Journal of Information System Research (JOSH)
Volume 3, No. 4, Juli 2022, pp 476−482
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v3i4.1885

Analyze Detection Depression In Social Media Twitter Using Bidirectional Encoder Representations from Transformers

Fikri Ilham*, Warih Maharani

Fakultas Informatika, Program Studi Informatika, Telkom University, Bandung
Jl. Telekomunikasi No. 1, Terusan Buahbatu - Bojongsoang, Sukapura, Kec. Dayeuhkolot, Kabupaten Bandung, Jawa Barat, Indonesia
Email: [1]fikriilham@student.telkomuniversity.ac.id
Email Penulis Korespondensi: fikriilham@student.telkomuniversity.ac.id
Submitted: **18/07/2022**; Accepted: **28/07/2022**; Published: **31/07/2022**

Abstrak−Kesehatan manusia merupakan bagian terpenting dari kesejahteraan suatu negara. Mendeteksi dinisuatu penyakit adalah sangat penting untuk mencegahnya suatu penyebaran dalam suatu wilayah. Media sosial kini menjadi perkembangan informasi yang pesat dan luas sehingga bisa memberikan kemudahan bagi masyarakat untuk melakukan komunikasi. Orang depresi memiliki berbagai macam gejala depresi dari setiap perilaku manusia. Dokter psikologis sering melakukan tatap muka wawancara pada diagnosis yang umum digunakan dan kriteria manual statistic gangguan jiwa. Depresi adalah gangguan mental yang umum muncul pada manusia dengan ciri-cirinya yaitu suasana hati yang tertekan, kehilangan minat dan kesenangan, energy tubuh yang tidak stabil, dan konsentrasi yang buruk. Dalam melakukan penelitian ini bertujuan untuk mendeteksi orang yang mengalami depresi dengan menggunakan metode BERT (Bidirentional Encoder Representations from Transformers) yang berbasis Machine Learning. BERT bisa mengklasifikasi secara biner teks di media sosial yaitu Twitter yang mengandung deteksi Depresi. Berdasarkan pengujian yang telah dilakukan didapatkan nilai akurasi terbaik sebesar 0,7176 atau 71%.

Kata Kunci: Penyakit Mental; Depresi; Twitter; BERT

Abstract−Human health is an essential part of the welfare of a country. Early detection of a disease is necessary to prevent it from spreading in an area. Social media is now a rapid and widespread development of information to provide convenience for the public to communicate. Depressed people have a variety of depressive symptoms from every human behaviour. Psychological doctors often conduct face-to-face interviews on commonly used diagnoses and statistical manual criteria for mental disorders. Depression is a mental disorder that typically appears in humans with the characteristics of depressed mood, loss of interest and pleasure, unstable body energy, and poor concentration. In conducting this research, the aim is to detect people who are depressed by using the Machine Learning-based BERT (Bidirectional Encoder Representations from Transformers) method. BERT can binarily classify text on social media, namely Twitter, which contains Depression detection. Based on the tests that have been carried out, the best accuracy value is 0.7176 or 71%.

Keywords: Mental Illness; Depression; Twitter; BERT

1. INTRODUCTION

Human health is an essential part of the well-being of a country. Early detection of a disease is necessary to prevent its distribution in an area. With the existence of social media mong In today's society, social media plays an essential role in detecting diseases in the form of depression through social media Twitter. Indonesia is one example where the different lives and their opinions about the present, In a study conducted by TMoslmi 2017 [1].

Social media is now a rapid and widespread information development, making it easier for people to communicate. In 2018 internet service users in Indonesia reached 171.17 million users compared to 2017, a significant increase of 143.26 million users, In a study conducted by G.Mahendra 2021[2]. Users use Twitter to express their emotions. Users of Twitter accounts used to communicate on Twitter, including Tweets. There's a Tweet quoted from Copernicus, with basically nothing opening. Tweets usually contain short messages, messages, statemen, ts and, in some cases, links to articles, blog spots, podcasts, or videos, In a study conducted by M.I.Maulana 2019[3]—Depression is the leading cause of disability worldwide. Depressed people have various kinds of depressive symptoms from every human behaviour. Psychological doctors often conduct face-to-face interviews on commonly used diagnoses and mental disorder statistics manual criteria. Globally it is estimated that 350 million people age staggering from depression, In a study conducted by G.Shen 2017 [4]. Depression is a common mental disorder that appears in humans with its characteristics, namely depressed mood, loss of interest and pleasure, unstable body energy, and poor concentration, In a study conducted by T.R.Ramadan 2021[5].

Comparison of a method with each other, namely the Naïve Bayes Classifier, is used to calculate the probability that the proposed data is correct. Still, Naïve Bayes it is not possible to measure the accuracy of the prediction. In addition, the naive Bayesian method has weaknesses in feature selection, In a study conducted by A.Pattekari 2019 [6], from Existing research using the Naïve Bayes method has resulted in 3049 total depression, and 15705 numbers do not show depression, In a study conducted by T.R.Ramadan 2021[5]. While Support Vector Machine(SVM) can only handle the classification of two classes, In a study conducted by F.Zikra 2021 [7], from research that has been done already exist SVM only produce a classification accuracy is 85.38% and also obtained 16 of the global burden of disease and injury for people aged 10 to 19 years experiencing mental disorders (Trisni Handayani, Dian Ayubi, 2020)[8]. As a result of the accuracy that has been studied in existing research, the SVM method is better than the Naïve Bayes method. The method used in this study is the BERT method (Bidirentional Machine Learning-based Encoder Representations from Transformers developed in 2018, the BERT method has

Journal of Information System Research (JOSH)
Volume 3, No. 4, Juli 2022, pp 476−482
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v3i4.1885

been developed with large Data so that BERT can handle various tasks in its field, namely NLP. This study aims to determine BERT's performance in classifying crawled tweets using acceptable adjustment strategies, In a study conducted by Y.Ajitama 2021 [9]. BERT can classify text binary on social media, namely Twitter, which contains Depression detection. Depression Detection in question is a Twitter account user currently experiencing symptoms of depression and then disclosed on social media, namely Twitter, In a study conducted by I.Prapitasari 2019[10]. BERT method can also read a long text well. Architecture The transformer uses fewer parameters than the CNN model to produce a good performance in a short time, In a study conducted by A.Khan 2020 [11].

2. RESEARCH METHODOLODY

2.1 System Design

At this stage, to carry out Depression Detection, which takes data from on Twitter social media, build a system with the flow that has been made. In making this classical system there are several stages consisting of data retrieval of datasets, preprocessing, data separation, training, testing, modeling, and evaluation. The following is a flow chart using the BERT method can be seen in **Figure 1**.

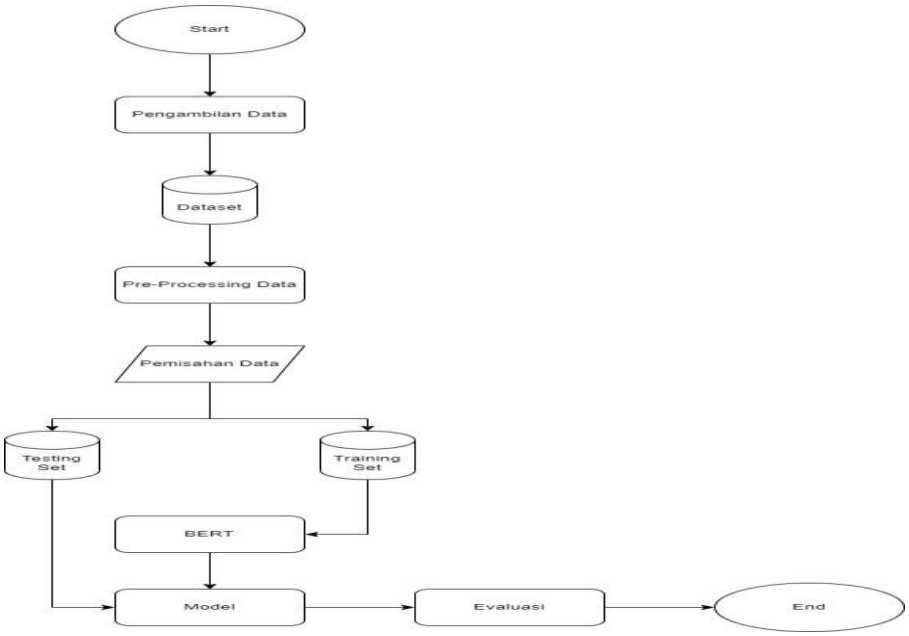

Figure 1. Stages of the system to be built BERT

2.1.2 Dataset

In the Data Collection step there are several stages In the first stage, respondents fill out a questionnaire form containing a Depression Detection using the DASS-42 assessment, the second stage is crawling on Twitter tweets taken from respondents' accounts who have filled out the questionnaire form. Tweets on Indonesian language twitter, there are 3867 datasets. The dataset contains two columns containing labels and tweets, and label 0 contains tweets containing elements of depression. In contrast, label 1 contains tweets that do not include aspects of depression, while tweets contain tweets that have been crawling. When doing the labelling is done manually.

2.2 Pre-Processing Data

Processing Data has several stages of the process, such as case folding, cleaning, stopword, stemming and tokenizing. In this process, it functions to tidy up data containing user respondents to make it easier to process at a later stage. Here are the procedures:

2.2.1 Case Folding

This process aims to convert all letters/words in the existing data into lowercase letters so that the data used becomes the same form, Case Folding can be seen in **Table 1**.

Tabel 1. Case Folding

Sentence Input	Sentence Output
Saya cuman Bisa Terpuruk dikmr krn Depresi	saya cuman bisa terpuruk dikamar karena depresi

Journal of Information System Research (JOSH)
Volume 3, No. 4, Juli 2022, pp 476−482
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v3i4.1885

2.2.2 Cleaning

This process aims to clean up existing data into a form that is easier to process to eliminate redundant symbols, numbers, punctuation marks and spaces, Cleaning can be seen in **Table 2**.

Tabel 2. Cleaning

Sentence Input	Sentence Output
Saya cuman bisa terpuruk dikamar karena depresi!!</3	saya cuman bisa terpuruk dikamar karena depresi

2.2.3 Stopword

This process aims to remove words that do not need to be entered but first see whether these words affect the data, Stopward can be seen in **Table 3**.

Tabel 3. Stopward

Sentence Input	Sentence Output
saya	terpuruk
cuman	kamar
bisa	depresi
terpuruk	
di	
kamar	
karena	
depresi	

2.2.4 Stemming

In this process, the aim is to remove affixes at the beginning and the end so that the goal is to avoid ambiguity in the data, Stemming can be seen in **Table 4**.

Tabel 4. Stemming

Sentence Input	Sentence Output
terpuruk	puruk
depresi	depresi
kamar	kamar

2.3 DASS-42

DASS is a 42-item, 3-item questionnaire designed to measure negative emotional states of depression, anxiety, and stress. The score for each respondent on each subscale was then graded according to severity. The severity of Depression pressure can be measured by the presence of DASS-42, In a study conducted by N.Syafitri 2020 [12]. This research is a data collection system based on the results of Twitter tweets that have been filled out in the DASS-42 respondent form. The data that has been collected is 15 accounts of Depression detected on Twitter with ratings of Very Severe, Severe, Moderate and Heavy. This data collection was carried out on July 25, 2021. The following is an example of a table which contains tweets from Twitter accounts with depression levels, In a study conducted by M.K.Neighbor 2021[13].

2.4 Twitter

Witter is one of the social media with total users. In 2018 there were many Twitter users (Wearesocial, 2015). The number of Twitter users in 2018 reached 330 million. Users use Twitter to express their emotions. Users of Twitter accounts used to communicate on Twitter, including Tweets (Maulina, 2015). There is a Tweet quoted from Copernicus with essentially no preamble (Copernicus, 2015). Tweets usually contain a short message, message, statement and in some cases, a link to an article, blog spot, podcast or video. The length of the character before is only 140 characters, but after 444, it increases to 280 characters, In a study conducted by M.I.Maulana 2019[3].

2.5 BERT

BERT (Bidirectional Encoder Representations from Transformers) is a language using a fine-tuning approach. BERT pre-trains in an unsupervised way by looking at the left and right context conditions simultaneously in each layer, In a study conducted by Y.Ajitama 2021[9]. BERT uses a new technique called Masked Language Modeling, which shows two-way training in the model. The state transformer consists of two mechanisms: an encoder that reads the input and a decoder that predicts the task, In a study conducted by EP.T.Kerja[14].

Journal of Information System Research (JOSH)
Volume 3, No. 4, Juli 2022, pp 476−482
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v3i4.1885

BERT only requires an encoder, utilizes the attention principle encoder and reads the entire text as input. BERT can build contextual relationships for each token well, In a study conducted by J.Devlin 2019[15]. BERT method can also read a long text well. The Transformer architecture uses fewer parameters than the CNN model, so it can produce good performance in a short time, In a study conducted by A.Khan 2020 [11].

2.5.1 Input BERT

BERT uses WordPiece embedding, which contains 30,000 syllables. At the beginning of each sequence, there is a CLS token. For each sentence is placed the SEP token. Then add embedding on each token to distinguish whether the token is included in the embedding segment. The position embed is the result of this sum with the same dimensions as the embedding token. In a study conducted by F.A.Pratama 2020[16]. To calculate embedding can be seen in Formula **(1) (2)**.

$$\vec{p}2k = \sin \frac{1}{10000^{2}k/d} t \quad (1)$$

$$\vec{p}2k + 1 = \cos \frac{1}{10000^{2}k/d} t \quad (2)$$

Each embedding has dimensions of 768 can be seen in Figure 2.

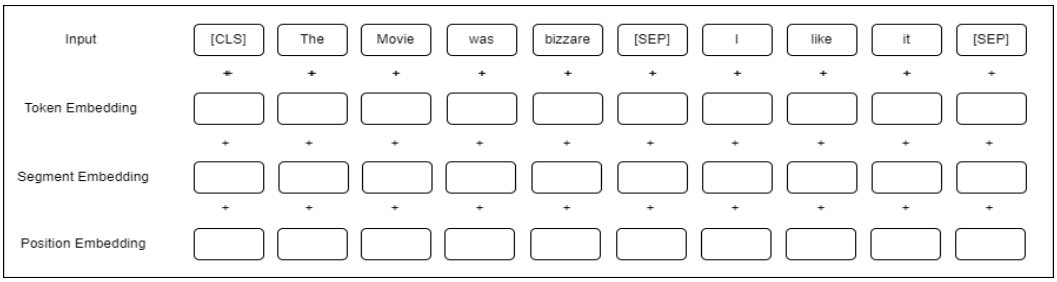

Figure 2. Input BERT

2.5.2 BERT Pre-training dan Fine Tuning

In the pre-training process, two tasks will be carried out by BERT, namely Masked Language Model (MLM) and Next Sentence Prediction (NSP). The function of MLM is so that the BERT method can simultaneously combine context from left and right. Then MLM masks some input tokens to predict the original vocabulary id of the masked vocabulary [15]. In comparison, NSP is so that the model can understand the relationship between two sentences. The BERT method is given two inputs, namely in the form of phrase pairs and a learning model for its use to identify whether the sentence is the following sentence in the original document. The input process is in the form of sentences A and B as an example for the pre-train, 50% probability that B is the following sentence from A (IsNext), and the other 50% is a random sentence from the corpus (NotNext). BERT uses pre-training data from BookCorpus, which contains 800 million words and from English Wikipedia, which has 2.5 billion words[15]. After encoding the data, the data that has been inputted into the BERT model is ready for fine-tuning. There are several strategies in doing fine-tuning, including [9]:

1. Text length
   In this study, we will use tweet data so that the sequence length will not exceed 512 due to the limited size of the tweet.
2. Layer selection
   In this study, an effective layer is needed to classify it.
3. Overfitting Problem
   In this study, overfitting chose the BERT optimizer with a reasonable learning rate.

2.6 BERT Model

BERT is a pre-trained example that has been trained using a large amount of data so that it has suitable parameters for use in tasks related to language understanding, one of which is sentiment analysis. For BERT to be used for sentiment analysis, it is hoped that the additional layer results can handle classification tasks. After that, fine-tuning was done using a dataset related to sentiment analysis in this final project in the form of a tweet related to the selection of depression detection. This analysis uses a BERT-base sample containing 110 million parameters using 12 layers, 768 hidden sizes and 12 attention heads. BERT only requires an encoder, utilizes the attention principle encoder and reads the entire text as input

2.7 Matriks Evaluasi

In this Depression Detection study, Scikit-Learn is used for its non-paid machine learning library software for python programming. This stage describes the Machine Learning process on models, data, algorithms and

Journal of Information System Research (JOSH)
Volume 3, No. 4, Juli 2022, pp 476−482
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v3i4.1885

functions. But the calculation process and sentiment analysis cannot be visualized in detail. For this stage, the metrics used as benchmarks for the results of the calculation of sentiment analysis data are:

a) F1 Score
b) Precision
c) Recall
d) Accuracy

F1 score is the average of precision and recall. Here is the formula:

$$F1 = 2.\frac{precision.recall}{precision+recall} \quad (3)$$

Precision is the ratio of correctly predicted positive observations to the total number of predicted observations. Here is the formula:

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

Recall is the ratio of positive observations that will be correctly predicted on actual class observations. Here's the formula:

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

Accuracy is the intuitive performance of the ratio of observations that will be correctly predicted on the total observations. Here is the formula:

$$Akurasi = \frac{TN+TP}{TP+FN+TN+FN} \quad (6)$$

3. RESULT AND DISCUSSION

In the text classification test, tweets on Indonesian language twitter, there are 3867 datasets. The dataset contains two columns containing labels and tweets, and label 0 contains tweets containing elements of depression. In contrast, label 1 contains tweets that do not include aspects of depression, while tweets contain tweets that have been crawling. When doing the labelling is done manually. On Test Scenario This final project focuses on the preprocessing stage and testing a BERT method. The first scenario is to do a test by changing the number of sample batch sizes based on the Neural Network, which aims: to determine the best classification for this test. The second scenario is to test split data which seeks to find the best performance on test data and train data from various data.

3.1 Results and discussion of the effect of Classification

In Stages classification, first, the input data obtained is converted into input form BERT can read. To be read by BERT, it is necessary to add the token [CLS] at the beginning of the sentence and pass [SEP] at the end of the sentence and determines the length of the sentence to add as many [PAD] token as remaining tokens. After [CLS], [SEP], and [PAD] tokens are added, BERT will convert each word token into ids token and returns input_ids, and attention_mask results for later will be passed into the BERT model. After converting data into an input that BERT can read, it will make the creation of a data loader to help speed up data retrieval. After creating the data loader, the step Next is modelling. After the model is created, it will be training and evaluation of the model that has been made, classification can be seen in **Figure 3**.

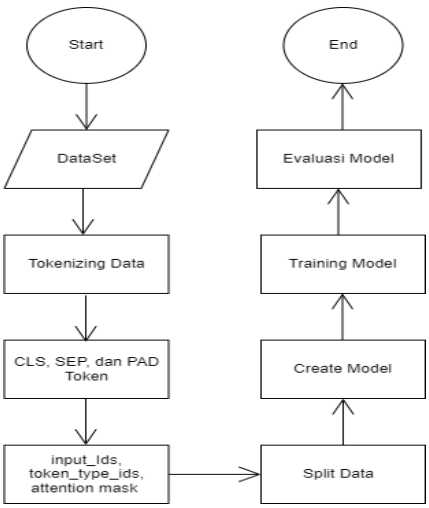

Figure 3. Stage of Klasifikasi

Journal of Information System Research (JOSH)
Volume 3, No. 4, Juli 2022, pp 476−482
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v3i4.1885

3.2 Results and discussion of the effect of Split Data

The first scenario, testing is done by changing the split data. The data is divided into two, namely train data and test data. data train serves to train the algorithm in finding the appropriate model in this test, while test data serves to test or determine the performance of the model to be obtained in this test. In this test, BERT classification can be used to determine whether or not split data has an effect on the accuracy value. The first scenario results can be seen in **Table 5**.

Tabel 5. Effect Split Data

Split Data (Data train : Data test)	Classification Accuracy BERT
60:40:00	0.6591
70:30:00	0.6883
80:20:00	0.6710
90:10:00	**0.7176**

The results of the tests that have been carried out are that each data split produces a different accuracy value can be seen in **Table 5**. In the above test, it can be seen that the train data and test data with a comparison 90:10 is the best accuracy value, which is 0.7176 or 71%, so it can be said that there is a lot of train data then the accuracy value is getting better.

3.3 Results and discussion of the effect of Batch Size

The second scenario, perform the test by changing the batch size. there are different batch size values, namely 16, 32 and 64, Batch size is a number of sample data values distributed to the Neural Network. In the first scenario doing testing changing the split data to 90:10 is the best value for accuracy. For this test by changing the batch size to 16, 32 and 64, the second scenario results can be seen in **table 6.**

Tabel 6. Effect of Batch Size

Batch Size	Classification Accuracy BERT Data 90:10
16	0.6943
32	0.6580
64	**0.7172**

The results of the tests that have been carried out in **Table 6**, each batch size has different values using split data 90:10 batch size 64 is the best value for this test. Based on Experiment above, it can be said that the batch size affects the increase or decrease in the accuracy value.

4. CONCLUSION

After the tester is done scenario testing that has do for the detection of depression on tweets on Indonesian-language twitter using the BERT method, it can be concluded that the best value results are 0.7176 or 71% with split data train data and test data 90:10 (90% train data and 10% test data) with batch size 64 and epoch 4. The best system performance is the result of sharing data between training data and test data with a total data sharing of 90:10, System performance in sharing data using the classification BERT method produces an accuracy of 0.7176 or 71% . In this test, the higher the train data, the better the accuracy value, while the more test data, the lower the accuracy value and the more batch sizes, the better the accuracy value. The evaluation and analysis stage using a confusion matrix with four combinations, namely True Negative (TN), True Positive (TP), False Negative (FN), and False Positive (FP) from these combinations get the calculation results for Accuracy, Precision, Recall and F1- Score. From the value that has been obtained will be a test. The following is a test set of values that have been obtained, Accuracy 71%, Precision 81%, Recall 71% dan F1-score 75%. For further research using the BERT method, more datasets should be needed so that the accuracy value obtained is much better.

REFERENCES

[1] T. Al-Moslmi, N. Omar, S. Abdullah, and M. Albared, “Approaches to Cross-Domain Sentiment Analysis: A Systematic Literature Review,” *IEEE Access*, vol. 5, pp. 16173–16192, 2017, doi: 10.1109/ACCESS.2017.2690342.

[2] 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 SENTIMENT ANALYSIS FOR MEASURING ENGAGEMENT ACCOUNT.” 2021.

[3] M. I. Maulana and A. A. Soebroto, “Klasifikasi Tingkat Stres Berdasarkan Tweet pada Akun Twitter menggunakan

Journal of Information System Research (JOSH)
Volume 3, No. 4, Juli 2022, pp 476−482
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v3i4.1885

Metode Improved k-Nearest Neighbor dan Seleksi Fitur Chi- square," vol. 3, no. 7, pp. 6662–6669, 2019. Available: http://download.garuda.kemdikbud.go.id/article.php?article=1479851&val=10384&title=Klasifikasi%20Tingkat%20Stres%20Berdasarkan%20Tweet%20pada%20Akun%20Twitter%20menggunakan%20Metode%20Improved%20k-Nearest%20Neighbor%20dan%20Seleksi%20Fitur%20Chi-square

[4] G. Shen *et al.*, "Depression detection via harvesting social media: A multimodal dictionary learning solution," *IJCAI Int. Jt. Conf. Artif. Intell.*, vol. 0, pp. 3838–3844, 2017, doi: 10.24963/ijcai.2017/536.

[5] T. R. Ramadan, E. Sutoyo, and O. N. Pratiwi, "MENDETEKSI GEJALA DEPRESI PENGGUNA TWITTER MENGGUNAKAN ALGORITMA NAÏVE BAYES CLASSIFIER DETECTING USER TWITTER DEPRESSION SYMPTOMS USING NAÏVE BAYES CLASSIFIER ALGORITHM," pp. 1–8. 2021

[6] A. Pattekari, S.A.; Parveen, "Prediction system for heart disease using Naïve Bayes," *Int. J. Adv. Comput. Math. Sci.*, vol. 3, no. 3, pp. 290–294, 2012, doi=10.1.1.1089.1654&rep=rep1&type=pdf

[7] F. Zikra, R. Patmasari, F. T. Elektro, and U. Telkom, "DETEKSI PENYAKIT CABAI BERDASARKAN CITRA DAUN MENGGUNAKAN METODE GRAY LEVEL CO-OCCURENCE MATRIX (GLCM) DAN SUPPORT VECTOR MACHINE (SVM) CHILI DISEASE DETECTION BASED ON LEAF IMAGE USING GRAY LEVEL CO-OCCURRENCE MATRIX (GLCM) AND SUPPORT VECTOR MACHI." Available: https://jurnal.darmajaya.ac.id/index.php/PSND/article/download/2920/1243

[8] L. Saletti-cuesta *et al.*, "No 主観的健康感を中心とした在宅高齢者における 健康関連指標に関する共分散構造分析Title," *Sustain.*, vol. 4, no. 1, pp. 1–9, 2020, [Online]. Available: https://pesquisa.bvsalud.org/portal/resource/en/mdl-20203177951%0Ahttp://dx.doi.org/10.1038/s41562-020-0887-9%0Ahttp://dx.doi.org/10.1038/s41562-020-0884-z%0Ahttps://doi.org/10.1080/13669877.2020.1758193%0Ahttp://sersc.org/journals/index.php/IJAST/article.

[9] Y. Ajitama, S. S. Prasetiyowati, and Y. Sibaroni, "Analisis Sentimen Terhadap Tweet Mengenai Pemilihan Presiden Amerika Serikat Tahun 2020 Menggunakan Metode BERT," 2021.

[10] I. Prapitasari *et al.*, "Bab Iii Metodologi Penelitian," pp. 62–76, 2019, [Online]. Available: https://www.neliti.com/publications/275650/elderly-visit-routines-to-elderly-integrated-service-post-in-the-working-area-of%0Ahttps://ejournal3.undip.ac.id/index.php/jkm/article/view/16353%0Ahttp://repository.unjaya.ac.id/id/eprint/3305.

[11] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, *A survey of the recent architectures of deep convolutional neural networks*, vol. 53, no. 8. Springer Netherlands, 2020, https://doi.org/10.1007/s10462-020-09825-6

[12] 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.

[13] M. K. Neighbor, A. P. Tirtopangarsa, W. Maharani, T. Informasi, and U. Telkom, "Sentiment Analysis of Depression Detection on Twitter Social Media Users Using the K-Nearest Neighbor Method," pp. 247–258, 2021. Available : http://103.23.20.161/index.php/semnasif/article/viewFile/6076/3935

[14] E. P. T. Kerja, "済無No Title No Title No Title," *Angew. Chemie Int. Ed. 6(11), 951–952.*, vol. 13, no. April, pp. 15–38, 1967.

[15] 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, https://doi.org/10.48550/arXiv.1810.04805

[16] F. A. Pratama and A. Romadhony, "Identifikasi Komentar Toksik Dengan BERT," vol. 7, no. 2, pp. 1–9, 2020. Available : https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/viewFile/13073/12728