



# Depression Levels Detection Through Twitter Tweets Using RoBERTa Method

Algi Erwangga Putra\*, Warih Maharani

School of Computing, Informatics, Telkom University, Bandung

Jl. Telekomunikasi No. 1, Terusan Buahbatu - Bojongsoang, Sukapura, Kec. Dayeuhkolot, Kabupaten Bandung, Jawa Barat, Indonesia

Email: <sup>1</sup>\*algiawangga@student.telkomuniversity.ac.id, <sup>2</sup>wmaharani@telkomuniversity.ac.id

Email Penulis Korespondensi : algi.erwangga99@gmail.com

Submitted: 99/99/999; Accepted: 99/99/999; Published: 99/99/999

**Abstrak**—Kondisi kesehatan mental adalah salah satu hal yang perlu diperhatikan sama pentingnya dengan kesehatan fisik. Depresi merupakan gangguan mental yang dapat mempengaruhi kehidupan sosial seseorang. Platform sosial media twitter merupakan salah satu tempat dimana para penggunanya dapat mencurahkan seluruh isi hati dalam bentuk cuitan. Hal tersebut sering kali melatar belakangi tingkat depresi seseorang. Interaksi yang bebas serta beragam pada twitter memiliki pengaruh yang cukup besar dalam kondisi psikologis penggunanya. Penelitian ini bertujuan untuk mendeteksi tingkat depresi melalui data yang didapatkan dari cuitan media sosial twitter. Metode yang digunakan adalah RoBERTa yang merupakan pelatihan ulang BERT yang dioptimalkan. Beberapa skenario dipakai untuk mendapatkan hasil akurasi terbaik pada klasifikasi teks. Untuk mendapatkan hasil evaluasi diperlukan pengukuran dengan metode Confusion Matrix. Nilai akurasi yang didapatkan yaitu 72% berdasarkan berbagai pengujian yang telah dilakukan.

**Kata Kunci:** Depresi; Twitter; BERT; RoBERTa; Confusion Matrix

**Abstract**—Mental health conditions are one thing that needs to be considered as important as physical health. Depression is a mental disorder that can affect a person's social life. The Twitter social media platform is where users can pour their hearts out in the form of tweets. This is often the background of a person's level of depression. Accessible and diverse interactions on Twitter considerably influence the psychological condition of its users. This study aims to detect the level of depression through data obtained from Twitter social media tweets. The method used is RoBERTa which is an optimized BERT retraining. Several scenarios are used to get the best accuracy results in text classification. To get the evaluation results, it is necessary to measure using the Confusion Matrix method. The accuracy value is 72% based on numerous tests that have been done.

**Keywords:** Depression; Twitter; BERT; RoBERTa; Confusion Matrix

## 1. INTRODUCTION

Mental health is just as essential as physical health, mental health concerns must be taken into account. Issues with mental health can affect one's physical and mental well-being. This can affect mood, behavior, feelings and thoughts. Symptoms of this condition are determined by how far the observation of daily behavior. Depression is one of the most prevalent mental health issues. Depression is an emotional condition that can interfere with mood because it is usually characterized by deep feelings of guilt, anti-social tendencies, loss of pleasure and interest in usual activities. Depression can also affect sleep patterns and is unacceptable in a short time[1].

Today, many people often use their hearts and complaints on social media, given the rapid and easily accessible development of social media. This has become a new habit in society in expressing themselves in cyberspace, and some people make social media one of their primary needs. One of the media that is often used is Twitter. Twitter is a social networking and microblogging service that allows users to share information, business, and share views with a limit of 140 characters called Tweets and allows users to interact with each other through comments and direct message features that are sent privately[2]. However, with this freedom of opinion, interactions on Twitter always lead to positive interactions, and many Twitter users are optimistic about ethics in social media. This is often the background of a person's level of depression. Free and varied interactions on Twitter have a considerable influence on the psychological condition of its users.

Research conducted in 2020 by Muhammad Ardhi examined the detection of depression. The study was conducted using data from Reddit forum posts. The analysis was carried out using the Support Vector Machine (SVM) classification. From the steps carried out, the resulting accuracy value is 97.94%, while if using chi-square, the accuracy value increases to 98.45%. This study concludes that using Chi-square feature selection provides an accurate accuracy value. This study's weakness is that it can only develop a classification into two classes[3]. Further research was conducted in 2020 by Ghalib Mahendra, who examined the detection of depression symptoms. The research was conducted using data from Twitter with analysis using the K-Nearest Neighbor classification. In this study, the accuracy value is 84%, precision is 82%, and recall is 84%, which produces an f1-measure 82%. This study concludes that the accuracy results are considered quite good[4]. Tubagus Rahman Ramadan conducted another study in 2020 regarding the detection of symptoms of depression. The research was conducted using data from Twitter using Naive Bayes classification and validation using k-fold cross with k = 10, which was then evaluated using a confusion matrix. The results obtained are 83.68% accuracy, 61.51% precision, 50.23% recall, and 46.20% f1-score. From the research conducted, the authors conclude that the accuracy results show that the Naive Bayes algorithm works quite well[5]. Yinhan Liu, et al. In 2020 conducted a study on the expansion of the BERT architecture, which was named RoBERTa. The study was carried out as a development of

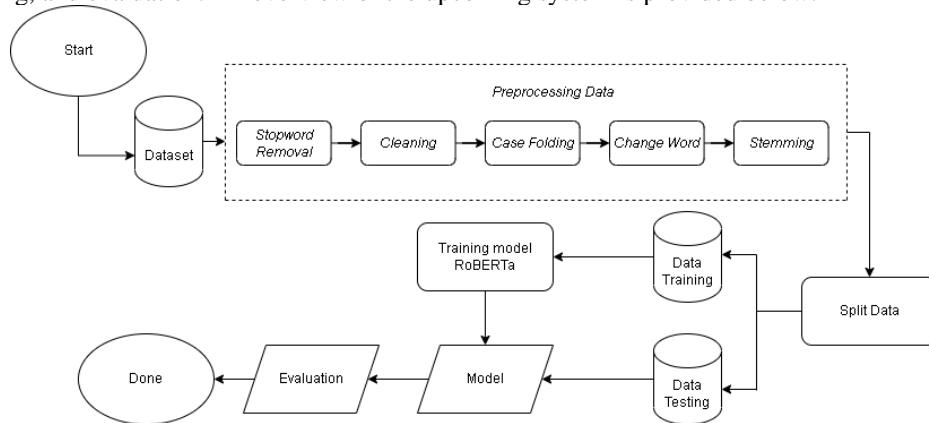
the BERT architecture, which showed a significant reduction compared to the models published after that. This is evidenced by the better performance of the RoBERTa model by offering a better F-score value than the BERT architecture[6]. Research conducted by Benny Richardson in 2021 regarding the Implementation of IndoBERT Lite and RoBERTa for Text Mining on the Jacob Chatbot Application. This study aims to design text mining through a web service on the Jacob chatbot with the aim that the Jacob chatbot gets information in Indonesian related to questions given by the user. The evaluation carried out in this study was carried out by measuring the accuracy value and F-score. The author tested three pre-trained on the RoBERTa model with unsatisfactory results on the three pre-trained. Meanwhile, the fine-tuned RoBERTa-1.5gb-tydiqa model produces the best value than the other fine-tuned RoBERTa models with an accuracy of 0.8 and an F-score of 0.87. At the same time, the fine-tuned indobert-lite-squad model gets an accuracy value of 0.8 and an F-score of 0.89. This study has a weakness in the pre-trained RoBERTa model because it does not test the large RoBERTa hyperparameter, which is predicted to further improve the results of the pre-trained RoBERTa to get maximum results[7].

Therefore, this study aims to detect a person's level of depression on social media, especially on the Twitter platform. The benchmark in this study is based on reciprocity and interaction on someone's tweet on Twitter. The method that will be used in this research is RoBERTa. RoBERTa is a BERT retraining with an improved training methodology, more data receptivity, and computational power. In addition, the method eliminates Next Sentence Prediction (NSP) on BERT and uses Dynamic Masking. With several advantages possessed by RoBERTa, this method is superior by increasing performance compared to the BERT method[6].

## 2. RESEARCH METHODOLOGY

### 2.1 Research Flow

The technique developed for this study, and the system built was able to detect the level of depression in Twitter tweets. There are several stages to detecting, including dataset preparation, data preprocessing, splitting of the data, modeling, and evaluation. An overview of the upcoming system is provided below:



**Figure 1.** Stages of the system to be built

#### 2.1.1 Dataset

The data collected is data obtained from Indonesian Twitter tweets. Internet technology has made significant development in human life. At this time, the development of internet technology does not only refer to browsing, chatting, or blogging but has penetrated social networking services. Apart from social networking services, we are often faced with blogging services. Now, the blog service has been developed into a microblog. Microblog is a blog service that allows users to update their status and upload it. Several microblogging services are currently available; one example is Twitter[8]. Twitter is a social networking and microblog service that allows users to share information, business, and views with a limit of 140 characters called Tweets[9]. Every day as many as 550 million tweets are posted by all Twitter users[10]. In Indonesia, the number of Twitter users in 2021 will reach 16.32 million [11]. If examined further, the number of users and the number of tweets uploaded per day can help research such as detecting depression levels through Twitter tweets.

Depression is a word that can describe a nuance of meaning. Someone must have felt sad, disappointed, frustrated, and lived life with problems that could lead to despair. This is an early phase of depression. Usually, depression occurs when a person experiences excessive stress[12]. Depression can also be said as a mental disorder that occurs in general[13]. This has affected 264 million people worldwide, or about 4.4% of the world's population, characterized by persistent sadness and a lack of interest in valuable activities. The World Health Organization (WHO) states that depression is more common in women (5.1%) than men (3.6%). The impact of this mental disorder can be fatal, one which is not a few people who end their lives due to depression. In 2015, data shows that as many as 788,000 people were killed by suicide[14]. That way, this needs attention in terms of



the various adverse risks that every person can experience with depression. When a person is detected as having symptoms of depression early on, appropriate treatment can be given immediately to minimize the significant risks caused by depression[12].

Before data collection, respondents were required to fill out the DASS-42 form containing questions about depression. Depression Anxiety Stress Scale 42, commonly known as DASS42, is a benchmark to determine a person's emotional state with 42 questions in the questionnaire developed by Lovibond PF and Lovibond SH. DASS-42 has three emotional levels that are categorized as depression, anxiety, and stress[15]. In the DASS, depression is characterized by low positive affect, loss of self-esteem to hopelessness. Concern includes things related to physical arousal, fear, and panic. At the same time, stress is characterized by tension, and irritability, to tend to do more in a situation. DASS can provide limits that aim to show the level of severity which can be seen in the following table 1[16].

Table 1. Category level DASS-42

Severity Level	Depression	Worried	Stress
Normal	0 – 9	0 – 7	0 - 14
Mild	10 – 13	8 – 9	15 – 18
Currently	14 – 20	10 – 14	19 – 25
Heavy	21 – 27	15 – 19	26 – 33
Very Heavy	>28	>20	>34

Next, crawl the Twitter tweets on the respondent's account who has filled out the approval in the previous state and combine the data into a csv file in which each tweet is labeled. Label one explains the tweet is included in the category of depressive symptoms, and label zero indicates that the tweet is included in the type of asymptomatic depression. The data obtained amounted to 3860, where there are 2010 tweets labeled 1 and 1855 tweets labeled 0.

### 2.1.2 Preprocessing

Preprocessing, which involves several phases is the first step in categorizing text data. This stage aims to change the text into better data to provide good quality information applied in the subsequent step. This study employs numerous preprocessing techniques such as Stopword Removal, Cleaning, Case Folding, and Change Word.

Stopword removal is the process of eliminating superfluous words that don't add any value. As the example in the table 2.

Table 2. Example of Stopword Removal

Sentence Input	Sentence Output
Terus aja nahan ego sendiri, egonya gak pernah ditahan.	Terus nahan ego sendiri, egonya gak pernah ditahan.

Cleaning is cleaning data by removing symbols, numbers, punctuation marks, and characters not in the alphabet. As the example in the table 3.

Table 3. Example of Cleaning

Sentence Input	Sentence Output
Terus nahan ego sendiri, egonya gak pernah ditahan.	Terus nahan ego sendiri egonya gak pernah ditahan

To facilitate searching, a procedure known as "Case Folding" involves changing all of the dataset's letters to lowercase. As the example in the table 4.

Table 4. Example of Case Folding

Sentence Input	Sentence Output
Terus nahan ego sendiri egonya gak pernah ditahan	terus nahan ego sendiri egonya gak pernah ditahan

Change Word is a process to improve the meaning of a word or improve a short word into the original word, such as "yg" to "yang". As the example in the table 5.

Table 5. Example of Change Word

Sentence Input	Sentence Output
terus nahan ego sendiri egonya gak pernah ditahan	terus menahan ego sendiri egonya tidak pernah ditahan

Stemming entails converting all suffixes to base words. The table 6 displays an illustration of the outcomes of modifications to the preprocessing procedure. As the example in the table 6.



Table 6. Exsample of Stemming

Sentence Input	Sentence Output
terus menahan ego sendiri egonya tidak pernah ditahan	terus nahan ego sendiri ego tidak pernah tahan

2.1.3 BERT

BERT is a language representation model designed to train a two-way representation of unlabeled text by conditioning left and proper contexts at all layers. Bidirectional Encoder Representations from Transformers (BERT) optimizes Masked Language Model (MLM) and Next Sentence Prediction (NSP) in the Pre-Trained process. The masked Language Model (MLM) is a process model to predict the words that will appear from the previous comments. Next Sentence Prediction (NSP) is a binary classification loss that functions to indicate two words that follow each other in a text[17]. BERT was the first finetuning-based representation model to outperform multiple task architectures[18]. The training analysis of the BERT method explores and quantifies the options that are important for training the BERT model while maintaining the architectural model. This starts from the training of the BERT model with the same configuration as BERT base (training<sup>2</sup>L = 12, H= 768, A = 12, 110M params)[6].

2.1.4 RoBERTa

Robustly optimized BERT approach (RoBERTa) is a replicated version of the Pre-Trained approach of BERT that has been optimized where the system can predict parts of text in annotated languages with a total dataset of 160GB. RoBERTa was built based on the research of replication of BERT pre-training concerning the impact of hyperparameters and the size of the training used. RoBERTa can predict parts of text in annotated languages. RoBERTa has two pre-trained models called RoBERTa large and RoBERTa base. The design of the two models requires the configuration of the BERT architecture[6]. The analysis was carried out by applying the pre-training and fine-tuning processes. Pre-training aims to get a pre-trained RoBERTa model. Before conducting training, the thing that needs to be done is to set it first by entering the hyperparameters used, the tokens obtained, and the training data that has been created. Fine-tuning is obtained from the previous pre-training process, and a pre-trained model is used to perform the masking process. After doing data preprocessing which aims to get data for training, testing, and validating. Then the configuration is carried out using the hyperparameters used for the fine-tuning process[7]. Four essential parts make RoBERTa's performance can be optimized[6].

First, this method is trained using the Dynamic Masking architecture, which functions to generate a mask pattern every time you enter a sequence into the model. This is important when pre-training for more steps or with more meaningful data.

Second, eliminating Next Sentence Prediction (NSP), It asks users to consider how two sentences relate to one another in order to better the performance of tasks that come after it, including Natural Language Inference. In the research tested by Liu (2009), it was stated that using one sentence decreased downstream task performance, making RoBERTa use entire sentences without NSP[6].

Third, training with a large number of batches can increase optimization speed and end-task performance. The RoBERTa architecture has increased the batch size from 256 BERT sequences to 8,000 sequences.

Fourth, the more significant Byte-Pair Encoding (BPE), BPE is a combination of character and word representation that allows for handling large, standard vocabularies. In the BPE implementation, RoBERTa uses bytes instead of Unicode characters as the basis of subword units, enabling it to study subword vocabulary with a size of up to 50,000 units[6].

With the optimization that has been done, this method has several advantages, namely, creating periods with more batches of data. In addition, this method eliminates the Next Sentence Prediction (NSP) task on BERT[17], which allows RoBERTa to use Dynamic Masking resulting in better performance. RoBERTa has two pre-trained models, namely RoBERTa large and RoBERTa base(L). Both designs follow the configuration on the BERT architecture where RoBERTa large and RoBERTa base use a training model with an extensive BERT architecture configuration (training<sup>2</sup>L = 24, H = 1024, A = 16, 355M params). The following are the hyperparameters used in the pre-training and refinement of RoBERTa[6].

Table 7. Hyperparameters pre-training RoBERTa large dan RoBERTa base

Hyperparam	RoBERTa large	RoBERTa base
Number of Layers	24	12
Hidden size	1024	768
FFN inner hidden size	4096	3072
Attention heads	16	12
Attention head size	64	64
Dropout	0.1	0.1
Attention Dropout	0.1	0.1
Warmup Steps	30k	24k
Peak Learning Rate	4,00E-04	6,00E-04
Batch Size	8k	8k

Hyperparam	RoBERTa large	RoBERTa base
Weight Decay	0.01	0.01
Max Steps	500k	500k
Learning Rate Decay	Linear	Linear
Adam	1,00E-06	1,00E-06
Adam	0.9	0.9
Adam	0.98	0.98
Gradient Clipping	0.0	0.0

**Table 8.** Hyperparameters fine-tuning RoBERTa large dan RoBERTa base

Hyperparam	RACE	SQuAD	GLUE
Learning Rate	1,00E-05	1.5e-5	{1e-5, 2e-5, 3e-5}
Batch Size	16	48	{16, 32}
Weight Decay	0.1	0.01	0.1
Max Epochs	4	2	10
Learning Rate Decay	Linear	Linear	Linear
Warmup ratio	0.06	0.06	0.06

### 2.1.4 Evaluation

A table called the confusion matrix is used to define, illustrate, and draw conclusions about the performance of a classification method. The results of the classification are divided into four characteristics, namely True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). The performance of a matrix is measured based on accuracy, precision, recall, and f1score, which can be calculated based on TP, TN, FP, and FN[19]. Accuracy represents the ratio to be classified by the formula:

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

Precision is the value obtained from the accuracy of a class with the total number of predictions for that class. The purpose of precision is to see the percentage of the relevance of the classification results with the formula:

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

Recall is a value obtained from the prediction accuracy of a class with the total number of facts for that class. The formula can calculate recall:

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

F1-score is an evaluation calculation performed by combining both precision and recall values. The formula can calculate F1-score:

$$f1 - score = \frac{2 * (Recall * Precision)}{Recall + Precision} \tag{4}$$

## 3. RESULT AND DISCUSSION

This study is a text classification test to detect a person's level of depression based on Twitter tweets. There are 3860 datasets divided into two categories of depression and non-depression. The following are the stages of implementing solution method. At first, The dataset will then be initialized by the RoBERTa tokenizer, which functions to replace text data with algorithmically generated numbers, also known as tokens. After tokenization, the next step is to set the hyperparameters. Hyperparameters are parameters set before the model learning process is carried out. This is done by reading the columns in the dataset, then adjusting the hyperparameters. In this study, the hyperparameter setting uses a random search. This is done because it has the advantage of being able to control the number of parameter searches. Parameters will be divided into three, namely train data, validation data, and test data, with a total comparison of 90:10.

The next step is to classify the RoBERTa model. RoBERTa has two pre-trained models called RoBERTa large and RoBERTa base. The analysis was carried out by applying the pre-training and fine-tuning processes. This study used RoBERTa base as pre-trained. This stage aims to get pre-trained from RoBERTa before training the model. Pre-training is carried out with the number of epochs = 6. After the value of the pre-trained is successful, the next step is to start testing the training model with epochs = 12. In the pre-training and training stages of this model, the data used are train data and evaluation data.



After the results of the training model are obtained, the next step is to test the results with test data to get an evaluation value. This study uses a confusion matrix that produces values for accuracy, precision, recall, and f1-score. This stage is the final stage of testing. The following are the results of the classification text of the RoBERTa model using a dataset derived from Twitter tweets with categories of depression and not depression which can be seen in the table 9.

Table 9. Classification Report

Classification Report			
	Precision	Recall	F1-Score
1	0.6500	0.7800	0.7091
0	0.6901	0.5385	0.6049
Accuracy	<b>0.6649</b>		

Considering the test outcomes in the previous table, might therefore say that the use of data preprocessing techniques and data separation with a comparison of 90:10 produces an accuracy of 66%. These results are unsatisfactory, so several other scenarios are needed to increase the accuracy value.

Several scenarios used in this research focus on batch size and preprocessing stage. The first scenario was tested on a batch size distribution of 16 and 32. The goal of this scenario was to ascertain how batch size affected the testing model's accuracy. In the second scenario, testing of the preprocessing technique was carried out on the removal of stemming and stopwords. It aims to determine the best combination of preprocessing stages to determine the performance of the preprocessing.

### 3.1 Result and discussion of the batch size effect

Changing the batch size is used for testing in the first scenario. This parameter is part of the hyperparameter that needs to be tuned to get the optimal model. This can affect learning significantly. The batch sizes tested in this scheme are 16 and 32. The following are the results of the first scenario in the table 10.

Table 10. Effect of batch size

Batch Size	Accuracy
16	<b>0.6597</b>
32	0.6230

Based on the results from the table above, the best accuracy is shown by the batch size of 16. This is because using a large batch size can harm the network's accuracy during training, reducing the stochastic gradient descent. Using a large number of batches has the advantage of faster computation time but has the disadvantage of gradient accuracy. Therefore, a small batch size can provide good performance on the model even though the computation time can be longer.

### 3.2 Result and discussion of the preprocessing effect

The second instance, a phased preprocessing test was conducted, which examined how stopword removal and stemming affected the accuracy of the results. The parameters used are the best parameters obtained from the previous test scenario. Stemming is changing all affixes into essential words that can affect a sentence. Stopword Removal removes unimportant words that have no meaning and only have a negligible effect on a sentence. Here is the result of the first scenario in the table 11.

Table 11. The result of preprocessing

Preprocessing	Preprocessing Accuracy against Split Data ( Data train : Data test )	
	90:10:00	
No Stemming	0.6806	
No Stopword Removal	0.6979	
No Stemming and No Stopword Removal	<b>0.7188</b>	
Full Preprocessing	0.6597	

Considering the test outcomes in the previous table, it can be concluded that the preprocessing technique can produce different accuracy values. In the tests, the best accuracy value is 0.7188, or about 72%. The scenario tested without stemming and removing stopwords had a better accuracy rate than the other generating scenarios by dividing the dataset's training data by 90% and test data by 10%. Using the best accuracy value as a benchmark, the stemming and stopword removal processes cannot be applied to each word in a Twitter tweet. Stemming can change the form of words into essential words that can change the meaning of the sentence.



Meanwhile, stopword removal can remove text that does not affect the sentence. However, because fewer words are analyzed when using stopword removal, computing time can be sped up, Therefore, the algorithm will quickly process the words in a sentence. Therefore, both processes can reduce the value of classification accuracy.

## 4. CONCLUSION

Considering the results of the tests that have been done on using the Roberta categorization system to classify tweets on Twitter and tested using several scenarios described in the previous chapter. It can be concluded that the best accuracy is generated by the scenario of a batch size of 16 without using stemming and stopword removal with the use of 90% of training data and 10% of test data producing an accuracy value of 72%. Batch size can affect network accuracy during training. This can affect gradient descent. Stemming should be avoided because it will alter the meaning of the statement based on the Twitter tweet dataset that was utilized. In addition, the use of omitting stopwords in the literature also needs to be avoided because it can make sentences less information and reduce the accuracy value. Then, choosing a split data percentage can slightly improve accuracy. This is because the classification results decline with increasing test data amounts while they rise with decreasing test data amounts, the higher the results from the classification obtained. The reason is because of the data dissemination on test results, so that it produces different classification values. However, the classification results in this study are considered unsatisfactory because they cannot show the maximum accuracy value. This is because the dataset in this study could not be appropriately processed by the RoBERTa classification method. Therefore, in future research, a class review can be carried out on the dataset used, which aims to avoid data imbalance. In addition, a more significant amount of data can produce even better classification results considering that the RoBERTa model is a Deep Learning that can learn more data.

## REFERENCES

- [1] J. P. Jimenez, A. Botto, and P. Fonagy, *Etiopathogenic Theories and Models in Depression*. Springer International Publishing AG, 2021.
- [2] J. B. Nancy K. Baym, *Twitter: A Biography*. NYU Press, 2020.
- [3] M. Ardhi, "Analisis Sentimen untuk Deteksi Depresi pada Forum Reddit dengan Fitur Seleksi Chi-Square Menggunakan Support Vector Machine," 2020.
- [4] G. Mahendra, E. Sutoyo, O. N. Pratiwi, F. R. Industri, and U. Telkom, "MENDETEKSI GEJALA DEPRESI PENGGUNA TWITTER BERDASARKAN ANALISIS SENTIMEN MENGGUNAKAN ALGORITME K-NEAREST NEIGHBOR," 2020.
- [5] T. R. Ramadan, E. Sutoyo, and O. N. Pratiwi, "MENDETEKSI GEJALA DEPRESI PENGGUNA TWITTER MENGGUNAKAN ALGORITMA NAÏVE BAYES CLASSIFIER," pp. 1–8, 2020.
- [6] Y. Liu *et al.*, "RoBERTa: A Robustly Optimized BERT Pretraining Approach," no. 1, 2019, [Online]. Available: <http://arxiv.org/abs/1907.11692>.
- [7] B. Richardson, "IMPLEMENTASI INDOBERT-LITE DAN ROBERTA UNTUK TEXT MINING PADA APLIKASI CHATBOT JACOB," 2021.
- [8] I. Weber, M. W. Macy, and Y. Mejova, *Twitter: A Digital Socioscope*. New York: Cambridge University Press, 2015.
- [9] A. Hodgkin, *Following Searle on Twitter*. Chicago: University of Chicago Press, 2017.
- [10] L. Mandloi and R. Patel, "Twitter sentiments analysis using machine learning methods," *2020 Int. Conf. Emerg. Technol. INCET 2020*, pp. 1–5, 2020, doi: 10.1109/INCET49848.2020.9154183.
- [11] J. Degenhard, "Forecast of the number of Twitter users in Indonesia from 2017 to 2025," *Statista*, 2021. <https://www.statista.com/forecasts/1145550/twitter-users-in-indonesia>.
- [12] Ms. Dr. Namora LumonggaLubis, *DEPRESI Tinjauan Psikologi*, Edisi Pert. Jakarta: KENCANA, 2016.
- [13] F. Benamara, V. Moriceau, J. Mothe, F. Ramiandrisoa, and Z. He, "Automatic detection of depressive users in social media," *Conf. en Rech. d'Informations Appl. CORIA 2018 - 15th French Inf. Retr. Conf. Proc.*, 2020.
- [14] H. F. Sandmire, S. D. Austin, and R. C. Bechtel, "Depression and Other Common Mental Disorders," *Obstet. Gynecol.*, vol. 48, no. 1, pp. 56–60, 2017.
- [15] N. Syafitri, Y. Arta, A. Siswanto, and S. P. Rizki, "Expert System to Detect Early Depression in Adolescents using DASS 42," no. ICoSET 2019, pp. 211–218, 2020, doi: 10.5220/0009158202110218.
- [16] R. Widiana, Sumiharso., and R. M. Safitri, "Psychometric properties of internet-administered version of Depression, Anxiety and Stress Scales (DASS-42) in sample Indonesian adult," *Talent Dev. Excell.*, vol. 12, no. 2s, pp. 1422–1434, 2020, [Online]. Available: <http://www.iratde.com>.
- [17] X. Wang, S. Chen, T. Li, W. Li, Y. Zhou, and J. Zheng, "Depression Risk Prediction for Chinese Microblogs via Deep Learning Methods : Content Analysis," vol. 8, pp. 1–10, 2020, doi: 10.2196/17958.
- [18] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," *NAACL HLT 2019 - 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - Proc. Conf.*, vol. 1, no. Mlm, pp. 4171–4186, 2019.
- [19] P. Singh, N. Singh, K. K. Singh, and A. Singh, "Diagnosing of disease using machine learning," *Mach. Learn. Internet Med. Things Healthc.*, pp. 89–111, Jan. 2021, doi: 10.1016/B978-0-12-821229-5.00003-3.