Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control
Journal homepage: http://kinetik.umm.ac.id
ISSN: 2503-2267
Vol. 5, No. 4, November 2020, Pp. 309-316

Mental disorder detection via social media mining using deep learning

Binti Kholifah*[1], Iwan Syarif[2], Tessy Badriyah[3]
Politeknik Elektronika Negeri Surabaya, Indonesia[1,2,3]

Article Info

Keywords:
Mental Disorder, Depression, Natural Languange Processing, Social Media Mining, Computational Linguistic

Article history:
Received 25 September 2020
Accepted 18 November 2020
Published 30 November 2020

Cite:
Kholifah, B., Syarif, I., & Badriyah, T. (2020). Mental Disorder Detection via Social Media Mining using Deep Learning. Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control, 5(4). doi:https://doi.org/10.22219/kinetik.v5i4.1120

*Corresponding author.
Binti Kholifah
E-mail address:
bintikholifah@pasca.student.pens.ac.id

Abstract

Due to the imperceptible nature of mental disorders, diagnosing a patient with a mental disorder is a challenging task. Therefore, detection in people with mental disorders can be done by looking at the symptoms they experience. One symptom in patients with mental disorders is solitude. Patients with mental disorders feel indifferent to their environment and mainly focus on their own thoughts and emotions. Therefore, the patient looks for a place that can accommodate his feelings. Twitter is one of the most widely used media in measuring one's personality through everyday statements. The symptoms as suggested by psychologists can be explored more broadly using Natural Languages Processing. The process involves taking a lexicon containing keywords that could indicate symptoms of depression. This study uses five criteria as a measure of mental health in a statement: sentiment, basic emotions, the use of personal pronouns, absolutist words, and negative words. The results show that the use of sentiments, emotions, and negative words in a statement is very influential in determining the level of depression. A depressed person more often uses negative words that indicate his self-despair, prolonged sadness, even suicidal thoughts (e.g. "sadly", “scared”, “die”, “suicide”). In the classification process, LSTM Deep Learning generates an accuracy of 70.89%; precision of 50.24%; recall 70.89%.

1. Introduction

Depression is a major cause of mental disorders and disabilities throughout the world. Patients with mental disorders that are not treated immediately could see their mental conditions deteriorate further, some even resort to suicide. World Health Organization data shows that more than 350 million people suffer from depression. This number increased by more than 18% between 2005 and 2015. Mental disorders pose a huge cost to the world economy. In 2010. spending due to mental disorders reached $ 2.5 trillion and is expected to increase to $ 6 trillion by 2030. On the other hand, mental disorder creates a burden on society in general. Unfortunately, the availability of mental health services is still not sufficient in accommodating the increase in the number of mental disorders patients [1]. Inadequate access to care is caused by a lack of support for people with mental disorders as well as fear of negative stigma about mental disorders.

Early detection of mental disorders may significantly help patients from developing more serious psychological problems. Mental disorders cannot be seen physically, hence they may be detected from symptoms experienced through behaviour or attitude. Patients may show some obvious symptoms of depression, but some patients behave normally. Depression changes the way people move, sleep, and how they interact with environment.

Detection can be based on experiences reported by patients themselves, behaviour reported by relatives or friends, and mental status checks. Most studies of mental disorders come from detection through social media. Scientists have tried to spell out the right relationship between depression and language, with natural language processing technology [2]. The majority of people with mental disorders prefer to be alone. Therefore, patients with mental disorders will find a place that can accommodate the overflow of the patient's mind. Social media is a place to tell stories for mental patients, especially Twitter [3]. Twitter is considered capable of describing the feelings of users more strongly. Statements on tweets will provide useful information as a tool to determine policies, so that text mining can be done on tweet data.

This research is based on evaluating emotions in a statement through a linguistic approach. There are three groups of symptoms of depression that can determine the severity, including psychological, social, and physical. The mental disorders that occur can affect a person's mood. It will last longer when it becomes the main symptom of depression [4]. Patients with mental disorders will feel sad for an extended period, despair, anxiety, and low self-worth. Moreover, patients may have little concern for their surroundings, including friends and even family. Mental disorders will change a person's life habits and in turn affect their physical health, such as diet and sleep patterns [5]. An effective method of detecting the severity of mental disorders can help health professionals treat depression patients early.

Deep learning is used to detect the severity of depression by analysing statements written by social media users Twitter. Several parameters will be extracted from the user's statement to show symptoms of depression. Statements written by mental patients will lead to more negative emotions and expressions [2]. This study will analyse user statements within a specific period using LSTM deep learning. The model produced by LSTM can store old information and relate it to new information, so that there is a link between information while still paying attention to the main information [6]. The purpose is not only to classify mental disorder patients and non-patients, but also to classify the patient's depression level, so that patients can receive treatment according to the level of depression.

2. Research Methods

The design of the system in this study can be seen in Figure 1. The system consists of 4 main parts, including the process of data collection, feature extraction, data labelling, and data classification using deep learning. The data collected is sourced from Twitter user statements. The data obtained is raw data, which still contains many characters that are not needed in the next stage. Therefore, after the data is collected, text pre-processing must be done. Feature extraction is the stage of obtaining information from features possessed by data sources (data tweets) collected. The features used in this study are features that serve as benchmarks for the mental health of research subjects. In this study, there are five features that will be extracted in the tweet data.

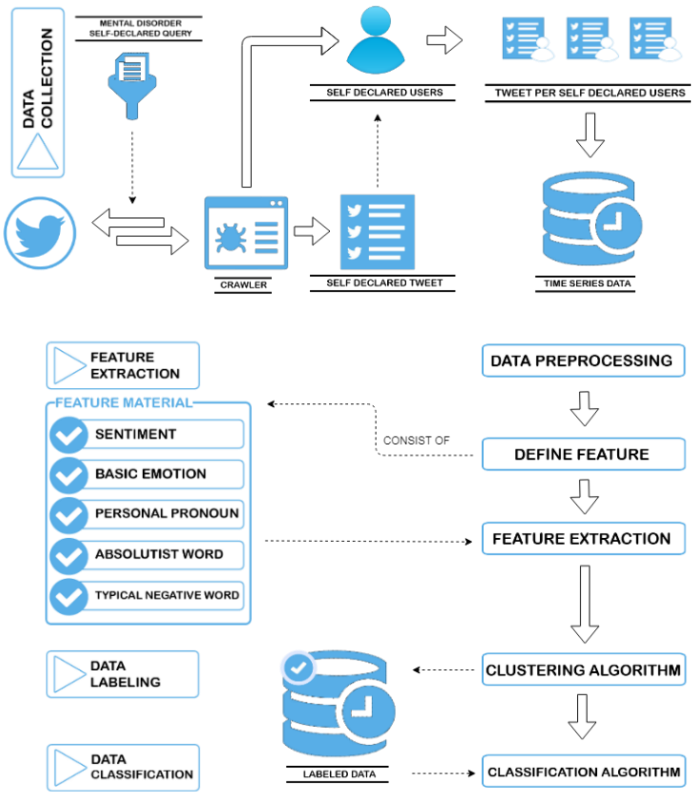

Figure 1. System Design

Material features that will be extracted can be seen in Figure 2 in the feature extraction section, including sentiment, basic emotion, personal pronoun, absolutist word, and typical negative word. The result of feature extraction is data that does not have a label. This study uses the help of the hierarchical clustering algorithm in categorizing data into three levels of depression, including low, moderate, and high. Data that has been labelled will be classified using LSTM deep learning classification. Every part of the system design will be explained in the next subsection.

2.1 Datasets

The data collection process in this study is presented in Figure 2 in the "Data Collection" section. This research uses a minimum of 75% of English tweet data. Data is collected from users who have been diagnosed with depression or states that users are depressed with a "mental disorder self-declared query". Examples of self-declared queries used are "I am diagnosed with depression disorder", "I am in depression", and several other queries. The crawling process generates self-declared tweets so that it can be known to self-declared users and users lists. The second data crawling is based on the username obtained from the user's list. Data was collected from Twitter for a specified period of time from accounts that were members of the self declared depression community. The final process is to combine all the tweet data from self-declared users to form complete data.

Table 1. Self-Declared Tweet

user	created_at	text
user1	26/03/2020 19:24	My reoccurring major depression disorder is kicking my ass.
user2	25/03/2020 12:30	The term is an useful adjective, so I'm glad there's now a distinction between the 2.
user3	25/03/2020 12:18	I knew instantly, (a realisation I wish had come years earlier) that the number on the scale was never going to fill the void that lay deep within me.

An overview of the results of the crawling process can be seen in Table 1. In the example; three users stated that they were diagnosed with depression. The user username obtained at this stage will be used in the next data crawling process so that there are two data crawling methods. The second data crawling does not use the query anymore but is done using the username obtained from Self-Declared Tweets. This process is carried out until several tweets are obtained per self-declared users.

Before entering the feature extraction stage, data processing must be done first. The tweet data obtained is raw and still contains unnecessary characters, so the data must be cleared first.

1. First, the entire text is converted into lowercase letters.
2. Second, contractions such as “won’t”, “can’t”, “I’m” are converted into their original form.
3. Third, characters such as punctuations and numbers are removed.
4. Fourth, tweets are separated into data per word. This process is also known as tokenizing.

The last process is stopword removal which is the process of removing words that are less important. This process is not required at several stages of feature extraction. For example, in the extraction of the personal pronoun feature, if a stopword removal is performed, then the personal pronoun feature will not be found in the tweet. This is because all pronouns or pronouns enter the stopword list, so they will definitely be deleted at the time of the stopword removal. Absolute words are the words contained in the stopword list. If stopword removal is done, the Absolutist words feature will not be extracted properly. So at this stage, stopword removal is not necessary.

2.2 Depression Detection

In mental health service standards, several techniques are needed to detect depression, such as filling out questionnaires, psychological tests, and interviews. But the response of this technique cannot represent the symptoms and the severity of depression. Therefore, detecting depression from daily living habits can solve the lack of existing methods. Online behaviour is considered to be an excellent source to describe the emotional state [7].

Natural Language Processing (NLP) on social media began to develop in mental health media. There are several studies that combine NLP with machine learning methods in identifying opportunities for someone to suffer from mental disorders [8]. There have been many studies related to NLP technology in the mental health sector, such as sentiment analysis and affective computing for depression monitoring [9], measure post traumatic stress disorder on twitter [10], and sentiment analysis of social networking site data using machine learning to measure depression [11].

Social media is widely used in measuring personality. The mental disorder symptoms can be explored more broadly using NLP. Before doing text mining, a dictionary that contains keywords is needed or in other words, a dictionary containing symptoms of depression in the form of words [12]. Textual data is the most widely used form of communication in analysing the emotions contained therein.

Syarif et. Al (2019) explores the basic emotional features contained in social media user statements. Basic emotions are considered to be very related to one's mental state [2]. There are innumerable types of emotions, such as joy, sadness, joy, love, hate, anger, fear, and anxiety. Six basic emotions will be used, including angry, bored, excited, fear, happy, and sad. From the tweet data obtained, extraction of the six basic emotions that have been determined will be carried out, so that the emotion tendency of the user can be seen in each tweet [13].

Besides, this study also explores the sentiment features contained in the text. Sentiment analysis is used to determine the attitudes, opinions, and emotions contained in a document. This technique is a way to evaluate statements into three categories: positive, negative, or neutralThe results showed that mental patients tended to make negative statements [11].

People who experience physical or emotional pain inclined are more to only focus on themselves [14]. Research on depression shows that the use of language can be used as a feature to recognise mental health. A person with a depressive disorder is more focused on himself and more often expresses himself with negative emotions and even uses words that lead to death. Someone with a depression disorder tends to use the first person singular pronouns (e.g., "I, me, mine") rather than someone who is not depressed [15].

Other studies suggest that words found absolutist words are better features than pronoun features or features of negative emotion words. Absolutist words are stress words in the form of delivering specific quantities or probable possibilities (e.g., "always, none, never"). Absolutist thinking is considered a deviation by most therapists in people with anxiety disorders and depression. It was found that patients who were depressed and had suicidal thoughts more often used absolute words in their statements [16].

In this study, basic emotions, sentiment analysis, and language style (negative words, first-person singular pronouns, and absolutist words), contained in a statement, are used as features to recognise the severity of depression.

2.3 Classification

This study uses LSTM (Long Short - Term Memory) as a method for classification. However, text data is obtained in the form of unsupervised, so to group the data into several levels of depression requires a long time. categorization of data is done using a clustering algorithm. Data will be grouped into three levels of mental disorder using Hierarchical Clustering. The purpose of clustering is to allocate each data object into the appropriate cluster [17]. The data will be grouped into 3 levels of mental disorder using Hierarchical Clustering.

In the Hierarchical Clustering method, there are two basic types, agglomerative and divisive. This research uses agglomerative clustering. Agglomerative clustering classifies the hierarchy of each object in a separate cluster and then forms an increasingly more massive cluster. So, the number of initial clusters is the same as the amount of data. The grouping technique that will be used is complete linkage. This technique combines clusters according to the distance between the farthest members of the two clusters [18]. The technique starts by making all points as individual clusters, then combining the two points that have a minimum distance. Next, the euclidian matrix is updated by calculating the distance between clusters that have been combined using the Equation 1 [19].

$$d(p,q)=\sqrt{(q_1-p_1)^2+(q_2-p_2)^2} \qquad (1)$$

The points that have the maximum Euclidean distance from the cluster must be combined with the next cluster. The process of finding the euclidean distance must be repeated until there is only one cluster with centroid i [18]. Data that has been labelled will be classified using deep learning. In this research, the deep learning algorithm that will be used is LSTM (Long-Short Term Memory). LSTM is one of the deep learning methods introduced by Hochreiter and Schmidhuber in 1997 until now has undergone many developments. LSTM can solve problems that cannot be explained by the Recurrent Neural Network algorithm [20].

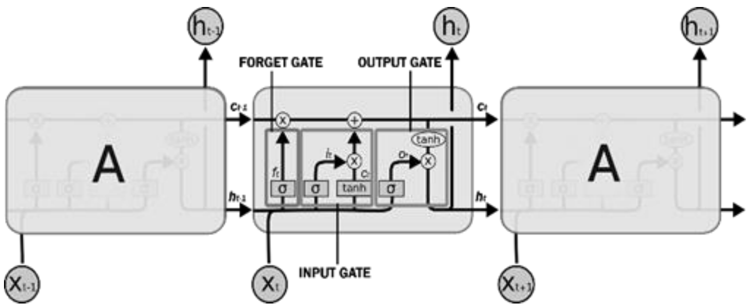

Figure 2. LSTM Network Layer, Olah (2015)

LSTM has a chain structure like in Figure 2. In contrast to ordinary neural networks, LSTM has four layers where one another is interconnected in determining policies. LSTM uses three gates namely input gate, forget gate, and output gate to control usage and update information [21].

In the first stage, the gate will choose whether information from $X_{(t-1)}$ and X_t can be passed on. Decisions are made by sigmoid layers which can be called "Forget Gate". Output 1 indicates that information can be continued and output 0 suggests the opposite. The calculation on forget gate can be seen in the Equation 2.

$$f_t= \sigma(W_f.[h_{t-1},x_t]+b_f) \qquad (2)$$

The next step is determining the new information that will be stored in a state cell. The sigmoid layer, called the input gate, decides which value to update. Then the tanh layer creates a new value vector (C˜_t) that can be added to the state. The two values are combined to create a new state. Calculation of input gate values is done using Equation 3, and new candidate values are calculated using Equation 4.

$$i_t=\sigma(W_i.[h_{t-1}, x_t]+b_i) \tag{3}$$

$$\widetilde{C}_t=tanh(W_c.[h_{t-1}]+b_c) \tag{4}$$

The next step is updating the old cell status (C_ (t-1)) to the new cell status C_t by multiplying the old cell condition with forget gate and adding i_t + C ̃_t as in Equation 5.

$$C_t=f_t * C_{t-1}+i_t * \widetilde{C}_t \tag{5}$$

The last stage is running the sigmoid layer where this process will determine which cells are the output. The next process is to put the state of the cell through the tanh and increase the output of the sigmoid gate. The gate output can be calculated by Equation 6 and Equation 7.

$$O_t=\sigma(W_O.[h_{t-1},x_t]+b_O) \tag{6}$$

$$h_t=O_t*tanh(C_t) \tag{7}$$

Sequence data prediction problems are different from learning other data. Information from each sequence must be retained when modeling with training data. Sequence prediction data is divided into four types, including: Sequence Prediction, Sequence Classification, Sequence Generation and Sequence-to-Sequence Prediction.

This research will use sequence classification from labelled time-series data. The purpose of sequence classification is to build a classification model using labelled datasets so that the model that has been built can be used to predict the label of the next sequence. The application of LSTM in sequence classification can be seen in research on DNA sequence classification, anomaly detection, and sentiment analysis [22].

In the case of sentiment analysis, LSTM has a method for extracting features from the text to be analyzed. The method used is word embedding. Hu et al. (2019) used Bi-LSTM with default word embedding to perform sentiment analysis on multi-label text data [23]. In addition, there are studies that combine LSTM with other word embedding models such as word2vec [24]. In this study, the feature extraction used was the symptom through the language style used by mental disorders patients.

2.4 Performance Measurement

Measurement of performance results is done using precision, recall and accuracy as evaluation criteria. The criteria are presented in a table called confusion matrix which is shown in Table 2.

Table 2. Confussion Matrix

	Positive Actual	Negative Actual
Positive Predicted	TP (True Positif)	FP (False Positif) Type I Error
Negative Predicted	FN (False Negative) Type II Error	TN (True Negative)

In measuring performance using a confusion matrix, there are four terms, namely TP, TN, FP, FN. TP is True Positive, namely the number of positive data classified correctly by the system. TN is True Negative, that is, the amount of negative data classified correctly by the system. Meanwhile, FN is False Negative, which is the number of positive data detected as negative data by the system. FP is False Positive, namely the amount of data [25].

Precision is a percentage of the amount of relevant information taken with the total amount of information taken [25]. Precision is obtained by Equation 8.

$$Precision=\frac{TP}{TP+FP}\times 100\% \tag{8}$$

Recall is a percentage of the amount of relevant information taken with the total amount of relevant information in the system [25]. Recall Precision is obtained by Equation 9.

$$Recall=\frac{TP}{TP+FN}\times 100\% \tag{9}$$

Meanwhile, accuracy is the percentage of closeness between the predicted value and the actual value. Accuracy is stated in Equation 10.

$$Accuracy=\frac{TP+FN}{TP+TN+FP+FN}\times 100\% \tag{10}$$

3. Result and Discussion

Sentiment feature extraction use the NLTK library (Natural Language Toolkit), SentimentIntensityAnalyzer dan Vader lexicon. The resulting weighting is the weighting of the three sentiment labels, namely negative, positive, and neutral. The results of weighting can be seen in Figure 3.

compound	neg	neu	pos
-0.2732	0.241	0.602	0.158
-0.7845	0.697	0.303	0.000
0.5859	0.107	0.533	0.360
-0.8020	0.507	0.493	0.000
0.7845	0.000	0.465	0.535

Figure 3. Sentiment Feature Extraction

While in the extraction of basic emotional features, this study utilises Lexicon WordNet. The basic emotions are divided into six, as shown in Figure 4.

Angry	Bored	Excited	Fear	Happy	Sad
0.153574	0.077361	0.086546	0.161340	0.089162	0.432017
0.428009	0.199924	0.091585	0.128574	0.063605	0.088303
0.067917	0.029652	0.139868	0.132515	0.300138	0.329911
0.224721	0.076051	0.014401	0.519544	0.010146	0.155136
0.140864	0.126820	0.213683	0.151062	0.258632	0.108940

Figure 4. Basic Emotion Feature Extraction

In the first person pronoun feature, the pronouns used are I am, I, me, my, myself, and mine. The results of the personal pronoun feature extraction are presented in Figure 5.

iam	i	me	my	myself	mine	personal
0	0	1	0	0	0	1
1	2	0	1	0	0	4
0	3	0	2	0	0	5
0	2	0	2	0	0	4
1	1	0	1	0	0	3

Figure 5. Personal Pronoun Feature Extraction

Then the weighting of each term will be carried out again to obtain one attribute; in other words, the total number of personal pronoun appearances. In the absolutist word feature, the word used is Always, never, none, all, every, only, completely. The results from the extraction of the absolutist word feature are presented in Figure 6. The results are again pursed from the weight of each term to obtain one attribute that is the total number of times the absolute word appears.

always	never	none	alle	every	only	completely	absolut
0	0	0	0	1	0	0	1
0	0	0	1	0	0	0	1
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	1

Figure 6. Absolutist Word Feature Extraction

The Typical Negative Word feature is extracted with a lexicon consisting of 220 negative wordsThen the system will calculate the frequency of occurrence of negative words in the text. Data is divided into three classes, including 0, 1, and 2 as in Figure 7. Therefore, to determine the categories on each label need further analysis. This labelled data will be used at the classification stage, using deep learning.

neg	neu	pos	Angry	Bored	Excited	Fear	Happy	Sad	personal	absolutist	negative	label
0.241	0.602	0.158	0.153574	0.077361	0.086546	0.161340	0.089162	0.432017	1	0	2	0
0.697	0.303	0.000	0.428009	0.199924	0.091585	0.128574	0.063605	0.088303	0	1	0	1
0.107	0.533	0.360	0.067917	0.029652	0.139868	0.132515	0.300138	0.329911	4	1	1	2
0.507	0.493	0.000	0.224721	0.076051	0.014401	0.519544	0.010146	0.155136	5	0	0	1
0.000	0.465	0.535	0.140864	0.126820	0.213683	0.151062	0.258632	0.108940	0	0	2	2

Figure 7. Data Labelled

Classification is done using deep learning LSTM with several optimizations. The results of the experiment are presented in Table 3. Data sharing was done using 4-fold cross validation and 500 epochs.

Table 3. Experiment Result

Optimizer	Accuracy	Precision	Recall
Adam	70,89%	50,24%	70,89%
RMSProp	70,89%	50,24%	70,89%
Adagrad	70,89%	50,24%	70,89%

3. Result Analysis

Labels generated from the results of clustering are still in the form of numbers, so they cannot describe the degree of depression. Analysis is needed to determine the mental disorder category based on the results of the clustering process. From the results of the clustering analysis, it can be concluded that label 0 is data with moderate levels of depression, label 1 is data with high levels of depression, and data with label 2 is data with low levels of depression.

The same research was conducted by Syarit et al. (2019). They explore 4 features of statements in tweets, namely tweet sentiment, basic emotion, first personal pronoun, and the number of positive and negative words contained in the statement. Labeling is done using the rule based tree method. The proposed method successfully categorized 8105 tweets into 3 levels of depression, 1028 tweets were categorized as high, 1.073 moderate, and 1605 low. However, their research results were not indicated by the system performance in measurable units.

We add 2 distinct features of the research conducted by Syarif et al. (2019), namely the use of absolute words and the use of negative words which indicate a mental disorder. Labeling is done using hierarchical clustering while classification is done using LSTM Deep Learning. At the classification stage, this study conducted trials with several optimizers, including Adam, RMSProp, SGD, and Adagrad. The experimental results in Table V show that all three optimizers get the same accuracy that is 70.89%. The resulting performance allows even better optimization of the LSTM and augmentation of the train data. Researchers also need to validate data to experts to ensure that the data used is the right data.

4. Conclusion

This study uses five criteria that can serve as benchmarks for detecting mental illness from a Twitter user statement. The features used are sentiments analysis and six basic emotions. This research builds a lexicon that contains personal pronouns, absolutist words, and negative words. The goal is to calculate the frequency with which words appear in a lexicon.

Our results show that the use of sentiments analysis, emotions, and negative words contained in a statement are very influential in determining a person's level of depression. Words that are often used are words that show negative feelings such as “sadly“,”scared”, “tragic” even words that indicate desire to die such as “die”, “death”, and “suicide”. Also, this study conducted a deep learning classification of data that has been through the process of labelling using hierarchical clustering. Classification is done using several optimizers. All optimizers obtain the same level of accuracy equal to 70.89%.

Future Work

In further research, the accuracy of the classification results can be increased again by optimizing the hyperparameter in the algorithm used and increase the amount of data. Furthermore, the labelling rules in this study can be further enhanced using clustering algorithms with better performance and validation from experts in the field of mental health.

References

[1] A. Wongkoblap, M. A. Vadillo, and V. Curcin, “Detecting and Treating Mental Illness on Social Networks,” *Proc. - 2017 IEEE Int. Conf. Healthc. Informatics, ICHI 2017*, p. 330, 2017. https://doi.org/10.1109/ICHI.2017.24

[2] I. Syarif, N. Ningtias, and T. Badriyah, “Study on Mental Disorder Detection via Social Media Mining,” *2019 4th Int. Conf. Comput. Commun. Secur.*, pp. 1–6, 2019. https://doi.org/10.1109/CCCS.2019.8888096

[3] E. Saravia, C. H. Chang, R. J. De Lorenzo, and Y. S. Chen, “MIDAS: Mental illness detection and analysis via social media,” *Proc. 2016 IEEE/ACM Int. Conf. Adv. Soc. Networks Anal. Mining, ASONAM 2016*, pp. 1418–1421, 2016. https://doi.org/10.1109/ASONAM.2016.7752434

[4] D. Ramalingam, V. Sharma, and P. Zar, “Study of depression analysis using machine learning techniques,” *Int. J. Innov. Technol. Explor. Eng.*, vol. 8, no. 7C2, pp. 187–191, 2019.

[5] N. Al Asad, M. A. Mahmud Pranto, S. Afreen, and M. M. Islam, “Depression Detection by Analyzing Social Media Posts of User,” *2019 IEEE Int. Conf. Signal Process. Information, Commun. Syst. SPICSCON 2019*, pp. 13–17, 2019. https://doi.org/10.1109/SPICSCON48833.2019.9065101

[6] X. Bai, “Text classification based on LSTM and attention,” *Thirteen. Int. Conf. Digit. Inf. Manag. (ICDIM 2018)*, pp. 29–32, 2018. https://doi.org/10.1109/ICDIM.2018.8847061

[7] K. Katchapakirin, K. Wongpatikaseree, P. Yomaboot, and Y. Kaewpitakkun, “Facebook Social Media for Depression Detection in the Thai Community,” *Proceeding 2018 15th Int. Jt. Conf. Comput. Sci. Softw. Eng. JCSSE 2018*, pp. 1–6, 2018. https://doi.org/10.1109/JCSSE.2018.8457362

[8] M. Deshpande and V. Rao, “Depression detection using emotion artificial intelligence,” *Proc. Int. Conf. Intell. Sustain. Syst. ICISS 2017*, no. Iciss, pp. 858–862, 2018. https://doi.org/10.1109/ISS1.2017.8389299

[9] C. Zucco, B. Calabrese, and M. Cannataro, “Sentiment analysis and affective computing for depression monitoring,” *Proc. - 2017 IEEE Int. Conf. Bioinforma. Biomed. BIBM 2017*, vol. 2017-Janua, pp. 1988–1995, 2017. https://doi.org/10.1109/BIBM.2017.8217966

[10] G. A. Coppersmith, C. T. Harman, and M. H. Dredze, “Measuring Post Traumatic Stress Disorder in Twitter,” 2014.

[11] A. U. Hassan, J. Hussain, M. Hussain, M. Sadiq, and S. Lee, “Sentiment analysis of social networking sites (SNS) data using machine learning approach for the measurement of depression,” *Int. Conf. Inf. Commun. Technol. Converg. ICT Converg. Technol. Lead. Fourth Ind. Revolution, ICTC 2017*, vol. 2017-Decem, pp. 138–140, 2017. https://doi.org/10.1109/ICTC.2017.8190959

[12] C. H. Chang, E. Saravia, and Y. S. Chen, “Subconscious Crowdsourcing: A feasible data collection mechanism for mental disorder detection on social media,” *Proc. 2016 IEEE/ACM Int. Conf. Adv. Soc. Networks Anal. Mining, ASONAM 2016*, pp. 374–379, 2016. https://doi.org/10.1109/ASONAM.2016.7752261

[13] P. Ekman, *Emotions revealed*, vol. 328, no. Suppl S5. 2004.

[14] Y. R. Tausczik and J. W. Pennebaker, “The psychological meaning of words: LIWC and computerized text analysis methods,” *J. Lang. Soc. Psychol.*, vol. 29, no. 1, pp. 24–54, 2010. https://doi.org/10.1177%2F0261927X09351676

[15] S. S. Rude, E. M. Gortner, and J. W. Pennebaker, “Language use of depressed and depression-vulnerable college students,” *Cogn. Emot.*, vol. 18, no. 8, pp. 1121–1133, 2004. https://doi.org/10.1080/02699930441000030

[16] M. Al-Mosaiwi and T. Johnstone, “In an Absolute State: Elevated Use of Absolutist Words Is a Marker Specific to Anxiety, Depression, and Suicidal Ideation,” *Clin. Psychol. Sci.*, vol. 6, no. 4, pp. 529–542, 2018. https://doi.org/10.1177%2F2167702617747074

[17] F. Cao and J. Liang, “A data labeling method for clustering categorical data,” *Expert Syst. Appl.*, vol. 38, no. 3, pp. 2381–2385, 2011. https://doi.org/10.1016/j.eswa.2010.08.026

[18] S. Sharma and N. Batra, “Comparative Study of Single Linkage , Complete Linkage , and Ward Method of Agglomerative Clustering,” *2019 Int. Conf. Mach. Learn. Big Data, Cloud Parallel Comput.*, pp. 568–573, 2019. https://doi.org/10.1109/COMITCon.2019.8862232

[19] A. Starczewski, “A New Hierarchical Clustering Algorithm,” pp. 175–180, 2012. https://doi.org/10.1007/978-3-642-29350-4_21

[20] S. Hochreiter, “Long Short-Term Memory,” vol. 1780, pp. 1735–1780, 1997.

[21] C. Olah, “Understanding LSTM Networks,” *Web Page*, pp. 1–13, 2015.

[22] J. Brownlee, *Long Short-Term Memory Networks With Python Develop Sequence Prediction Models With Deep Learning*, V1.0. Jason Brownlee, 2017.

[23] J. Hu, X. Kang, S. Nishide, and F. Ren, “Text multi-label sentiment analysis based on Bi-LSTM,” *Proceeding CCIS*, pp. 16–20, 2019. https://doi.org/10.1109/CCIS48116.2019.9073727

[24] L. Xiao, G. Wang, and Y. Zuo, “Research on Patent Text Classification Based on Word2Vec and LSTM,” *Proc. - 2018 11th Int. Symp. Comput. Intell. Des. Isc. 2018*, vol. 1, pp. 71–74, 2018. https://doi.org/10.1109/ISCID.2018.00023

[25] Y. Luan and S. Lin, “Research on Text Classification Based on CNN and LSTM,” *Proc. 2019 IEEE Int. Conf. Artif. Intell. Comput. Appl. ICAICA 2019*, pp. 352–355, 2019. https://doi.org/10.1109/ICAICA.2019.8873454