OPEN ACCESS 41

ISSN 2460-9056
http://socj.telkomuniversity.ac.id/ijoict/

Intl. Journal on ICT
Vol. 8, No. 1, June 2022. pp.41-51
doi:10.21108/ijoict.v8i1.634

General Depression Detection Analysis Using IndoBERT Method

Ilham Rizki Hidayat[1*], Warih Maharani [2]

[1,2]*School of Computing, Telkom University*
Jl. Telekomunikasi No. 1 Terusan Buahbatu, Bandung, Jawa Barat, Indonesia 40257

[1]ilhamrizkihidayat@student.telkomuniversity.ac.id, [2]wmaharani@telkomuniversity.ac.id

Abstract

Many of the tweets we find on Twitter are about feelings of depression caused by various things and the number is increasing. To be able to decide how depressed a user is, analyzing the tweets of a user can facilitate this and help to provide applicable treatment for users who are detected to be depressed. In this paper, the analyzed users are users who have more than 1000 tweets which are in Indonesian. Then, crawling / retrieval of user tweet data is carried out. After that, data pre-processing is carried out. After that, classification of the data obtained using the IndoBERT method and the model then provides an accuracy value of this detection analysis using the IndoBERT method with an accuracy value of 51% and F1-Score of 31%.

Keywords: IndoBERT, Depression, Analysis, Twitter, Social Media

I. INTRODUCTION

Nowadays, the development of technology continues to grow with no limitation. What we want to do, we can do with one or two clicks on the Internet. This conjointly applies once communicating or sharing something in a platform on the Internet, like on social media. Twitter is one of the most popular social networks in the world. According to data published by Mansoor Iqbal on BusinessofApps, the quantity of Twitter users in 2020 had reached 186 million, compared to 139 million in 2019[1].

Social media has become a place for everybody to share what they are going through on any given day, particularly once they feel stressed or exhausted from work or their daily activities. Twitter is additionally often a place to vent regarding life's problems. We can find this out by using the hashtag #Depression, #Depressed or #Depresi to point out that the user writing the tweet is feeling depressed and the number is not small. Per a survey report published by the Labor Force Survey (LFS) within the United Kingdom, there were 828,000 cases of work-related depression, stress and anxiety in 2019/20. This suggests that 2,440 out of 100,000 workers experience such cases out of a total of around 17.9 million workers[2]. In Indonesia, the number of depression cases, particularly throughout the COVID-19 pandemic era, is no less. Based on the results of a survey conducted by the Indonesian Psychiatric Association (PDSKJI), out of 1,522 respondents, 68% of respondents felt anxious, 67% were depressed and 77% admitted to experiencing psychological trauma[3].

Depression is a mental health disorder characterized by a persistent sad and depressed mood and loss of interest in activities, resulting in a decrease in the quality of daily life. Depression testing is necessary, for example, that companies can provide their workers with appropriate workloads and the workers themselves can better understand what is happening to themselves. One of the depression tests that can be taken is the Dass-42 (Depression Anxiety and Stress Scale)[4]. Dass-42 is a depression screening test where the person has to fill

Received on 4 July 2022. Revised on 25 July 2022. Accepted and Published on 27 July 2022.

out 42 questions/statements to determine what the person's level of depression is[4]. Based on the results obtained, this can help doctors to help the person overcome their depression.

Several previous studies have conducted depression detection with numerous methods [5][10]. One of them is research conducted by Md. Rafiqul Islam, et al [5] who used numerous methods such as Decision Tree, k-Nearest Neighbor, Support Vector Machine (SVM) and Ensemble using comments taken from Facebook. From this study, it was found that comments from Facebook can be classified based on LIWC. In the results, the utilization of the SVM method gets the biggest F-Measure results in the use of all features, that is between 0.63 to 0.73. The weakness in this research is that the SVM, which got the best results in this research, only separates the data into two different classes with a large separation. This will be an obstacle if we are going to classify it into more than two different classes. There is also a research conducted by Sergio G. Burdisso, et al[10] who used a method called SS3 which they designed to conduct this research. The results obtained using this SS3 method are reaching a value of 0.61 in the F-Measure results which indicates that the method is not better or the same as previous research using the SVM method.

Some studies also attach weaknesses that occur in using Deep Learning for sentiment analysis. This study was conducted by Lei Zhang, et al[13] where this study used the Bag-of-Words model (BoW) to perform sentiment classification at the document level, where this classification was to determine the sentiment value of an entire document consisting of many sentences. This BoW model had several weaknesses, where this model could experience data sparsity and high dimensionality problems. This means that the BoW model ignored the order of the words, resulting in two or more documents can have the same sentiment value if they had the same words, regardless of the order in which they appeared even though they might have different meanings.

One of the new methods developed that can be used to detect symptoms of depression on social media is IndoBERT[9]. IndoBERT is the Indonesian variant of a pre-trained model developed by a number of researchers from Google AI Language[6], namely BERT. BERT itself can read text at once so that it can be used for sequential data. IndoBERT has a train model with more than 220 million words obtained from several main sources [7]. IndoBERT can also be used to conduct sentiment analysis. In a research conducted by Fajri Koto, et al [8], it was found that sentiment analysis using the IndoBERT method produced a fairly high F-Measure value compared to previous studies, which was 0.84.

Pre-processing can also be a problem in sentiment analysis research. In the research conducted by Md. Rafiqul Islam, et al [5], it was mentioned that after getting raw data from Facebook, the data was immediately analysed using LIWC Software and feature extraction was carried out before the learning process. In the research conducted by Sergio G. Burdisso [10], publicly available data was directly implemented in the proposed SS3 framework. Meanwhile, the research that introduced IndoBERT [8] itself mentioned that the used data had been pre-processed but it did not mention what the pre-processing that has been done. So, it cannot be concluded what kind of pre-processing can help in conducting sentiment analysis. This is also the reason why the author proposed the IndoBERT method because the author wanted to test whether pre-processing would affect the results when using IndoBERT or not. In addition, because the focus of the data used is mostly in Indonesian, this research can also test the depth of the Indonesian corpus that has been owned by IndoBERT.

The purpose of this study is to conduct depression detection based on user activities and profiles on Twitter using the IndoBERT model and DASS-42 as a depression assessment scale. Twitter users using the IndoBERT model and DASS-42 as a depression rating scale. The results of using The results of using the model will then be evaluated using a confusion matrix to find the value of accuracy, precision, recall and F1-Score values.. The remainder of this journal is structured as follows. Section 2 is the explanation of related studies and definitions of the methods used. Section 3 is the explanation of the method or system built for the research. Section 4 is the presentation of research and evaluation results. Section 5 is the conclusions and suggestions based on this research.

II. LITERATURE REVIEW

Lovibond PF and Lovibond SH have developed the Depression, Anxiety and Stress Scale (DASS-42) to define, understand and measure the levels of these three negative emotional states[12]. The DASS-42 is a measure of depression, anxiety and stress consisting of a 42-point questionnaire containing 14 questions each to assess depression, anxiety and stress. Measurement using the DASS-42 has been shown to be highly internally consistent and provide meaningful results in many settings.

One studies using their self-designed method called SS3 to detect depression in a text on social media[9]. This method obtained an F1-Score matrix value of 0.55 and is the same value when compared to the use of the SVM method on the same dataset. SS3 is superior because it takes less time, which is 3.7m compared to SVM of 73.9m. In addition, paper [5] is to detect depression in Facebook comments and tweets on Twitter. This dataset is used because it provides a variety of posts that can be analyzed and is directly related to the users of each social media account. This research uses several Machine Learning groups such as SVM, Decision Tree, k-Nearest Neighbor, and Ensemble. The results of this study show that the SVM method group provides the best F1-Score matrix results using All Features, with a value range of 0.63 to 0.73.

However, not all Deep Learning method is good to use for the sentiment analysis. Refer to Lei Zhang, et al[13], one of the most commonly used models, the BoW model, has weaknesses. If two or more documents have some of the same words, then the sentiment value generated is more likely to be the same, even though the order in which the words appear is different. This can be limiting, as the order of the words may affect the actual meaning of the sentence or document.

In paper [6], it is explained that BERT (Bidirectional Encoder Representation from Transformers) is a transformer-based technique for NLP first developed by Jacob Devlin and his colleagues at Google and published in 2018. BERT is used to handle bidirectional representation of anonymized text in a way that the left and right sides are combined into a context in all sections. Making small changes to the existing BERT model can provide answers to many problems. BERT excels in simplicity and good observability. This is also why BERT can work and be understood in 11 programming languages, resulting in a GLUE score of 80.5%, MultiNLI accuracy of 86.7%, SQuAD v1.1 F1 test of 93.2 and SQuAD v2.0 F1 test of 83.1. The result of standardized data can be seen in Fig.1.

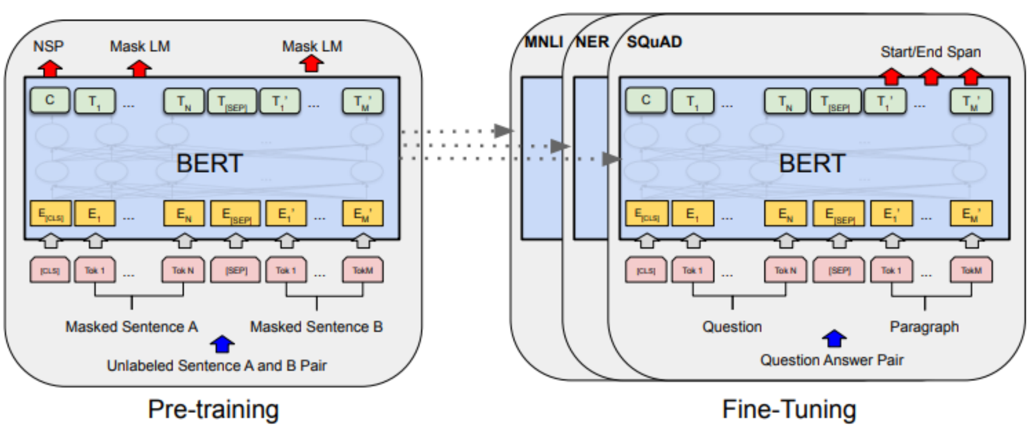

Fig. 1. Pre-Training & Fine-Tuning Model BERT[6]

In the Pre-Trained process, BERT uses two unsupervised tasks as shown in Figure 1. The first process is called Masked LM, which is a process where the model uses other words in the surrounding context to try to predict the [MASK] word. The model is trained by assigning [MASK] to a random percentage of input tokens, and then predicting the [MASK] tokens. In [6], a mask was applied to 15% of all randomly consecutive Word Piece tokens. The drawback of this model is that it may cause a mismatch between pre-training and fine-tuning because the [MASK] token does not appear during fine-tuning. This can be overcome by not always replacing masked words with actual [MASK] tokens. Instead, 80% are replaced with [MASK] tokens, 10% are replaced with arbitrary words and the other 10% are not changed at all [6].

The next process is Next Sentence Prediction (NSP), where the model receives a pair of sentences as input and learns them to predict whether the second sentence is the next sentence in the actual document. In the training process [6], 50% of the inputs are pairs where the second sentence is the next sentence in the original document, while the other 50% are arbitrary sentences from the corpus selected as the second sentence. The assumption is that the arbitrary sentence will be disconnected from the first sentence [6]. Figure 2 shows a representation of the input process performed on the BERT model. The result of standardized data can be seen in Fig.2.

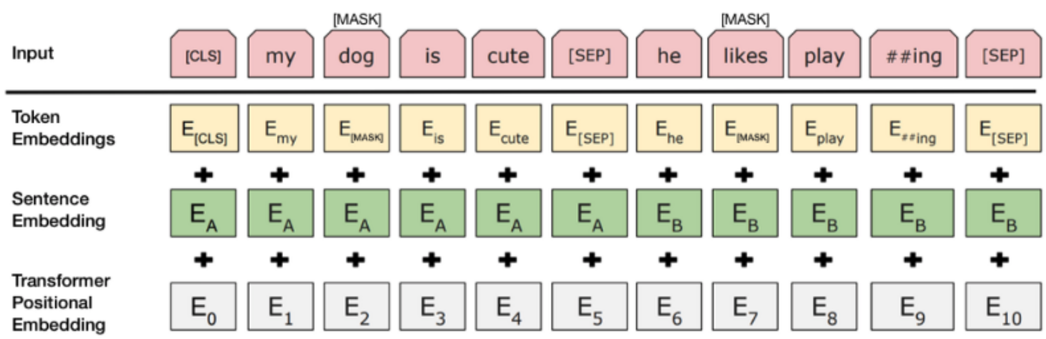

Fig. 2. BERT Input Representation[6]

In paper [7], this research aims to show the performance that can be generated in the use of several tasks using the IndoBERT model. IndoBERT itself is the result of the Indonesian variant and modification of the BERT method that was previously formed in 2018 by a number of researchers from Google AI Language and has been applied to predict the next sentence in the search column on Google itself. In the Sentiment Analysis task, the use of the IndoBERT method gets an F1-Score matrix value of 84.13. This value is higher than the use of other methods on the same dataset, namely Naïve Bayes, Logistic Regression, BiLSTM w/ fast Text, MBERT, and MalayBERT.

III. RESEARCH METHOD

In this study, the system built can detect symptoms of depression in tweet data from Twitter. There are several stages in this detection system as can be seen in Figure 3, namely dataset, *pre-processing* data, *splitting* data, model building, and analysis of the model performance results.

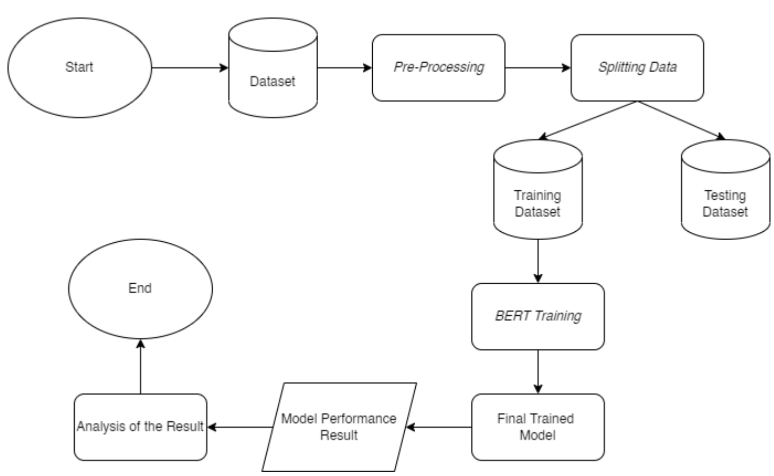

Fig. 3. Flow of the Built System

A. Dataset

In this research, the dataset comes from crawling the tweets of users on Twitter. Users whose tweets are crawled are selected based on criteria, namely having filled out the DASS-42 form first and are twitter users with at least 1000 tweets. The dataset consists of two labels, namely Symptoms of Depression and Not Depressed. The labelling process on the Training Set is done by taking opinions from 3 people to determine the appropriate label for the tweet, while the Testing Set is determined based on the results of the DASS-42 questionnaire that has been conducted. Table I shows an example of the data for the Training Set and Table II shows a short example of the data for the Testing Set which is the combination of dozens of tweets written by the user that were successfully obtained. The result of standardized data can be seen in Table I and Table II.

TABLE I
EXAMPLE OF THE TRAINING DATASET

Sentence	Label
RT @jfarxq ternyata lebih eNak semuanya dipendem sendiri, lagian nangis juga nangis sendiri, kuat juga kuat sendiri, depresi juga depresi sendiri, nyakitin diri juga diri sendiri. Ngapain juga cerita masalah sendiri ke orang lain, gaada yang peduli:) #capek https://t.co/9OI9odMxA2	Symptoms of Depression
Gemeteran tangan ini nge chat dosen, bego bet lagian dinda	Not Depressed

TABLE II
SHORT EXAMPLE OF THE TESTING DATASET

User	Sentence	Label
kata17_00	@literarybase Tidak pernah ada yg bertanya, karena aku memang terlihat baik-baik saja meski tak baik-baik saja You'll still defend her. Cause your kind nature or cause u love her. Or maybe both. Marah sampai mau nangis. Tapi g boleh nunjukin marahny. G boleh juga nangis. Tapi nyatanya, aku tetap hancur berantakan Aku ngerasa aku ngelakuin itu juga buat diriku sndri soalnya Baru sadar, kalau temanku nganggap aku toksik, ngebawa pengaruh buruk buat org lain Dan mgkn...sekat itu adalah pikiranku Tapi...terkadang dengan sadar, aku menjauh secara otomatis dari mereka Kenapa semuanya "negatif" sekali? @collegemenfess Syudah selesai dungss I am happy. But i am scared Satu²nya mimpiku yang belum menjadi abu Adalah menjadi seorang Ibu	Symptoms of Depression
bacaadong	@novandhiSP @The_RedsIndo Haha true😂 😂 @registaco Meriam lontong🤣 @Arsenal @idextratime 🤣🤣 @Bola_Jakarta Ga kaya yg onoh ya min🤣 Yang penting sabar hehe @melovsmealot Gas @melovsmealot Kan makan lah Jangan kan orang, yang punya badan aja kalo ngeliat badan nya sendiri rasanya pen nangis😌 @mimpi_van_arte @kang_kasur @BungFasoo @SiaranBolaLive Haha lord maguire El kontolero😂 @Bola_Jakarta @bepe20 Template ajg😂 @HaiMagazine Tau ajee bang boril😂 😂 @collegemenfess Kaco lu dosen Ampe nyariin begitu🤣🤣	Not Depressed

B. Pre-Processing Data

The data that has been collected is then subjected to a Data Pre-Processing process. Table III shows an example of a sentence in the pre-processing phase. The steps taken in this process include the following steps.

1) Data Cleaning

The process of removing signs that the tweet is the result of retweets and existing hyperlinks.

2) *Case Folding*
The process of converting all letters into lowercase letters.
3) *Removal Punctuation*
The process of removing punctuation marks in the sentence.
4) *Stemming*
The process of removing affixes on words.

TABLE III
EXAMPLE OF THE PRE-PROCESSING

Pre-Process	Sentence
Raw Data	RT @jfarxq ternyata lebih eNak semuanya dipendem sendiri, lagian nangis juga nangis sendiri, kuat juga kuat sendiri, depresi juga depresi sendiri, nyakitin diri juga diri sendiri. Ngapain juga cerita masalah sendiri ke orang lain, gaada yang peduli:) #capek https://t.co/9OI9odMxA2
Data Cleaning	@jfarxq ternyata lebih eNak semuanya dipendem sendiri, lagian nangis juga nangis sendiri, kuat juga kuat sendiri, depresi juga depresi sendiri, nyakitin diri juga diri sendiri. Ngapain juga cerita masalah sendiri ke orang lain, gaada yang peduli:) #capek
Case Folding	@jfarxq ternyata lebih enak semuanya dipendem sendiri, lagian nangis juga nangis sendiri, kuat juga kuat sendiri, depresi juga depresi sendiri, nyakitin diri juga diri sendiri. ngapain juga cerita masalah sendiri ke orang lain, gaada yang peduli:) #capek
Removal Punctuation	jfarxq ternyata lebih enak semuanya dipendem sendiri lagian nangis juga nangis sendiri kuat juga kuat sendiri depresi juga depresi sendiri nyakitin diri juga diri sendiri ngapain juga cerita masalah sendiri ke orang lain gaada yang peduli capek
Stemming	jfarxq nyata lebih enak semua dipendem sendiri lagi nang juga nang sendiri kuat juga kuat sendiri depresi juga depresi sendiri nyakitin diri juga diri sendiri ngapain juga cerita masalah sendiri ke orang lain gaada yang peduli capek

C. *IndoBERT Modelling*

IndoBERT is a modified result of BERT Base that has followed the settings of BERT-Base (uncased). In the process, IndoBERT uses a transformer mechanism, which is a mechanism that learns the relationship between words in a text/sentence. Formally, the transformer consists of two mechanisms, namely an encoder that functions to read the input text and a decoder that functions to generate predictions. Unlike other sequential-type models (left to right or vice versa) that read the text sequentially, the transformer encoder reads the entire word at once. These features can enable the model to learn the context of a word based on its surroundings (left and right words).

In the pre-training stage carried out by IndoBERT, there are two stages that must be carried out. First is Masked Language Modeling (Masked LM), which is inserting [MASK] tokens randomly in a sentence, where the model will try to predict the original value of the words given the [MASK] token. An example as in Table IV.

TABLE IV
MASKED LM PROCESS

Sentence	Encode Token Masked LM
pernah nggk ngetawain diri sendiri karna sering banget ngelakuin halhal bodoh yang ujungujungnya nyakitin diri sendiri	['pernah', 'ng', '##g', '##k', 'nget', '##awai', '##n', 'diri', 'sendiri', 'karna', 'sering', 'banget', 'ngel', '##aku', '##in', 'hal', '##hal', 'bodoh', 'yang', 'ujung', '##ujung', '##nya', 'nyak', '##iti', '##n', 'diri', 'sendiri']

The second stage is Next Sentence Prediction (NSP) where the model will try to guess the next sentence of the inputted sentence with the help of some tokens before being inserted into the model. Some tokens used are such as [CLS] and [SEP] tokens which are inserted at the beginning and end of the sentence respectively. In addition, there is also the token [PAD] which is a special token used for padding. An example is shown in Table V.

TABLE V
NSP PROCESS

Encode Token Masked LM	Encode Token NSP
['pernah', 'ng', '##g', '##k', 'nget', '##awai', '##n', 'diri', 'sendiri', 'karna', 'sering', 'banget', 'ngel', '##aku', '##in', 'hal', '##hal', 'bodoh', 'yang', 'ujung', '##ujung', '##nya', 'nyak', '##iti', '##n', 'diri', 'sendiri']	['[CLS]', 'pernah','ng', '##g','##k', 'nget','##awai', '##n', 'diri', 'sendiri', 'karna', 'sering', 'banget', 'ngel', '##aku', '##in', 'hal', '##hal', 'bodoh', 'yang', 'ujung', '##ujung', '##nya', 'nyak', '##iti', '##n','diri', 'sendiri','[SEP]','[PAD]','[PAD]','[PAD]']

The next step is to perform sentiment analysis with the model and since there is a similarity with the Next Sentence Classification task, the model will add a layer above the Transformer output. To perform fine-tuning, it is necessary to refine the parameters used. Table VI shows the parameter values used in this experiment.

TABLE VI
INDOBERT MODEL PARAMETER VALUES

Parameter	Value
Batch Size	32
Dropout	0.1
Epoch	15
Learning Rate	2e-5

D. Performance

At this stage, evaluation and analysis of the results obtained from the previous stages are carried out. Evaluation is carried out using the Confusion Matrix method and looking for accuracy, precision, recall, and F1-Score values. This Confusion Matrix method consists of 4 types of values, namely TP (True Positive), FP (False Positive), FN (False Negative), and TN (True Negative). These values are then used to analyse the results of the model work by calculating the accuracy, precision, recall and f1-score values.

In Confusion Matrix, TP means a case where the predicted value is yes and the actual value is also yes, while TN is a case where the predicted value is no and the actual value is no as well. For FP and FN are cases where the predicted value and the actual value are not the same[14]. In this study, yes indicates depressive symptoms while no means not depressed. Based on that, we able to get the following formula.

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

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

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

$$F1 - Score = \frac{2 * (\mathrm{Recall} * \mathrm{Precision})}{\mathrm{Recall} + \mathrm{Precision}} \tag{4}$$

Accuracy represents the number of correctly classified data out of the total data. Accuracy may not be a good measure if the dataset is unbalanced (different number of positive and negative data). Therefore, other calculations are needed, namely precision (positive predictive value) and recall (true positive rate). The more the precision and recall values approach the value of one, the better the classification model is. The ideal classification model is one precision and recall, where the FP and FN values are zero. Therefore, we need a matrix that takes precision and recall into account, namely F1-Score. F1-Score is a harmonic mean of precision and recall and is a better assessment of accuracy[11].

IV. RESULTS AND DISCUSSION

Testing is done by comparing the accuracy value and the value obtained from the Confusion Matrix method where two test scenarios are carried out, namely the Testing Set with preprocessing and those without. The results of the value comparison can be seen in Table VII.

TABLE VII
TESTING COMPARISON RESULTS

Dataset	Accuracy	Precision	Recall	F1-Score
With *Pre-Process*	0.5111	0.4839	0.2308	0.3125
Without *Pre-Process*	0.5111	0.4828	0.2154	0.2979

The comparison results show that both datasets get the same accuracy. However, it can be seen based on another assessment component, namely F1-Score, that data that is pre-processed gives a better value than those that are not. This indicates that the model is better at detecting depressive tweets if the data is preprocessed. This shows that pre-processing has an effect on how the IndoBERT model detects symptoms of depression in tweets issued by users. The results of this experiment also assume that not all people who are considered to show symptoms of depression when filling out the DASS-42 form will also show it in their tweets. For example, a user with the username "antidamay_" got a result showing symptoms of depression based on the DASS-42. However, detection based on the tweets the user has written shows that it is not depressed. This case also occurred for several users, causing the low performance value.

TABLE VIII
PERFORMANCE MEASURE

With *Pre-Process*				**Without *Pre-Process***			
		Predicted Class				**Predicted Class**	
		Positive	**Negative**			**Positive**	**Negative**
Actual Class	Positive	15	16	Actual Class	Positive	14	15
	Negative	50	54		Negative	51	55

This study used data consisting of 135 users who had filled out the DASS-42 questionnaire first, where based on the results of the questionnaire, 65 of them showed symptoms of depression. Each user consisted of 20-100 tweets combined per user. From the results shown in Table VIII, it can be seen that there are 66 data or almost 50% of the data that are misclassified. After being analyzed, the Confusion Matrix value on data that is pre-processed and those that are not is not much different, where the value is only one number away. This result can provide an answer that pre-processing in sentiment analysis using the IndoBERT model, at least in this research, does not significantly affect the results obtained. This can also explain why the accuracy value on both datasets has exactly the same value, while the reason why the value obtained is fairly low is back to the quality of the data obtained and used in this experiment. The parameters used are also influential, such as the learning rate and epoch, which when changed can give different results from those already obtained. The parameters used and already mentioned in the previous section have given the best results in this study compared to other combinations of parameter values.

Based on the test results, some analysis can be done on several tweets that were tested that may cause the model to not predict correctly and the following reasons were obtained. First, there are words that make it difficult for the model from recognizing the context of the whole sentence, such as in the sentence " duh masalah timbul mulu, cape dah, pengen mati aja kyknya wkwkwkwkkwkw". This word wkwkwkwkkwkw is still not recognized by the model to determine the context of the sentence. The second one, there are some words that change its meaning during pre-processing. An example is the word "bajingan" which when the affix is removed changes to "bajing" and becomes a word with a different meaning from the actual one. The next one, there are some tweets that consist of a combination of Indonesian and other languages, even some tweets only consist of foreign languages, such as English, Korean, Japanese and also regional languages, such as Sundanese and Javanese. In the case of Japanese and Korean languages that do not use alphabetic letters, the tweets were completely removed in the pre-processing phase. The last one, overfitting occurs, where during training, the model gets an accuracy value of 0.8796. When the model was used to detect using the testing data, the model was not able to do its job well.

V. CONCLUSION

This study was conducted to determine whether the use of the IndoBERT model can detect sentences with symptoms of depression optimally on the dataset analyzed, namely data in the form of tweets from Twitter. Based on the research conducted, the IndoBERT model in detecting sentences with symptoms of depression obtained an accuracy value of 51% for both datasets. These results were successfully obtained because IndoBERT itself is an Indonesian variant of the pre-existing BERT model and can perform depression detection as it has a mechanism for learning the context of words in a sentence so that in the learning phase it can assess words more contextually.

This study also found some limitations such as the emergence of words that cannot be classified such as "wkwkwkwk". In addition, often the tweets obtained contain words that come from other than Indonesian, making the detection process more difficult. In future research, it is recommended not to use tweets that contain languages other than Indonesian or to first translate the words into Indonesian.

VI. References

[1] Iqbal, Mansoor.. "Twitter Revenue and Usage Statistics (2021)," Business of Apps. November 2021. [Online]. Available: https://www.businessofapps.com/data/twitter-statistics/.

[2] The Health and Safety Executive. "Work-related stress, anxiety or depression," 2020.

[3] Mutia, Annisa. "Survei: 68% Orang Depresi Akibat Covid-19," Databoks – Katadata. 17 June 2021. [Online]. Available: https://databoks.katadata.co.id/datapublish/2021/06/17/survei-covid-19-menggangu-kesehatan-jiwa-68-orang-depresi.

[4] Healthfocus Clinical Psychology Services. "Depression Anxiety and Stress Scale DASS (-42) – Healthfocus Clinical Psychology Services," Healthfocus Clinical Psychology Services. [Online]. Available: https://www.healthfocuspsychology.com.au/tools/dass-42/.

[5] Islam, Md. R., Kabir, M. A., Ahmed, A., Kamal, A.R. M., Wang, H. and Ulhaq, A. "Depression detection from social network data using machine learning techniques," *Health Information Science and Systems/* no. 6. 2018.

[6] Devlin, J., Chang, M. -W., Lee, K. and Toutanova, K. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers).* Minneapolis, Minnesota. Association for Computational Linguistics. 2019. pp. 4171-4186.

[7] Koto, F., Rahimi, A., Lau, J. H. and Baldwin, T. "IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained," in *Proceedings of the 28th International Conference on Computational Linguistics.* Barcelona, Spain. International Committee on Computational Linguistics. 2020. pp. 757-770.

[8] Koto, F., Lau, J. H. and Baldwin, T. "Liputan6: A Large-scale Indonesian Dataset for Text Summarization," in *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing.* Suzhou, China. Association for Computational Linguistics. 2020. pp. 598-608.

[9] Burdisso, S. G., Errecalde, M. and Montes-y-Gómez, M. "A Text Classification Framework for Simple and Effective Early Depression Detection Over Social Media Streams," *Expert Systems with Applications.* vol. 133. pp. 182-197. 2019.

[10] Leonard, L. C.. "Web-Based Behavioral Modeling for Continuous User Authentication (CUA)," *Advances in Computer.* 2017.

[11] B, H. N.. "Confusion Matrix, Accuracy, Precision, Recall, F1 Score," Analytics Vidhya | Medium, 11 December 2019. [Online]. Available: https://medium.com/analytics-vidhya/confusion-matrix-accuracy-precision-recall-f1-score-ade299cf63cd.

[12] Lovibond, P. F., Lovibond, S. H., "The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories." in *Behaviour Research and Therapy*. vol. 33. pp 335-343. 1995.

[13] Zhang, L., Wang, S., and Liu, B. "Deep learning for sentiment analysis: A survey," in *WIREs Data Mining and Knowledge Discovery*, vol. 8, no. 4, 2018.

[14] Siburian, V. W., Mulayana, I. E., "Prediksi Harga Ponsem Menggunakan Metode Random Forest," in *Prosiding Annual Research Seminar 2018*, vol. 4, no. 1, 2018.