embeddings generated in this model do not strictly follow the aforementioned architectures. The Support Vector Classification is a machine learning classification algorithm that is often used to classify data with multiple labels. The generated word embedding was then fed into the Support Vector Classification model to obtain the classified results. Figure 2 shows the working of this system.

The LinearSVC and TF-IDF Vectorizer are used in the final strategy. Similar to the Word2Vec model, the Term Frequency-Inverse Document Fre-

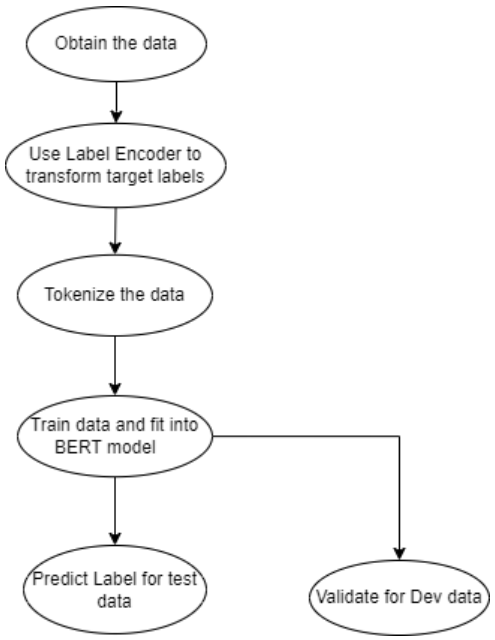

Figure 1: Flowchart for BERT

Parameters	Score
Number of epochs	1
Batch size	16
Learning rate	2e-5
Maximum sequence length	512

Table 5: Hyperparameters: BERT

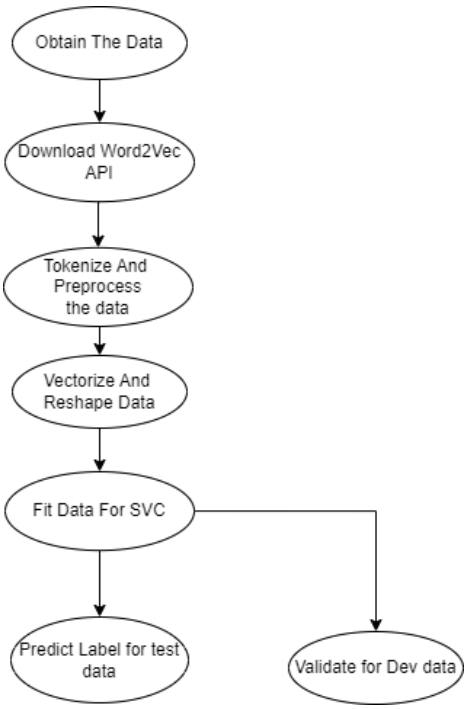

Figure 2: Flowchart for Word2Vec and SVC

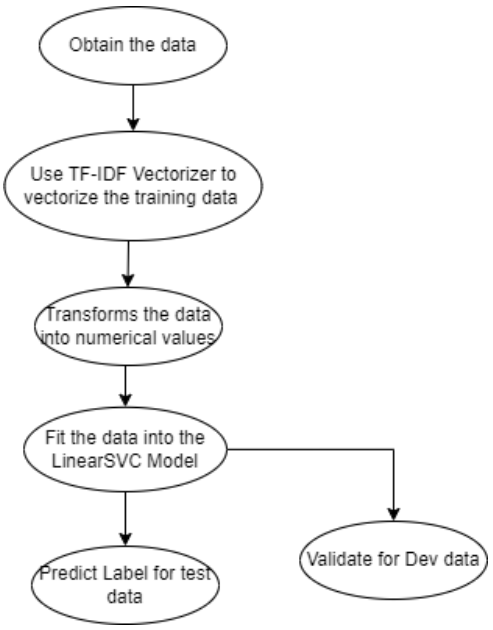

Figure 3: Flowchart for TFIDF Vectorizer and Linear SVC

quency text vectorizer combines the concepts of term frequency (TF) and document frequency (DF). No pre-processing steps were undertaken in this model. The term frequency is an indication of the frequency count of words in a document and gives an idea about the importance of a given word. The quantity of documents that use a particular term is known as the document frequency. LinearSVC is a machine learning algorithm implemented in Python's scikit-learn library, based on the Support Vector Machine (SVM) algorithm that aims to find a linear decision boundary that maximizes the margin between different classes. Figure 3 depicts the steps in developing this model.

6 Results

6.1 Test Data Results

Out of the three above-mentioned models, the BERT model gave the highest F1-score of 0.407 for the predicted labels on the test data the organizers gave, and an accuracy of 55 percent. On further analysis, the following metrics were obtained for the three models. Tables 6, 7 and 8 show the evaluation metrics value for accuracy, precision, recall and F1 score.

It is inferred from the results that SVC could not perform well when compared to the BERT model. This is our maiden attempt for applying the machine learning algorithms for a real life problem in the form of a challenge. We assume that the performance of the BERT model can be increased by training the model for more number of epochs.

Parameters	Score
Accuracy	0.55
Macro Precision	0.54
Macro Recall	0.42
Macro F1-score	0.41
Weighted precision	0.55
Weighted recall	0.55
Weighted F1-score	0.50

Table 6: Evaluation Metrics for Test Data: BERT

7 Conclusion

In conclusion, an analysis was experimented with the provided social media texts using three models: BERT, Word2Vec with SVC, and TF-IDF with LinearSVC. Among these models, BERT demonstrated the highest accuracy in predicting the target labels. It is worth noting that accurately analyzing and predicting signs of depression is challenging for any model, as it may struggle to comprehend the deeper layers of a sentence and grasp the nuanced tone of the text.

Improving the accuracy of the model can be achieved by ensuring a more balanced distribution of the dataset, where the number of cases for all three target labels is nearly equal. This would help

Parameters	Score
Accuracy	0.46
Macro Precision	0.47
Macro Recall	0.39
Macro F1-score	0.37
Weighted precision	0.49
Weighted recall	0.46
Weighted F1-score	0.44

Table 7: Evaluation Metrics for Test Data: LinearSVC and TFIDF

Parameters	Score
Accuracy	0.52
Macro Precision	0.65
Macro Recall	0.38
Macro F1-score	0.35
Weighted precision	0.59
Weighted recall	0.52
Weighted F1-score	0.46

Table 8: Evaluation Metrics for Test Data: Word2Vec and SVC

prevent biases and provide a more comprehensive understanding of different degrees of depression.

Moving forward, our aim is to advance our knowledge in the field of Natural Language Processing (NLP) by exploring various models and implementing them in different use cases to analyze signs of depression from social media texts effectively.In addition,we have planned to work on the various variants of BERT models for better understanding, such as DistilBERT. By undertaking such endeavors, we can enhance our understanding and contribute to the development of more robust and accurate NLP techniques in detecting mental health conditions.

References

Karun Anantharaman, S Rajalakshmi, S Angel Deborah, M Saritha, and R Sakaya Milton. Ssn_mlrg1@ lt-edi-acl2022: Multi-class classification using bert models for detecting depression signs from social media text. *LTEDI 2022*, page 296, 2022.

Aaron T Beck and Brad A Alford. *Depression: Causes and treatment*. University of Pennsylvania Press, 2009.

Ana-Maria Bucur and Liviu P Dinu. Detecting early onset of depression from social media text using learned confidence scores. *arXiv preprint arXiv:2011.01695*, 2020.

Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. Predicting depression via social media. In *Proceedings of the international AAAI conference on web and social media*, volume 7, pages 128–137, 2013.

Sarika Esackimuthu, H Shruthi, Rajalakshmi Sivanaiah, S Angel Deborah, R Sakaya Milton, and TT Mirnalinee. Ssn_mlrg3@ lt-edi-acl2022-depression detection system from social media text using transformer models. *LTEDI 2022*, page 196, 2022.

Constance Hammen. Stress and depression. *Annu. Rev. Clin. Psychol.*, 1:293–319, 2005.

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

S Rajalakshmi, S Milton Rajendram, TT Mirnalinee, et al. 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, 2018.

Kayalvizhi S, Thenmozhi Durairaj, Bharathi Raja Chakravarthi, and Jerin Mahibha C. 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*, Dublin, Ireland, May 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.ltedi-1.51. URL https://aclanthology.org/2022.ltedi-1.51.

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

Kayalvizhi Sampath, Thenmozhi Durairaj, Bharathi Raja Chakravarthi, Jerin Mahibha C, Kogilavani Shanmugavadivel, and Pratik Anil Rahood. 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, 2023. Recent Advances in Natural Language Processing.

Alina Trifan, Rui Antunes, Sérgio Matos, and Jose Luís Oliveira. Understanding depression from psycholinguistic patterns in social media texts. In *Advances in Information Retrieval: 42nd European Conference on IR Research, ECIR 2020, Lisbon, Portugal, April 14–17, 2020, Proceedings, Part II 42*, pages 402–409. Springer, 2020.

Wei-Yao Wang, Yu-Chien Tang, Wei-Wei Du, and Wen-Chih Peng. Nycu_twd@ lt-edi-acl2022: Ensemble

models with vader and contrastive learning for detecting signs of depression from social media. In *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 136–139, 2022.