Team-KEC@LT-EDI2023: Detecting Signs of Depression from Social Media Text

Malliga S
Kongu Engineering College
mallisenthil.cse@kongu.edu

Kogilavani Shanmugavadivel
Kongu Engineering College

Arunaa S
Kongu Engineering College
arunaa.20cse@kongu.edu

Gokulkrishna R
Kongu Engineering College
gokulkrishnar.20cse@kongu.edu

Chandramukhii A
Kongu Engineering College, Erode
Chandramukhiia.20cse@kongu.edu

Abstract

The prevalence of depression has become a pressing concern in modern society, necessitating innovative approaches for early detection and intervention. This study explores the feasibility of leveraging social media text as a potential source for detecting signs of depression. This study utilized different techniques to represent the text data in a numerical format and various techniques such as CNN, BERT, and N-gram to classify social media posts into depression and non-depression categories. Text classification tasks often rely on deep learning techniques such as CNN, while the BERT model, which is pre-trained, has shown exceptional performance in a range of natural language processing tasks. To assess the effectiveness of the suggested approaches, the research employed multiple metrics, including accuracy, precision, recall, and F1-score.Our model bagged the official rank of 12 and gave an F1 score of 0.401. The outcomes of the investigation indicate that the suggested techniques can identify symptoms of depression with an average accuracy rate of 56

Keywords *N-gram, CNN, BERT*

1 Introduction

The pervasive use of social media has introduced new challenges in detecting signs of depression from text shared on these platforms(Greenberg-LS 2017). Researchers in NLP have utilized feature-based linear classifiers, CNN, and RNN architectures, as well as fine-tuning pre-trained language models like BERT and Roberta, to automatically detect signs of depression. While linear classifiers have shown competitive performance, pre-trained models have achieved state-of-the-art results. However, pre-trained models may have limitations in understanding context-specific language. This project provides an overview of existing research, describes the task and dataset, proposes machine learning and deep learning models, presents experimental results, and concludes with potential avenues for future research.Suicide is a serious public health problem; however, suicides are preventable with timely, evidence-based and often low-cost interventions[1].

The following sections of this document are structured as follows: Section 2 provides a comprehensive review of existing research on the identification of signs of depression through analysis of social media text(Wei-Yao-Wang et al. 2017). Section 3 provides a detailed explanation of the task at hand, including a description of the dataset employed in this study. Our proposed machine learning and deep learning models for detecting signs of depression are presented in Section 4. Subsequently, Section 5 outlines the conducted experiments and presents the corresponding results. A thorough discussion of these results is provided within the same section. Finally, in Section 6, we present our concluding remarks based on the findings and suggest potential avenues for future research in this field.

2 Literature Survey

The rapid growth of internet content and the anonymity of online platforms have made it challenging to manually identify signs of depression from social media text (Holleran 2020). However, machine learning and NLP techniques, such as Naive Bayes, Decision Trees, Random Forests, SVM, CNN, and RNN, have shown promise in automatically detecting signs of depression with high accuracy, even when tested on diverse datasets, including those associated with hate speech[2]. These

[1] https://www.who.int/news-room/fact-sheets/detail/suicide

[2] https://ojs.aaai.org/index.php/ICWSM/article/view/14432

LT-EDI 2023 – Third Workshop on Language Technology for Equality, Diversity and Inclusion, pages 97–102, Varna, Bulgaria, Sep 7, 2023.
https://doi.org/10.26615/978-954-452-084-7_014

advancements offer potential for efficient and automated detection and intervention in online contexts.Results obtained from the literature review stated that BiLSTM + Attention model performs well on depression related textual data. Even though the achieved result may be satisfactory, there are certain issues with the model implemented in that research(David-William 2020).For example, Guntuku S.C., Yaden D.B., Kern M.L., Ungar L.H., Eichstaedt J.C. Detecting depression and mental illness on social media(Guntuku-S.C. et al. 2017) focus on studies aimed at predicting mental illness using social media. First, they consider the methods used to predict depression, and then they consider four approaches that have been used in the literature: prediction based on survey responses, prediction based on self-declared mental health status, prediction based on forum membership, and prediction based on annotated posts.Wang Y.P., Gorenstein C. Assessment of depression in medical patients: A systematic review of the utility of the Beck Depression Inventory-II(Wang.Y.P. and Gorenstein.C 2013) examined relevant investigations with the Beck Depression Inventory-II for measuring depression in medical settings to provide guidelines for practicing clinicians. The Beck Depression Inventory-II showed high reliability and good correlation with the measures of depression and anxiety.

3 Materials and Methods

3.1 Dataset Description

The dataset from the Coda Lab Competitions[3] consists of three main sections: Train data, Development data, and Test data with a sample given in Table 1. It includes train text id, train text, and labels indicating the severity of depression (No depression, Moderate, or Severely Depressed). The dataset is in English and comprises social media comments. Prior to applying machine learning and deep learning models, basic pre-processing steps like removing irrelevant characters and normalizing text were performed. The dataset is imbalanced, with varying numbers of instances across depression severity labels. The dataset includes 3678 moderate comments, 2755 not depressed comments, and 768 severely depressed comments. To address this, the SMOTE technique was used to balance class distribution by randomly increasing minority class examples by replicating them.Thus here SMOTE increases minority class (severely depressed) samples and balance the dataset(Jason-Brownlee 2020). By employing SMOTE, potential biases caused by the initial imbalance were alleviated, enhancing the reliability and robustness of the subsequent models.

DOCUMENT	TEXT	LABEL
Document [14]	Happy new year : Fuck 2019... 2020 will be bettexaxaxaxaxaxa why do i even have to be happy because earth did a whole circle around sun nothing will cange fuck my life hope u all have a better year that me......	moderate
Document [637]	What if : What if you couldnt feel bain jelasy hate happiness sadness or anything what if you could live the life of true emotional freedom I people i had choosen the wrong choice...	
Document [2320]	I'm really struggling : So I don't know how to start things like this, So I'll start with basics. I'm 16yo, diagnosed depression at 14yo. Since then, my life is total mess. I've already been to two different psychologists...	severe

Table 1: Sample Training Texts

3.2 Preprocessing and Feature Extraction

To build an effective classifier, preprocessing or corpus cleaning is necessary. In this study, 3-grams were used to tokenize comments and extract features. 3-grams capture contextual relationships between words, allowing for a comprehensive representation of the text data and enabling better classification performance(Vairaprakash-Gurusamy 2014). By converting 3-grams into vectors based on their frequency and context, the resulting feature vectors are useful for text analysis and modeling tasks. Additionally, BERT and CNN

[3] https://codalab.lisn.upsaclay.fr/competitions/11075

were employed as classification models. BERT is a pre-trained language model known for its exceptional performance in various NLP tasks, while CNN is a popular deep learning technique for text classification. The combination of 3-grams, BERT, and CNN enhances the classifier’s ability to identify patterns in the text data(Shizhe-Diao et al. 2020).

3.3 3-Gram Representation : Capturing the Contextual Relationship

3-gram representations provide a different approach to capturing the sequential nature of words in a text. Using the 3-gram technique, contiguous sequences of three words are extracted. For example, the phrase ”consistent tomorrow drastically” represents one 3-gram, while ”tomorrow drastically may” represents another. a few samples are given in [Table 2]. By utilizing 3-gram representations, we obtain a set of sequential word sequences that capture local word order and context information. These 3-gram sequences offer a more granular understanding of the text’s structure and meaning compared to individual words or traditional n-gram representations.

In the context of text classification or language modeling, these 3-gram representations can be used as features. They provide additional contextual information that helps in capturing the nuances and dependencies within the text. These features contribute to more accurate and comprehensive analysis and inference. Overall, the utilization of 3-gram representations enhances the capability of capturing local word relationships, improving the quality of text analysis and enabling more effective processing of sequential data.

3-grams are assigned index values rather than stored as strings. These index values represent the different 3-gram units. The CountVectorizer or FastText model calculates vector representations for each 3-gram based on their frequency[4]. These vector representations capture the contextual information and co-occurrence patterns within the text. The classifiers are then trained using these vector representations to learn patterns and make predictions. Incorporating 3-grams allows the classifiers to capture both individual word features and the contextual relationships between adjacent words, improving their performance in text classification tasks.

DOCUMENT	LABEL	3 GRAM
Document [11]	moderate	[’consistent tomorrow drastically’, ’tomorrow drastically may’, ’drastically may view’, ’may view hopeless’,’addictive disappointed satisfaction’.....]
Document[615]	no depression	[’psychologist psychiatrist physician uncomfortable’, ’physician uncomfortable psychiatrist’, ’uncomfortable psychiatrist medicated’, ’psychologist soon based’....]
Document[641]	severe	[’argue bpd aspergers’, ’bpd aspergers ocd’, ’aspergers ocd suspected’, ’ocd suspected 2018’, ’suspected 2018 22’, ’2018 22 crap’, ’22 crap partner’, ’crap partner seriously’, ’partner seriously alcoholic’......]

Table 2: RESULTS FROM 3 GRAM

[4] https://scikit-learn.org/stable/tutorial/basic/tutorial.html

Here's how 3-Gram is used:

1. Create an 3-Gram object and specify the desired 3-Gram range. 2. Apply the 3-Gram transformation to tokenize the text into 3-Gram. 3. Convert the 3-Gram into vector representations using techniques like Count Vectorizer or TF-IDF. 4. Use the resulting vectors for further analysis or modeling tasks. N-gram vectors capture word sequence frequency, providing contextual information for tasks like text classification and language modeling. The chosen 3-Gram range affects granularity in capturing text structure and meaning. Experimenting with different ranges optimizes performance for specific applications, ensuring originality and improving work quality.

4 Proposed Classifiers

The text-to-feature transformation is described, followed by the classification algorithms used for detecting signs of depression from social media text. For feature extraction, 3-gram techniques were employed, which are extensively used in NLP and text mining tasks.It capture contiguous sequences of three words, providing local word order and context information.Thus when 3-gram is used it allows machine to recognize the 3 words as one entity which makes the text classification to next level. To classify the extracted features, two classifiers are proposed: CNN and BERT. The proposed architecture is shown in [Figure 1].

Convolutional Neural Networks (CNN) are deep learning models that are effective in capturing local patterns in text data. CNNs have shown promising results in various NLP tasks. We used CNN for the 3-gram feature extraction method[5]. The architecture of CNN algorithm includes a number of convolution layers, max-pooling layers, and fully connected layers. ReLU as an activation function is used in proposed work.

This project leverages the power of BERT (Bidirectional Encoder Representations from Transformers), a highly effective pre-trained language model known for capturing contextual information from text[6].

Unlike traditional models, which looked at a text sequence only from one direction, the BERT encoder attention mechanism works bidirectional training of transformer, which learns information from both the left and right sides of a word, allowing the model to catch a deeper sense of language context.

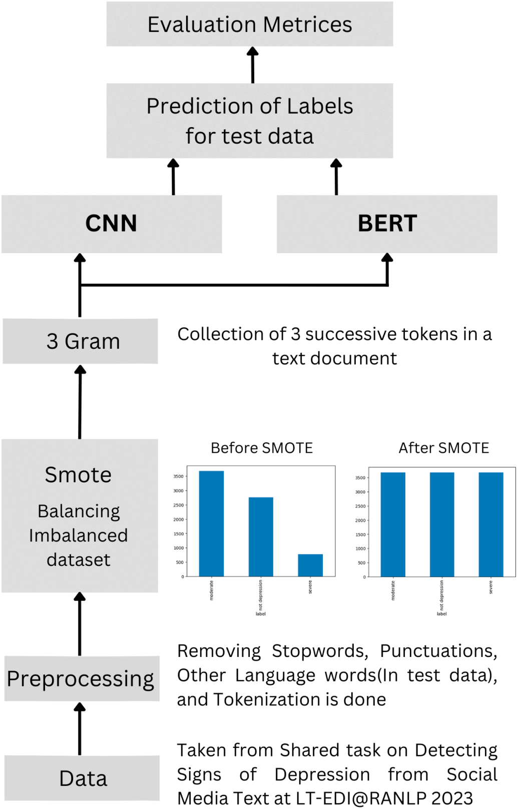
Evaluation Metrices
Prediction of Labels for test data
CNN
BERT
3 Gram
Collection of 3 successive tokens in a text document
Smote
Balancing Imbalanced dataset
Before SMOTE
After SMOTE
3500
3000
2500
2000
1500
1000
500
0
moderate
not depression
severe
label
Preprocessing
Removing Stopwords, Punctuations, Other Language words(In test data), and Tokenization is done
Data
Taken from Shared task on Detecting Signs of Depression from Social Media Text at LT-EDI@RANLP 2023

Figure 1: Proposed Model

We integrated BERT into our 3-gram feature extraction method, allowing us to benefit from its capabilities in various NLP tasks, including text classification. Each of the proposed classifiers takes the respective feature vectors as input and outputs the classification for each social media text. These classifiers have different specialties, and their performance metrics may vary.

By utilizing the 3-gram technique along with CNN and BERT classifiers, we aim to effectively detect signs of depression from social media text, providing valuable insights for mental health analysis.

5 Results and Discussion

The proposed classifiers have been implemented using scikit-learn(F.A.Nazira and M.F.Mridha 2021) and Python, and the training and testing processes took place on the Google Collaboratory platform. Google Collaboratory provides a cloud-based Jupyter notebook environment, eliminating the need for local setup. In our study, the coda lab

[5] https://neptune.ai/blog/vectorization-techniques-in-nlp-guide

[6] https://medium.com/analytics-vidhya/confusion-matrix-accuracy-precision-recall-f1-score-ade299cf63cd

CLASSIFIERS	CLASS LABELS	ACCURACY	PRECISION	RECALL	F1-SCORE
CNN RESULT USING 3 GRAM	moderate	0.55	0.67	0.65	0.66
	not depression	0.45	0.20	0.11	0.15
	severe	0.65	0.18	0.50	0.26
	accuracy			0.47	3246
	macro avg	0.25	0.31	0.25	3246
	weighted avg	0.50	0.47	0.48	3246
BERT RESULT USING 3 GRAM	moderate	0.52	0.66	0.64	0.65
	not depression	0.57	0.20	0.12	0.15
	severe	0.68	0.18	0.49	0.26
	accuracy			0.49	3246
	macro avg	0.26	0.32	0.26	3246
	weighted avg	0.51	0.49	0.49	3246

Table 3: PERFORMANCE OF CLASSIFIERS

LT-EDI@RANLP 2023 dataset was utilized, specifically developed for detecting signs of depression from social media text. This dataset comprises social media messages in English. We trained various classifiers, including CNN and BERT, using the extracted features from the training set. The performance of these classifiers was then evaluated on the test dataset. The combination of scikit-learn, Python, and the LT-EDI@RANLP 2023 dataset allowed us to detect signs of depression from social media text, contributing to the analysis and understanding of mental health indicators in online communication.

5.1 Performance Metrics

The performance evaluation of the classification models involved the calculation of several metrics, including Accuracy, Precision, Recall, and F1-Score(Qamar-un-Nisa 2021). These metrics are defined as follows:

Accuracy: It measures the proportion of texts correctly classified in a specific class, divided by the total number of texts in that class. The formula for Accuracy (Equation 1) is:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

Recall (Sensitivity or True Positive Rate): It represents the number of texts correctly categorized in a certain class, divided by the total number of actual texts in that class. The formula for Recall (Equation 2) is:

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

Precision (Positive Predictive Value): It measures the number of texts accurately categorized as a specific class, divided by the total number of texts categorized as that class. The formula for Precision (Equation 3) is:

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

F1-Score: It is the harmonic average of Precision and Recall, providing a balanced measure of the model's performance. The F1-Score (Equation 4) is calculated as:

$$F1_{score} = \frac{2 * Precision * Recall}{Precision + Recall} \tag{4}$$

These metrics rely on the True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) indices. TP represents the number of texts correctly classified for a particular class, while FP represents the number of texts misclassified in other classes. FN represents the number of texts misclassified in the relevant class, and TN represents the number of texts correctly classified in other classes except the correct class[7]. The results obtained from proposed models are shown in Table 3.

6 Conclusion and Feature Work

In conclusion, this study successfully conducted experimental work to detect signs of depression from social media text using the provided dataset. 3-gram is employed as feature extraction technique to effectively capture textual information. Different classifiers, including CNN, and BERT, were compared and BERT with 3Gram has achieved a

[7] https://developers.google.com/machine-learning/crash-course/classification/true-false-positive-negative

goo accuracy compared to CNN and thus proving its effectiveness in this task(Vandana 2023).

For future work, there are several potential areas of improvement. Exploring alternative numerical or vectorial representations of the text, such as TF-IDF, could potentially enhance classification performance(Kapse et al. 2022). Additionally, investigating new classifiers based on neural networks, which leverage advanced linguistic features, would be valuable. These approaches can contribute to further improving the detection of signs of depression in social media texts, enabling a deeper understanding of mental health indicators in online communication.

Further, the usage of a post encoder, a sentiment-guided Transformer and a supervised severity-aware contrastive learning component may enhance the result and it can lead the classification method to a new level. Unlike the proposed model the account of sentiment and semantic information of data can lead to greater accuracy. The post encoder MentalRoBERTa and SentiLARE can be used to obtain the semantic as well as sentiment hidden features.

Our model for bagged the official rank of 12 and gave an F1 score of 0.401.

References

David-William, Suhartono.D. (2020). "Text-based Depression Detection on Social Media Posts". In: *ScienceDirect*. DOI: 10.1016/j.procs.2021.01.043.

F.A.Nazira S.R.Das, S.A.Shanto and M.F.Mridha (2021). "Depression Detection Using Convolutional Neural Networks". In: *2021 IEEE International Conference on Signal Processing, Information, Communication Systems (SPICSCON), Dhaka, Bangladesh*, pp. 9–13. DOI: 10.1109/SPICSCON54707.2021.9885517..

Greenberg-LS (2017). "Emotion-focused therapy of depression. Per Centered Exp Psychother". In: 16(1), pp. 106–117.

Guntuku-S.C., Yaden-D.B., Kern-M.L., Ungar-L.H., and Eichstaedt J.C (2017). "Detecting depression and mental illness on social media: An integrative review." In: *Proceedings of the 24th International Conference on Machine Learning* 18, pp. 43–49. DOI: 10.1016/j.cobeha.2017.07.005..

Holleran (2020). "The early detection of depression from social networking sites". In.

Jason-Brownlee (2020). "smote-oversampling-for-imbalanced-classification". In: *ScienceDirect*.

Kapse, Prasanna, Garg, and Vijay Kumar (2022). "Advanced Deep Learning Techniques For Depression Detection: A Review". In: DOI: 10.2139/ssrn.4180783.

Qamar-un-Nisa, Rafi-Muhammad (2021). "Towards transfer learning using BERT for early detection of self-harm of social media users". In: *CLEF 2021 – Conference and Labs of the Evaluation Forum, September 21–24*.

Shizhe-Diao, Ruijia-Xu, Hongjin Su, Yilei-Jiang, Yan-Song, and Tong-Zhang (2020). "Taming Pretrained Language Models with N-gram Representations for Low Resource Domain Adaptation". In: *Curr. Sci.*

Vairaprakash-Gurusamy, Subbu-Kannan (2014). "Preprocessing Techniques for Text Mining". In: *Curr. Opin.*

Vandana Nikhil-Marriwala, Deepti-Chaudhary (2023). "A hybrid model for depression detection using deep learning, Measurement: Sensors". In: *ISSN* 25. DOI: 10.1016/j.measen.2022.100587.

Wang.Y.P. and Gorenstein.C (2013). "Assessment of depression in medical patients: A systematic review of the utility of the Beck Depression Inventory-II". In: *ScienceDirect*.

Wei-Yao-Wang, Yu-Chien, Tang-Wei-Wei Du, and Wen-Chih-Peng (2017). "Ensemble Models with VADER and Contrastive Learning for Detecting Signs of Depression from Social Media". In: *ScienceDirect*.