Journal of Engineering Science and Technology
Special Issue on ICIT2022 Vol. 19, No. 1 (2024) 12 - 24

DEPRESSION DETECTION ON SOCIAL NETWORKS WITH NATURAL LANGUAGE PROCESSING

IRIS NG, HARPRITH KAUR RAJINDER SINGH*

Faculty of Data Science and Information Technology, INTI International University, 71800 Nilai, Negeri Sembilan, Malaysia
Corresponding author: harprith.randhawa@newinti.edu.my

Abstract

Depression has become a major concern in public health as it is the leading cause of disease worldwide. Depression is a growing issue due to measures being taken to stop the spread of COVID-19 disease by limiting people's ability to interact directly. At worst, depression can lead to suicide. People with depression also suffer from anxiety disorders and many other symptoms that affect their physical health and social relationships. In addition, current methods for early screening of patients are time-consuming and costly. Therefore, effective early detection of depression is significant to provide adequate medical attention and support in order to save lives before it is too late. It was noticed that people with mental illness tend to reveal their mental condition on social media as a relief. Thus, this study was conducted to develop a web-based dashboard application with an automated model for analyzing and estimating depression risk posts from social media. This paper introduced techniques that help prediction models to recognize the difference between depressive and non-depressive posts. A new technique is proposed to detect depressive symptoms with a custom LIWC dictionary. This application can assist mental health professionals to assess indicators of depression and raise a flag when there are patients who are potentially depressed. Furthermore, patients can use the system to analyze their mental health condition and contact their doctors when necessary. The system can also spread mental health awareness as it can be used by anyone who wants to self-screen their mental health.

Keywords: Data mining, Depression detection, Human, Health, Natural language processing, Neural network, Sentiment analysis, Twitter.

1.Introduction

Depression has become a major concern in public health as it is the leading cause of disease globally and is often characterized by a lack of social connection [1]. According to the World Health Organization (WHO), about 280 million people of all ages suffered from depression in 2019 [2]. Moreover, depression is a growing issue due to measures taken to stop the spread of COVID-19 disease by limiting people's ability to interact directly [3].

Depression can lead to suicidal ideation or suicide attempts if left unaddressed. People with mental disorders are often reluctant to seek professional help, resulting in fewer people receiving appropriate treatment in coping with depression [4]. More than 70% of people in the early stages of depression do not consult a psychologist because they are ashamed of their depression [5]. Even during consultations with mental health professionals, 30 – 50% of people with depression often go unrecognized and not properly treated [6, 7].

Depression is diagnosed through face-to-face sessions with a clinical psychologist, based on his/her judgment on diagnostic criteria in manuals and questionnaires like the Electronic Health Records (EHRs) and the Beck's Depression Scale (BDI) [6]. However, it is difficult to accurately automate the diagnosis because there are limited datasets due to challenges in obtaining data in such format [8]. Furthermore, it is difficult to diagnose depression efficiently because of the complexity of mental illness [4]. Thus, there is a need for an effective approach to improve early screening for depression to mitigate the negative effects of the disorder.

Meanwhile, social media has grown and developed into a popular platform that provides a supportive environment. People are increasingly relying on social media like Twitter to disclose their mental condition as a relief [9]. This allows researchers to examine the potential of social media as a tool to predict depression earlier and identify depressive symptoms manifested in user posts [10].

This project aims to develop a web-based dashboard application with an automated model capable of analysing and recognizing social media posts with depression-related expressions. The sentiments expressed in the tweets will give an idea of the user's underlying emotions and feelings [6]. Therefore, this study will implement sentiment analysis to obtain disease characteristics from tweets posted by users. Finally, the results can serve as a reference for mental health professionals to diagnose clinical depression [11]. It can also assist mental health professionals in effectively managing and monitoring patients with depression through their social media activities. Specifically, whenever a depression risk was detected, a message advised them to take immediate action for a particular patient, such as scheduling an appointment with the patient immediately. Similarly, patients are allowed to use the system for self-screening and contact their doctors if necessary. This system also helps spread mental health awareness among public users.

2.Literature Reviews

A similar research project was completed by Ahmad et al. [12]. They applied deep learning models to identify depression from social media text by training the corpus and classifying tweets as normal or depressed. Instead of using techniques like word n-gram, character n-gram, and GloVe, they proposed Bidirectional Long Short-Term Memory (BiLSTM) to identify depression risk posts efficiently without any exhaustive feature engineering. By training the model for 30 epochs and a batch size of 300, the model achieved a maximum accuracy of 93.5%. BiLSTM achieves the highest performance among other deep learning models, as shown in Table 1.

Table 1. Comparative results of BiLSTM with other deep learning models [12].

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
RNN	0.82	0.83	0.82	0.83
CNN	0.87	0.84	0.86	0.82
LSTM	0.79	0.81	0.77	0.79
GRU	0.84	0.81	0.84	0.83
BiLSTM	0.93	0.89	0.91	0.90

The next similar research project is suicidal profile detection with machine learning completed by Mbarek et al. [13]. They proposed various machine learning classifiers including Bayes Net, Adaboost, Sequential Minimal Optimization (SMO), J48, and Random Forest to detect suicidal profiles on Twitter. They applied different feature extraction techniques such as using frequent words, n-grams, word length, and more to extract linguistic features. They also looked at emotional, emojis, and timeline features to further analyse suicidal profiles. Apart from that, they also analysed account features such as availability of the country, language, profile description, and facial features. The best-performing model was Random Forest. The comparison results are shown in Table 2.

Table 2. Comparative results of machine learning classifiers [13].

Classifier	Precision (%)	Recall (%)	F1-Score (%)
Bayes Net	0.73	0.74	0.73
Adaboost	0.78	0.78	0.78
SMO	0.79	0.79	0.78
J48	0.80	0.81	0.81
Random Forest	0.83	0.83	0.83

3. Methods

3.1. Data collection

For this study, the dataset was collected using Twint, an advanced Twitter scraper that scrapes tweets without using Twitter's API. Two datasets were required, a normal (non-depressed) dataset and a depressed dataset.

3.1.1. Normal dataset (Non-depressed)

The normal dataset was collected in 4 batches of one day each from 16th to 18th December 2021, and 31st December 2021. 3000 tweets are collected per day, only 1000 tweets on the last day. It has 10,000 tweets in total. The dataset was collected from different days to avoid similar topics discussed on the same day, and to augment the normal dataset with topics diversity.

3.1.2. Depressed dataset

A range of keywords were chosen to target depressive tweets, including different hashtags related to depression, such as “depressed,” “depression,” “anxiety,” “bipolar,” “DepressionIsReal,” and more. Depressed candidates were manually selected from the dataset collected using these hashtags. Their profiles were scrutinized to make sure they showed depressive tendencies such as suicidal, loneliness, and self-hatred.

3.2. Data screening

In this step, irrelevant tweets from the dataset are manually removed using different rules listed below. When done, an additional column “target” is added to each dataset. The “target” column in the normal dataset is set to “normal”, while the depressed dataset is set to “depressed”.

3.2.1. Normal dataset (Non-depressed)

- Remove non-English tweets.
- Remove tweets with duplicate topics, such as NFTs, giveaways, and birthday wishes.
- Remove tweets with commercial purposes, typically from business accounts.
- Remove tweets that share songs, which usually contain only the song title and an external link direct to the song.
- Remove tweets with depressive tendencies.

3.2.2. Depressed Dataset

- Remove non-English tweets.
- Remove tweets with motivational purposes.
- Remove tweets without depressive tendencies.

3.3. Data Preprocessing

Data preprocessing prepares quality data for analysis by removing or modifying duplicate, irrelevant, incorrect, and incomplete data. If data is poorly prepared, it will not produce high-quality results and support meaningful decision making. In this study, data preprocessing was done on the tweet column. The processed tweet column was then combined to the dataset so that it has both the original and cleaned tweets. The data preprocessing techniques applied in the dataset are listed below:

- Extract 4 required columns, including datetime, username, tweet, and target.
- Remove duplicate tweets, non-English characters, and stop words.
- Remove punctuation, numbers, and whitespaces.
- Remove links, emails, mentions, hashtags, Unicode, and special characters.
- Convert emoji to text and text to lowercase.
- Expand contractions and jargon.
- Fix word lengthening.
- Lemmatizing.

3.4. Feature engineering

Linguistic features are extracted in this step to describe and demonstrate depressive and non-depressive tweets. After feature engineering, the dataset has a total of 10547 rows and 30 columns.

3.4.1. Valence aware dictionary and sentiment reasoner (VADER)

VADER is a lexicon and rule-based sentiment analysis tool designed for social media sentiment [14]. It can detect polarity within a text, as either positive, neutral, or negative. It also tells the intensity of emotion by considering the emphasis on capitalization and punctuation. VADER returns a dictionary that contains negative, positive, neutral, and compound scores. The original tweets are chosen to calculate VADER scores rather than cleaned tweets because original tweets retain the intensity of emotion such as capitalization and punctuation. The return value "compound" is added to the dataset as a summary of polarity scores.

3.4.2. National research Council Canada affect lexicon (NRCLex)

NRCLex is an extended NRC lexicon of WordNet synonyms based on the NLTK library, ranging from approximately 10,000 words to 27,000 words. It can extract 10 emotional effects behind the text, including fear, anger, anticipation, trust, surprise, positive, negative, sadness, disgust, and joy [15]. Since NRCLex requires text cleaning, the cleaned tweets are selected to calculate the emotion lexicon. The returned affect frequencies are shown in Fig**.** 1.

fear	anger	anticipation	trust	surprise	positive	negative	sadness	disgust	joy
0.25	0.25	0.000000	0.000000	0.000000	0.250000	0.25	0.0	0.0	0.000000
0.00	0.00	0.000000	0.000000	0.000000	0.000000	0.00	0.0	0.0	0.000000
0.00	0.00	0.000000	0.000000	0.000000	0.000000	0.00	0.0	0.0	0.000000
0.00	0.00	0.142857	0.428571	0.142857	0.142857	0.00	0.0	0.0	0.142857

Fig. 1. Example result of NRCLex.

3.4.3. Linguistic inquiry and word count (LIWC)

LIWC is a text analysis program that counts words in over 80 categories with psychological meaning. The dictionary is one of the essential components of the LIWC program as it specifies exactly what words are to be counted [16]. A customized version of the dictionary was created by referring to the concept of the original LIWC dictionary to detect depressive symptoms from tweets.

Figure 2 shows the content of the customized LIWC dictionary created using an Excel spreadsheet. The first two categories of personal pronouns are included because depressed people tend to use first person pronouns, while non-depressed people tend to use second or third person pronouns [17]. Next, categories 3 to 9 represent the 7 depressive symptoms, and categories 10 to 13 represent the 3 psychological stressors. Psychological stressors refer to the social and physical environmental circumstances that affect one’s mental health [18]. The dictionary contains only 142 words, because the collection of words was manually collected from a limited number of papers and resources on this topic were limited.

%	
1	firstperson
2	second/thirdperson
3	weight/appetite
4	disturbedsleep
5	agitation/retardation
6	fatigue/lossenergy
7	worthlessness/guilt
8	troubleconcentrate
9	suicidal
10	primarysupport
11	occupational
12	housing
13	sadness/loneliness
%	
i	1
my	1
me	1

Fig. 2. Content of customized LIWC dictionary.

3.4.4. Gibbs sampling Dirichlet multinomial mixture (GSDMM)

GSDMM is a short text clustering model modified from Latent Dirichlet Allocation (LDA). Compared to LDA, it proves to be more suitable for detecting topics in shorter texts like tweets due to their different generative assumptions. Parameters used to initialize GSDMM include K=10, alpha=0.1, beta=0.1, and n_iter=30. Topics are assigned to tweet only if the threshold is equal or higher than 0.3. Otherwise, the tweet will be assigned topic 10 to denote another topic.

3.5. Classification model

This stage constructs classification models to identify depressive tweets. In this work, four popular classifiers are employed: Linear Regression (LR), Support Vector Machine (SVM), Random Forest (RF), and Neural Network (NN). Models were trained using different word embedding methods, including Term Frequency-Inverse Document Frequency (TF-IDF), Word2Vec, and Bidirectional Encoder Representations from Transformers (BERT). Besides, models are built with and without linguistic features extracted from previous steps to compare performance. After model training, 10-fold cross validation was used to evaluate the mean performance of the model. To create training samples, the dataset was shuffled and then 5000 samples were selected from normal and depressed tweets. Four columns including username, datetime, original tweet, and target were removed from the training sample. Lastly, split the sample into 80% training data and 20% testing data.

3.5.1. Logistic regression

Logistic Regression is a commonly used classifier that is similar to Linear Regression but with binomial variables, such as true/false, yes/no, and 1/0. The predicted value of Linear Regression can exceed the range of 0 and 1, while Logistic Regression lies within the range of 0 and 1.

3.5.2. Support vector machine (SVM)

SVM has been found to be the most widely used machine learning classifier for mental disorder prediction. It is widely used for text categorization and performs well with linear kernels, especially when there are a lot of features [19]. SVM creates a hyperplane that partitions two categorical data as wide as possible [9]. It can reduce the probability of overfitting by finding the perfect balance while adjusting several features [20]. In this study, Linear SVC was chosen to implement the SVM algorithm because it is more efficient in terms of execution time, and it scales almost linearly to a large number of samples [21].

3.5.3. Random forest (RF)

RF is a combination of multiple decision trees trained simultaneously. The dataset is first split into a random subset of features. These random subsets are then used to build smaller trees and finally merged together to form more accurate and stable predictions.

3.5.4. Neural network (NN)

Neural Network (NN) is a machine learning algorithm whose structure mimics the neurons of the human brain. Machine learning models make decisions based on what they learn from data, while neural networks can make intelligent decisions on their own by learning from their own mistakes. The neural network created in this study has two relu layers, two dropout layers, and one sigmoid layer. The first relu layer has 64 dimensions and the second layer has 32 dimensions. Both dropout layers have 0.4 dropout rate, and the sigmoid layer has 1 dimension. The relu activation function is used for the hidden layer as it is the standard activation function. Moreover, it does not suffer from the vanishing gradient problem, thus guaranteeing the accuracy of the output prediction. The sigmoid activation function is used to normalize the output to a range between 0 and 1 for binary classification. Furthermore, dropout is a method that force the neural network to learn more robust features using different random subsets of neurons to prevent overfitting. Other parameters include optimizer=Adam, learning_rate=0.001, loss=binary_crossentropy, metrics=accuracy, epochs=20, and batch_size=16.

3.6. Word Embedding Techniques

3.6.1. Term frequency-inverse document frequency (TF-IDF)

TF-IDF is a technique for transforming text into meaningful numerical representations to fit a model. It is a common word embedding method to compute the numerical weightage of words in a document. TF-IDF can be initialized using unigram, bigram, and trigram. Trigram is chosen to initialize TF-IDF in this study to generate a more robust model.

3.6.2. Word2Vec

Word2Vec is different from TF-IDF which uses a simple machine learning algorithm. It is a word embedding method trained on the Google News dataset using a feed-forward neural network with one hidden layer. Word2Vec is able to capture the semantic and syntactic meaning of words in the dataset it was trained on because it creates a vector space where each unique word shares a common context located close to each other [22]. The parameters for Word2Vec model training include vector_size=300, window=8, min_count=1, sg=1, and epochs=30.

3.6.3. Bidirectional encoder representations from transformers (BERT)

BERT is a dynamic word embedding, also known as a transformer. Compared to static word embeddings like Word2Vec, transformers can produce better contextual word representations by using attention to capture the relationships between words in a sentence. Attention will assign high scores to word representations with high similarity and vice versa. Hence, it is able to focus on specific and related features. DistilBERT was chosen to be applied in this study. It is a smaller, faster, and lighter version of BERT for implementing BERT in a more efficient manner [23, 24].

3.7. System requirements

Functional and non-functional system requirements are listed as below. Functional requirements describe what a system is required to do, while non-functional requirements concern how well the system is expected to perform.

3.7.1. Functional system requirements

- The system shall allow the administrator to add doctor and patient information to the database.
- The system shall allow doctors to login to the system using the username and password provided by the administrator.
- The system shall allow patients and public users to register accounts in this system.
- The system shall be able to connect to the Twint API and scrape tweets for analysis.
- The system shall be able to predict depressive tweets using the model deployed to it.
- The system shall be able to visualize the analysis results on the dashboard with table and charts.
- The system shall be able to calculate the depression score and display an alert on the dashboard after analysis.
- The system shall be able to store the analysis results into the database.
- The system shall allow doctors to perform mental health analysis on their patients' Twitter accounts.
- The system shall allow patients and public users to perform mental health analysis on their own Twitter accounts.
- The system shall allow doctors to review the analysis history of each patient.
- The system shall allow doctors and patients to manage their appointments.

3.7.2. Non-functional system requirements

- The system shall have a clean and organized user interface to enhance readability. The elements used shall be familiar and used consistently throughout the system.
- The system shall only be accessed by authorized users.
- The system shall be accessible from anywhere with an internet connection.

- The analysis results shall possess high accuracy in detecting depressive tweets.
- The system shall load the page within 4 seconds after clicking, and the analysis results shall display within 30 seconds of processing.

4. Results

Different models are composed of different algorithms, word-embedding techniques, and training features. Table 3 demonstrates the comparison of model performance. The overall results show that training with all linguistic features leads to better performance than training with only tweets. By comparing the performance of three machine learning models with the same embeddings and features, SVM performed the best, whereas RF performed the worst. RF is the least efficient model in this study because it takes the longest training time. The SVM model is trained again with a more complex word embedding, Word2Vec, and the result showed improvement. Furthermore, the neural network model can further enhance the performance, reaching the highest accuracy of 93.90%.

Table 3. Comparison of model performance.

Model	Label	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Logistic Regression + Trigram TF-IDF + tweets feature	Depressed	87.10		-	
	Normal				
Logistic Regression + Trigram TF-IDF + all features	Depressed	88.70	86.00	92.00	89.00
	Normal		91.00	86.00	89.00
SVM + Trigram TF-IDF + all feature	Depressed	89.80	87.00	94.00	90.00
	Normal		93.00	86.00	90.00
Random Forest + Trigram TF-IDF + all feature	Depressed	88.80	84.00	92.00	88.00
	Normal		92.00	83.00	87.00
SVM + Word2Vec + tweets feature	Depressed	90.20	90.00	91.00	90.00
	Normal		91.00	90.00	90.00
SVM + Word2Vec + all features	Depressed	90.40	90.00	91.00	90.00
	Normal		91.00	90.00	90.00
Neural network + BERT + tweets feature	Depressed	92.85		-	
	Normal				
Neural network + BERT + all features	Depressed	**93.90**		-	
	Normal				

The neural network model was selected as the final model and deployed into the system for use by end users. The system interface is shown in Figs. 3 to 5. Figure 3 represents the dashboard visualization including the patient’s name, date range, analysis table, sentiment over time, word clouds, depression score, and alert message. The depression score is calculated by dividing the total number of depressive tweets by the total number of tweets. Moreover, an alert message is displayed according to the depression score range. Next, Fig. 4 represents the appointment list that the users can request, accept, reject, and cancel appointments. Lastly, Fig. 5 represents the history of analysis results for a specific patient.

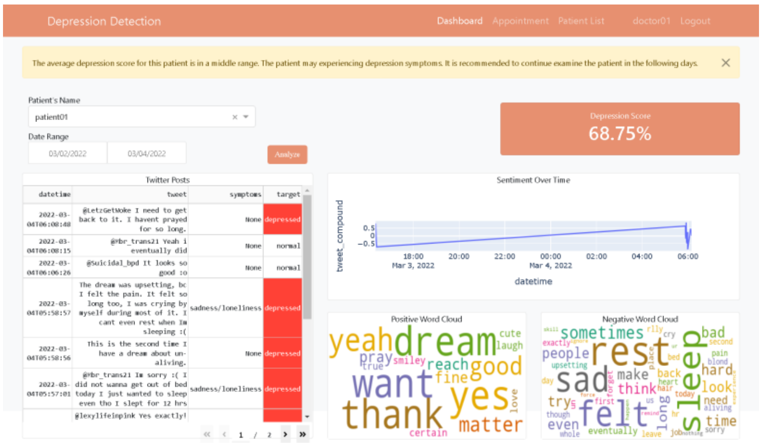

Fig. 3. Dashboard visualization.

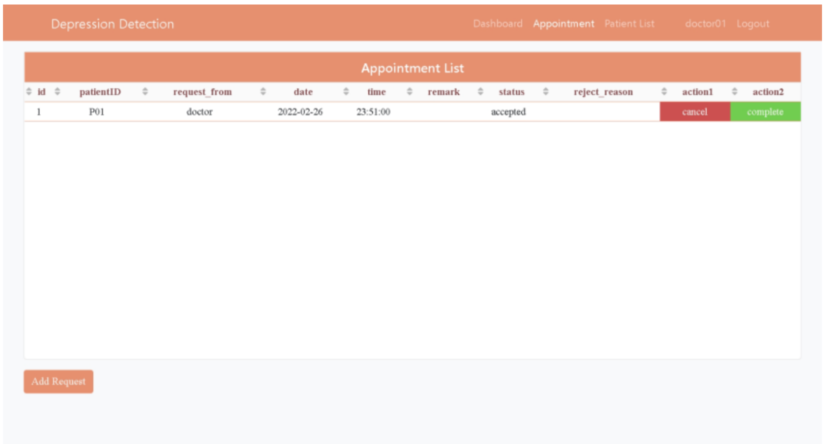

Fig. 4. Appointment list.

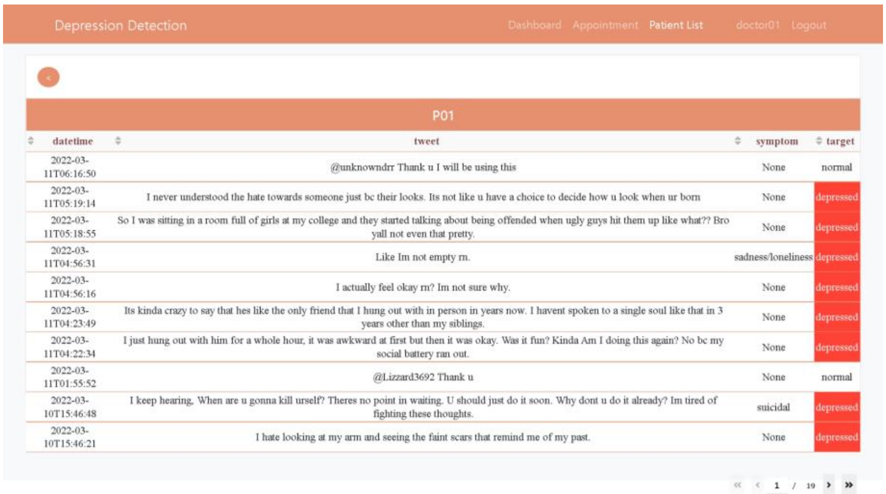

Fig. 5. History of analysis result.

5. Discussions

To better understand the general depressive characteristics exhibited on social media posts, different techniques were applied to extract significant features from the dataset. This includes VADER for calculating polarity scores, NRCLex for obtaining 10 emotional affects, LIWC for identifying depressive symptoms, and GSDMM for detecting topics in shorter texts. Besides, different classification models such as Linear Regression, Support Vector Machine, Random Forest, and Neural Network were applied to identify depressive tweets. In addition, word embeddings like LIWC, Word2Vec, and BERT were used to train the models. Moreover, this study successfully demonstrated the ability of LIWC to detect depressive symptoms from text. These techniques can successfully extract important characteristics from the dataset and show high accuracy in detecting depressive tweets. It can be observed that neural networks give the best outcome, and they can outperform the performance of simple machine learning models. It is important to note that the system developed in this study does not diagnose depression but facilitates the screening process for further diagnosis.

6. Conclusions

This paper demonstrates the capability of using Twitter as a tool to detect depressive tweets. It can assist mental health professionals with the screening process, such as understanding a patient's past experiences and inner thoughts. The system facilitates effective understanding and identifying of the underlying causes of the illness, enabling mental health professionals to make quick decisions and provide patients with appropriate treatment. Additionally, the system has the ability to increase mental health awareness. It offers access to people who are afraid to ask for help but still want to know about their mental health from their Twitter profiles. The system is also able to provide advice and support hotlines to persuade those who are suffering from mental illness to seek professional support.

References

1. DeJohn, A.D.; Schulz, E.E.; Pearson, A.L.; Lachmar, E.M.; and Wittenborn, A.K. (2018). Identifying and understanding communities using Twitter to connect about depression: cross-sectional study. *JMIR Mental Health*, 5(4), e61.
2. Pathirathna, M.L.; Nandasena, H.M.R.K.G.; Atapattu, A.M.M.P.; and Weerasekara, I. (2022). Impact of the COVID-19 pandemic on suicidal attempts and death rates: a systematic review. *BMC Psychiatry*, 22, 506, 1-15.
3. Varma, P.; Junge, M.; Meaklim, H.; and Jackson, M. L. (2021). Younger people are more vulnerable to stress, anxiety and depression during COVID-19 pandemic: A global cross-sectional survey. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 109, 110236.
4. William, D.; and Suhartono, D. (2021). Text-based depression detection on social media posts: A systematic literature review. *Procedia Computer Science*, 179, 582-589..
5. Shen, G.; Jia, J.; Nie, L.; Feng, F.; Zhang, C.; Hu, T.; .Chua, T,-S.; and Zhu, W. (2017). Depression detection via harvesting social media: A multimodal dictionary learning solution. *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence Main Track*, Melbourne, Australia, 3838-3844.

6. Stephen, J.J.; and Prabu, P. (2019). Detecting the magnitude of depression in Twitter users using sentiment analysis. *International Journal of Electrical and Computer Engineering*, 9(4), 3247.
7. Aswathy, K.S.; Rafeeque, P.C.; and Murali, R. (2019). Deep learning approach for the detection of depression in twitter. *Proceedings of the International Conference on Systems*, *Energy Environment* (*ICSEE*) 2019, GCE Kannur, Kerala
8. Haque, A.; Reddi, V.; and Giallanza, T. (2021). Deep learning for suicide and depression identification with unsupervised label correction. In Farkaš, I.; Masulli, P.; Otte, S.; and Wermter, S. (Eds.) *Artificial neural networks and machine learning – ICANN* 2021. *ICANN* 2021. Lecture Notes in Computer Science, vol 12895. Springer, Cham.
9. Leiva, V.; and Freire, A. (2017). Towards suicide prevention: early detection of depression on social media. *Lecture Notes in Computer Science*. Springer; Berlin/Heidelberg, Germany, 428-436.
10. Cacheda, F.; Fernandez, D.; Novoa, F.J.; and Carneiro, V. (2019). Early detection of depression: social network analysis and random forest techniques. *Journal of Medical Internet Research*, 21(6): e12554.
11. Ma, L.; Wang, Z.; and Zhang, Y. (2017). Extracting depression symptoms from social networks and web blogs via text mining. In Cai, Z.; Daescu, O.; and Li, M. (Eds.). *Bioinformatics research and applications*. *ISBRA* 2017. Lecture Notes in Computer Science, vol 10330. Springer, Cham.
12. Ahmad, H.; Asghar, M.Z.; Alotaibi, F.M.; and Hameed, I.A. (2020). Applying deep learning technique for depression classification in social media text. *Journal of Medical Imaging and Health Informatics*, 10(10), 2446-2451.
13. Mbarek, A.; Jamoussi, S.; Charfi, A.; and Hamadou, A.B. (2019). Suicidal profiles detection in Twitter. *Proceedings of the* 15*th International Conference on Web Information Systems and Technologies* (*WEBIST* 2019), Setubal, Portugal, 289-296.
14. Hutto, C.J.; and Gilbert, E.E. (2020, May 22). vaderSentiment. Retrieved March 3, 2022, from PyPI: https://pypi.org/project/vaderSentiment/
15. Bailey, M.M. (2019). NRCLex. Retrieved March 3, 2022, from PyPI: https://pypi.org/project/NRCLex/
16. Pennebaker, J.W.; Boyd, R.L.; Jordan, K.; and Blackburn, K. (2015). *The development and psychometric properties of LIWC*2015. Austin, TX: University of Texas at Austin.
17. Hussain, J.; Satti, F.A.; Afzal, M.; Khan, W.A.; Bilal, H.S.M.; Ansaar, M. Z.; Ahmad, H.F.; Hur, T.; Bang, J.; Kim, J.-I.; Park, G.W.; Seung, H.; and Lee, S. (2020). Exploring the dominant features of social media for depression detection. *Journal of Information Science*, 46(6), 739-759.
18. Mowery, D.; Smith, H.; Cheney, T.; Stoddard, G.; Coppersmith, G.; Bryan, C.; and Conway, M. (2017). Understanding depressive symptoms and psychosocial stressors on Twitter: a corpus-based study. *Journal of Medical Internet Research*, 19(2), e48.
19. Paul, S.; Kalyani, J.S.; and Basu, T. (2018). Early detection of signs of anorexia and depression over social media using effective machine learning

frameworks. Proceedings of Conference and Labs of the Evaluation Forum *CLEF* (*Working notes*), Avignon - France.

20. AlSagri, H.S.; and Ykhlef, M. (2020). Machine learning-based approach for depression detection in Twitter using content and activity features. *IEICE Transactions on Information and Systems*, 103(8), 1825-1832.
21. scikit-learn (2021). 1.4. Support Vector Machines. Retrieved March 7, 2022, from scikit-learn: https://scikit-learn.org/stable/modules/svm.html
22. Di Gennaro, G.; Buonanno, A.; and Palmieri, F.A. (2021). Considerations about learning Word2Vec. *The Journal of Supercomputing*, 77(11), 12320-12335.
23. Sanh, V.; Debut, L.; Chaumond, J.; and Wolf, T. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:*1910.01108*v*4.
24. Surono, S.; Goh, K.W.; Onn, C.W.; and Marestiani, F. (2023). Developing an optimized recurrent neural network model for air quality prediction using K-means clustering and PCA dimension reduction. *International Journal of Innovative Research and Scientific Studies*, 6(2), 330-343.