Building of Informatics, Technology and Science (BITS)
Volume 4, No 2, September 2022 Page: 369−376
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i2.1850

Depression Detection on Twitter Using Bidirectional Long Short Term Memory

Putri Ester Sumolang[1], Warih Maharani[2]

Faculty of Informatics, Informatics Department, Telkom University, Bandung, Indonesia
Email: [1]putriester@student.telkomuniversity.ac.id, [2]wmaharani@telkomuniversity.ac.id
Correspondent Author Email: wmaharani@telkomuniversity.ac.id
Submitted: **11/07/2022**; Accepted: **14/08/2022**; Published**: 30/09/2022**

Abstract− The usage of social media as a platform for individual expression is one example of how the advancement of technology has an impact on society. Therefore, many Twitter users show symptoms of depressive disorders through their tweets. It is crucial to be aware of the need to consult a doctor or other specialists to prevent suicides. However, leveraging Twitter user tweet data to detect depression early on can be avoided by using the Bidirectional Long Short Term Memory (BILSTM) approach and the word2vec feature extraction method. The dataset utilized in this study was obtained from respondents who agreed to have their data used in research after completing a questionnaire based on the Depression Anxiety and Stress Scales - 42 (DASS-42). The whole data from 159 users of Twitter who have been classified as depressed or normal based on the results of the DASS-42 labeling are then preprocessed so that the data can be entered into the word2vec feature extraction and modeled by BiLSTM as a classification. The evaluation revealed an accuracy of 83.46 % and an f1-score of 87.11 %. By increasing the number of neurons, accuracy increased by 2.36 %, and f1-score climbed by 1.64 %.

Keywords: Twitter; Depression; DASS-42; Word2Vec; BiLSTM

1. INTRODUCTION

The usage of social media as a platform for individual expression is one example of how the advancement of technology has an impact on society. Instagram, Twitter, and Facebook have all become part of our daily routines for sharing moments, ideas, status updates, and feelings. 75% of people utilize one of the many social networking services, like Twitter. Twitter, a well-known social media network introduced in 2006 that allows users to send and read published posts, is the ideal platform for data collection.

In recent years, the importance of mental health in accomplishing global development goals has become more widely recognized [2]. One of the significant factors in the overall global disease load and one of the leading causes of disability worldwide is depression [2]. According to Kamus Besar Bahasa Indonesia, a person with depression has a mental disease marked by a sense of decline (such as gloomy, sad, or depressed feeling). Depressive disorders have a prevalence of 6.2 percent, according to the 2018 Riskesdas (Riset Kesehatan Dasar) data, and they first appear in adolescents (15–24 years) [3]. Most of the time, depression is a health issue at the association level that necessitates awareness of the symptoms; those who experience severe depression will also exhibit suicidal and self-harming tendencies [3]. Suicide is the second most common cause of death for those between the ages of 15 and 29, according to the WHO [2]. Therefore, it is critical to recognize if someone is depressed or not to lessen the effects. In [1], it is claimed that social media posts can indicate the possibility of depression, with the dataset utilized coming from Twitter social media tweet posting activity. If a user tweets something that appears to be depressive and no steps are made to prevent it, the person may self-harm or, in the worst-case scenario, attempt suicide. The article [4] provides evidence of this.

Sangeeta R. Kamite and Dr. V. B. Kamble used machine learning techniques like naive bayes and random forests in earlier studies to identify depression-related concerns on social media [5]. The model must be continuously trained by naive bayes, producing many complex results [5]. According to this study, it is clear that having a large amount of data is necessary for good sentiment tweet prediction because acquiring annotated datasets and preparing the data for analysis are the most difficult tasks. Deep learning techniques are something that Sanggeta R. Kamite and Dr. V. B. Kamble mentioned they intend to apply for better outcomes [5].

In their study [6], Faisal M. S., Farzad A., Sajib K. S. J., Sifat A., Samir S., Rimon S., and Md. Hasanul K. applied word2vec as a feature extraction to deep learning techniques to identify depression on social media. The use of word2vec embed and meta-features has been done well and can classify correctly, but it takes a long time to detect, according to his research [6]. Aulia, R.I., Agus S., Yohanes S., Bidirectional long short-term memory (BiLSTM) and word2vec extraction were used to detect hate speech. Based on their findings, it can be said that employing BiLSTM and word2vec approaches with CBOW architecture can provide high accuracy [7].

Deep learning techniques have been utilized in previous studies to identify problems on social media, like in the research [5] by Sangeeta R. K. with Dr. V. B. Kamble and the research [6] by Faisal M. S., Farzad A., Sajib K. S. J., Sifat A., Samir S., Rimon S., and Md. Hasanul K. to identify depression on social media. Machine learning techniques like support vector machines and naive bayes must train the model continually, leading to various complex outcomes [6]. Using Twitter as a source of knowledge and data on social media [6]. Supported by the LSTM method's prowess in categorizing and forecasting time series data with ambiguous time scales [8][9].

Our study uses a BiLSTM approach to analyze tweets from Twitter users to identify depression using Bidirectional Long Short Term Memory (BiLSTM). By using DASS-42 as labeling of depression or not [10]. BiLSTM

Building of Informatics, Technology and Science (BITS)
Volume 4, No 2, September 2022 Page: 369−376
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i2.1850

is a two-layer LSTM neural network [6][7]. The excellent ability of the LSTM approach in categorizing and predicting time series data with uncertain time duration is used in conjunction with Twitter social media as a source of information and data [5][7]. We hope that this research will make it easier to determine whether someone is depressed or not earlier.

The following writing organization forms the basis of this research's structure: The first part explains the introduction, the second part explains the research techniques utilized, the third part provides an explanation of the findings and an analysis, and the fourth part provides a summary.

2. RESEARCH METHODOLOGY

2.1 Research Stages

This study used word2vec feature extraction and the BiLSTM approach to build a system that can identify depression. According to Figure 1, the system flow begins with the collection of Twitter data, following which the data is stored in Excel in.csv format. The data is then processed at the preprocessing stage, which includes case folding, normalization, stopword removal, stemming, and tokenization. During the feature extraction process, the preprocessed data are converted using word2vec into a vector. The word is transformed into a vector, and then the data is split between training and test sets. The train data are classified using the BiLSTM approach and evaluated alongside the testing data during the model evaluation phase. Figure 1 below shows the system's logical flow in the following steps.

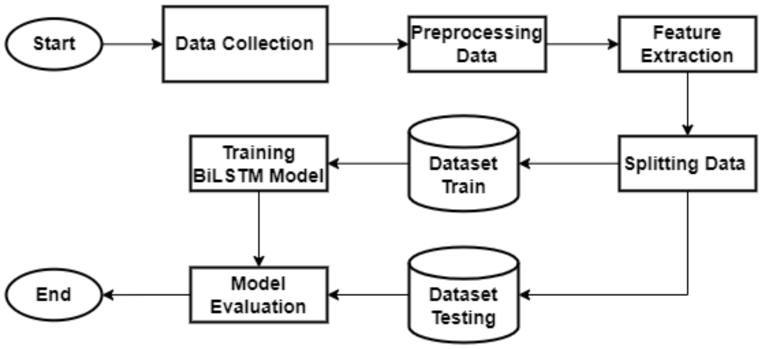

Figure 1. Flowchart System

2.2 Data Collection

In this study, the dataset was used primary dataset, so twitter data was collected from participants who responded to a questionnaire about their mental health using the DASS-42. The data taken are tweets and usernames from respondents and stored in excel files in .csv format. We took 100 tweets for each respondent account. The data collected is 159 users from 342 respondents who filled out the questionnaire and labeled it according to the questionnaire with questions based on DASS-42 [10]. According to Table 1, users who score between 0 and 9 are classified as normal, whereas those who score more than nine are classified as depressed.

Table 1. The Severity of The Disorder [10]

Disorder	Severity level				
	Normal	Low	Medium	Heavy	Extreme
Depression	0 – 9	10 – 13	14 – 20	21 – 27	28+

2.3 Preprocessing Data

This process cleans the raw data from noise to produce structured data. Text data with redundant and unusual words, such as URL, mention, hashtag, and others, are eliminated throughout this process. The processes include case folding, normalization, stopword removal, stemming, and tokenization. Depending on the dataset's conditions, this step may need to be expanded. The preparation processes are described in the section below.

a. Case folding: Case folding changes all capital letters to lowercase or uppercase letters in a character [11].
b. Normalization: Normalization is a step in the data cleaning process. Removed noise includes hashtags, emoticons, mentions, rt/cc, whitespace, repeat words, URLs, numbers and symbols, and single characters. Slang words are transformed into appropriate Bahasa Indonesia through this process of normalization.
c. Stopword removal: Stop word removal is a step in removing unnecessary words. Nltk is the library utilized for stop word removal [12]. Additionally, stopword removal eliminates frequently used words such as “aku”, “kamu”, “dan”, “yang”, “ke”, “yang”, “di”, “ini”, “itu”.
d. Stemming: Stemming is the process of returning the affixes to the original word. The library used is sastrawi [11]. For instance, the root word "bawa" is shared by the words "membawa" and "bawain".

Building of Informatics, Technology and Science (BITS)
Volume 4, No 2, September 2022 Page: 369−376
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i2.1850

e. Tokenization: The tokenization process involves dividing the sentence into a list of words. For instance, after the tokenization process, the phrase "tidak guna" becomes "tidak", "guna".

2.4 Feature Extraction with Word2Vec

Feature extraction is the process of turning textual information into numbers. The input data must be a matrix or vector used in the BiLSTM procedure. In this study, we used word2vec as Feature Extraction. Word2Vec is a one-word embedding technique for representing words in a vector [13]. For instance, the word "hujan" and the word "dingin" have a similar vector representation because the word2vec model recognizes that the two words have a similar association with weather and temperature [13].

2.5 Splitting Data

Before utilizing BiLSTM to classify the dataset, it is split into training and test data sets. By doing this, the risk of the model overfitting to the training data is decreased. For train data, data sharing is 80%, and for test data, data sharing is 20%. The results of data splitting are 127 train data and 32 testing data. The data distribution was chosen randomly to maintain a balance between negative labeling, which represents normal, and positive labeling, which represents depression from the twitter data. Table 2 below provides a visual representation of the data distribution.

Table 2. Distribution of Labeled Data in Train Data and Test Data

	Data Training		Data Testing	
Jumlah	Positive	Negative	Positive	Negative
	77	50	17	15

2.5 Training BiLSTM Model

The Bidirectional Long Short Term Memory (BiLSTM) method categorizes the training twitter data as depressed or not. BiLSTM allows for more precise prediction because each step can include data from past and future conditions [6]. BiLSTM is a neural network of Long Short-Term Memory (LSTM) consisting of 2 layers, namely, forward LSTM, which models the previous context, and backward LSTM, which models the following context [7][14]. Figure 2 below shows the BiLSTM architecture.

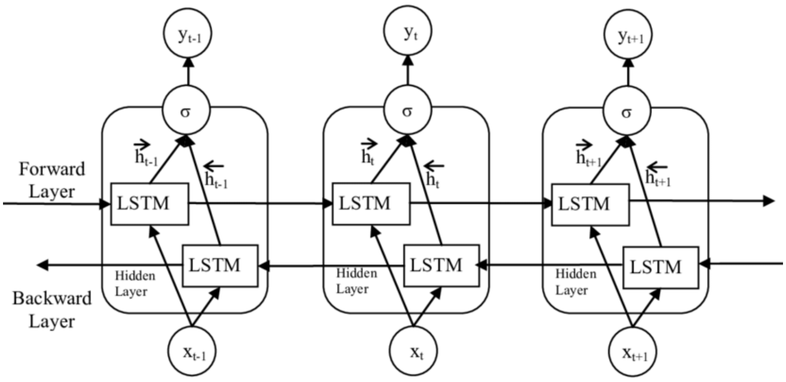

Figure 2. BiLSTM Architecture [15]

Long Short-Term Memory (LSTM) is a Recurrent Neural Network (RNN) architecture. Long-term dependencies, disappearing gradients, or explosive gradients can all be handled with LSTM [9]. Figure 3 illustrates the several gates that make up the LSTM. The gate's operation and how LSTMs function is explained as follows [8] [9].

a. Forget Gate
The forget gate (f_t) determines how much information to erase after receiving the output of the previous state from state h_{t-1} and outputs a value of zero to indicate that the information will be discarded and a value of one to indicate that the information will be continued. Equation 1 can be used to calculate the forget gate value (f_t) as shown below:

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{1}$$

b. Input Gate
The input gate (i_t) then chooses to update the cell state ($\tilde{C}_t$)with the new item from the current input. In addition to the input gate (i_t), there is a tanh layer that generates a vector for new candidates to be added to the current cell state ($\tilde{C}_t$). For the gate input (i_t) and tanh, equations 2 and 3 are as follows:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{2}$$

$$\tilde{C}_t = tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{3}$$

Building of Informatics, Technology and Science (BITS)
Volume 4, No 2, September 2022 Page: 369−376
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i2.1850

To eliminate redundant data, the forget gate's output (f_t) is multiplied by the prior cell state (C_{t-1}). The information to be used for the new cell state (C_t) is then determined by multiplying the gate input (i_t) and the current cell state. The multiplied results are then added. The following is how Equation 4 is applied:

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \quad (4)$$

c. Output Gate
The output gate (o_t) then determines what data should be output from the cell state. Then the result of the output gate (o_t) is multiplied by the tanh layer which produces an output that is valued between -1 or 1. Following are equations 5 and 6 that are used here:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = (o_t \times tanh\ C_t) \quad (6)$$

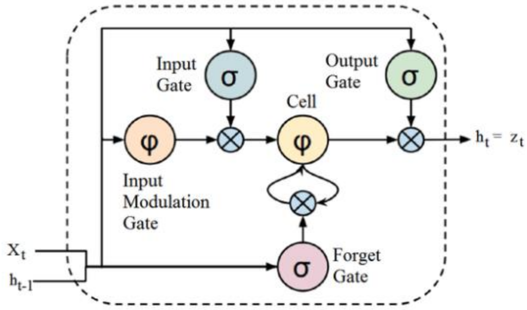

Figure 3. LSTM Gate [8]

2.5 Model Evaluation

In this research, the performance of the model that has been made is measured using the confusion matrix. Four terms in the confusion matrix represent the classification process: true positive (TP) indicates a positive prediction and is positive, true negative (TN) indicates a negative prediction and is negative, false positive (FP) indicates a positive prediction but is negative, and false negative (FN) indicates a negative prediction but is positive [16]. More information can be found in table 3.

Table 3. Confusion Matrix

	Actually Positive	Actually Negative
Predicted Positive	TP	FP
Predicted Negative	FN	TN

A confusion matrix is utilized as the evaluation tool to determine how accurately machine learning algorithms classify the input into the proper labels [17]. The illustrations in Table 3 are beneficial for measuring recall, precision, accuracy, and F1-score, which will be calculated by formulas 7,8,9,10. The recall is the ratio of all predicted positive classes to positive. Precision is the ratio of all predicted positive classes to the number of true positives. Accuracy is the ratio of all classes (positive and negative) that are correctly predicted. F1-measure is the average value for measuring recall and precision together. Equations 7, 8, 9, and 10 can be understood as follows [17][18].

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

$$Precision = \frac{TP}{TP+FP} \quad (8)$$

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (9)$$

$$F1 - measure = \frac{(2 \times Recall \times Precision)}{(Recall+Precision)} \quad (10)$$

3. RESULTS AND DISCUSSION

3.1 Data Preprocess Result

Data was collected from 159 Twitter users, with 100 tweets collected from each user for 15,900 tweets. Data preprocessing contains case folding, normalization, stopword removal, stemming, and tokenizing. Table 4 shows how case folding converts all words in input to a lowercase character in output. After entering the normalization process, the data is removed from hashtags, emoticons, mentions, whitespace, URLs, digits, and symbols, leaving only existing words as can be seen in table 4. Next, proceed to the stopword removal process, which removes useless and

Building of Informatics, Technology and Science (BITS)
Volume 4, No 2, September 2022 Page: 369−376
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i2.1850

meaningless words. Table 4 shows that "gua" has been removed. The next step is the stemming process, where words with affixes are converted back to their original words. Table 4 shows how "mengantuk" is transformed into "kantuk". And the next step is tokenization, where each word is separated into a list of words such as “cinta senang” into “cinta” and “senang”. The results of the preprocessing procedure are shown in Table 4 below.

Table 4. Preprocessing Data Result

Process	Input	Output
Case folding	Luvv??????#HappyTaehyungDay https://t.co/oD3C7UxUFO Hbd ya ???? @BTS_twt GUA BUTUH HOLIDAY ngantuk pars	luvv??????#happytaehyungday https://t.co/od3c7uxufo hbd ya ???? @bts_twt gua butuh holiday ngantuk pars
Normalize	luvv??????#happytaehyungday https://t.co/od3c7uxufo hbd ya ???? @bts_twt gua butuh holiday ngantuk pars	cinta senang taehyung hari selamat ulang tahun gua butuh liburan mengantuk parah
Stopword Removal	cinta senang taehyung hari selamat ulang tahun gua butuh liburan mengantuk parah	cinta senang taehyung hari selamat ulang tahun butuh liburan mengantuk parah
Stemming	cinta senang taehyung hari selamat ulang tahun butuh liburan mengantuk parah	cinta senang taehyung hari selamat ulang tahun butuh liburan kantuk parah
Tokenize	cinta senang taehyung hari selamat ulang tahun butuh liburan kantuk parah	‘cinta’, ‘senang’, ‘taehyung’, ‘hari’, ‘selamat’, ‘ulang’, ‘tahun’, ‘butuh’, ‘liburan’, ‘kantuk’, ‘parah’

3.2 Feature Extraction Result

In this study, we use word2vec as feature extraction, as explained in point 2.4, which converts the word into a vector. We employ the Word2vec settings window size 5, vector size 300, min_count 10, and workers 4. The extraction results for the words "sedih" and "hidup," as shown in Table 5, are transformed into a vector with a 300-vector length. There are 300 vectors for each extracted word. The extraction results can be seen in table 5.

Table 5. Word2Vec Result

Words	Word2Vec Vector
sedih	[0.58588785, 0.18515052, … ,0.0463098]
hidup	[0.99907225, 0.0240291, …, -0.00181407]

Table 6 shows that, according to vector values, the word "sedih" is comparable to several other terms. The most similar terms are only picked from 10 words, according to the min_count. Word2vec identifies that similar relationships exist between words. The most similar words are displayed in table 6 below.

Table 6. The Most Similar Words

Words	Most Similar Words
sedih	[('manis', 0.7995769381523132), ('hati', 0.7870360612869263), ('parah', 0.762904167175293), ('curhat', 0.746485710144043), ('sesak', 0.7116708755493164), ('lewat', 0.711042046546936), ('tahan', 0.7095839381217957), ('gelap', 0.6995427012443542), ('sibuk', 0.698088526725769), ('matahari', 0.6941350698471069)]
hidup	[('gelap', 0.8763992786407471), ('damai', 0.8274452686309814), ('buruk', 0.7951920032501221), ('mudah', 0.7920188903808594), ('tahan', 0.7770768404006958), ('lelah', 0.7569860219955444), ('sadar', 0.7561659216880798), ('hadap', 0.73512202501297), ('kanan', 0.7308627963066101), ('sukses', 0.7222609519958496)]

3.3 Classification

Data enters the embedding layer process with a size of 300 dimensions after it has become a vector due to the feature extraction phase. The classification process starts from the input layer, LSTM layer, Conv1D layer, and output layer. The results of the input layer that has been completed enter the LSTM layer using the Bidirectional LSTM process and a dropout of 0.1. After that, Enter the conv1d layer, which contains the conv1d process to increase or decrease the intensity of the value. The GlobalMaxPool1D process is then used to downsample the input representation by taking the maximum value across other dimensions. The outcome is sent to the output layer, which has two dense layers: one using the rectified linear unit (ReLU) activation function and the other using the sigmoid activation function. Dense with activation function ReLU has 64 dimensions, and dense with sigmoid activation has one dimension.

As a result, we use the five levels previously mentioned in the classification procedure. Here, we use default Keras API parameters, such as the Adam optimizer with a 32-batch size. Table 7 below shows the dimensional parameters for dense.

Building of Informatics, Technology and Science (BITS)
Volume 4, No 2, September 2022 Page: 369−376
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i2.1850

Table 7. Dense Dimension Setting

Layer	Type	Neuron
Input Layer	Bidirectional LSTM	100
	Dropout	0.1
LSTM Layer	Bidirectional LSTM	100
	Dropout	0.1
Conv1D Layer	Filters	100
	Kernel size	5
Output Layer	Dense	64,1

3.4 Model Evaluation and Experiment Result

BiLSTM-classified data is trained with an epoch eight and batch size of 32. Table 8 shows test and training data's accuracy, precision, recall, and f1-score values. The data used for the train set was 127 with 77 positive and 50 negative labels, while 32 were used for the test set with 17 positive and 15 negative labels. Table 8 shows that the classification of the training data resulted in high accuracy and f1-score, with 83.46% accuracy and 87.11% f1-score. In contrast, the accuracy and f1-score for the testing data tend to be low, with an accuracy of 56.25% and f1-score of 61.11%.

Table 8. Model Result

Data	Accuracy(%)	Precision(%)	Recall(%)	F1-Score(%)
Training	83.46%	82.55%	92.20%	87.11%
Testing	56.25%	57.89%	64.70%	61.11%

So, using train data, test data, and batch size, we tested the model by hyperparameter tuning the dimensions of each model layer. Bidirectional LSTM and conv1d parameters with dimension sizes of 64, 128, and 256 and batch sizes of 16, 32, and 64 with 8 epochs were the subjects of experiments. The comparison of accuracy and f1-score in Table 9 below shows the experimental outcomes.

Table 9. The Experiment on Neurons and Batch Size Results

Batch Size	Neuron	Data Training		Data Testing	
		Accuracy	F1-Score	Accuracy	F1-Score
16	64	74.80%	79.74%	40.62%	53.65%
	128	91.33%	92.90%	53.12%	63.41%
	256	88.18%	90.44%	46.87%	45.16%
32	64	78.74%	82.11%	53.12%	61.53%
	128	86.61%	89.03%	0.50%	66.66%
	256	85.82%	88.75%	53.12%	68.08%
64	64	74.01%	80.92%	0.50%	61.90%
	128	65.35%	77.31%	0.50%	61.90%
	256	83.46%	84.67%	46.87%	48.48%

The results of the experiments were conducted by comparing the accuracy and f1-score of the training and testing data. Based on Table 9, We found that a batch size of 32 with 256 dimensions can obtain good results with an accuracy of 53.12% and an f1-score of 68.08% on the testing data. However, the best accuracy and f1-score on the training data is when the batch size is 16 and 128 dimensions, with an accuracy of 91.33% and f1-score of 92.90%. Consequently, we consider the best experimental results when the batch size is 32 and 256 dimensions because the accuracy and f1-score of the training and testing data are not much different from each other compared to other dimension and batch size experiments. The accuracy difference between the training and testing data is 32.7%, while the f1-score difference is 20.67%. The confusion matrix results with batch size 32 and dimension 256 can be seen in Table 10.

Table 10. Confusion Matrix Result

	Actually Positive	Actually Negative
Predicted Positive	50.00%	43.75%
Predicted Negative	3.12%	3.12%

Based on Table 10, 16 out of 50% of twitter users are TP (True Positive) individuals detected as depressed and actually depressed. TN (True Negative) users who were detected as not depressed and were indeed not depressed were one twitter user out of 3.12%. One twitter user out of 3.12 % FP (False Positive) was recognized as depressed, but whose diagnosis was not depressed. The prediction was wrong for 14 out of 43.75 % FN (False Negative) users who were identified as not depressed; they're actually depressed. According to the test results, batch size and dimension in bidirectional LSTM and conv1d have an impact on the previous classification results. The previous f1 score in the

Building of Informatics, Technology and Science (BITS)
Volume 4, No 2, September 2022 Page: 369−376
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i2.1850

training data increased from 87.11% to 88.75%, and in the testing, data increased from 61.11% to 68.08%. After that, we predict whether the user is depressed or not. Table 11 below displays the user's prediction results.

Table 11. Prediction Result

Username	Tweet	Predict
Username1	'sial', tiktok, 'suka', 'nurul', 'budi', 'aluka', 'cowok', 'anjing', 'menantea', 'lari', 'hujan', 'bikin', 'mens', 'tugas', 'suka', 'tebak', 'banget', 'libur', 'hotel', 'buruk', 'gila', 'tutup', 'hati', 'pikir', 'tes', 'detak', 'hati', 'karma', 'karma', 'kwangso', 'drama', 'korea', 'gelap', 'gelap', 'ya', 'wn', 'yuhu', 'manis', 'cepat', 'banget', 'opa', 'armin', 'bawa', 'toge', 'obat', 'batuk', 'tawa', 'pusing', 'dunia', 'juara', 'semangat', 'takut', 'moga', 'mas', 'sepupu', 'senang', 'gilbe', 'hidup', 'gemas', 'banget', 'suka', 'mas', 'bri', 'suami', 'badai', 'cinta', 'paham', 'yami', 'culik', 'tonton', 'budak', 'cinta', 'selingkuh', 'selingkuh', 'sabar', 'tanggal', 'senang', 'hobi', 'selamat', 'ulang', 'norman', 'hidup', 'astaga', 'penyamarataan', 'bikin', 'anime', 'karma', 'pikir', 'narak', 'bayar', 'menang', 'halusinasi', 'astaga', 'gila', 'yuzu', 'kencan', 'ban', 'mantap', 'cinta', 'selamat', 'ulang', 'ya', 'butuh', 'libur', 'senang', 'cerah', 'ekspektasi', 'makan', 'latar', 'kayak', 'kenal', 'bunda', 'harap', 'coba', 'dasar', 'ntnya', 'disney', 'lef', 'hijau', 'soh', 'habis', 'seru', 'parah', 'fakta', 'banyak', 'cabang', 'tahan', 'pusat', 'moto', 'hidup', 'zidan', 'sociola', 'tugas', 'astaga', 'tolong', 'hancur', 'rebut', 'laki', 'orang', 'pergi', 'dos', 'suzy', 'jipyeong', 'panggil', 'jin', 'soh', 'penuh', 'memori', 'bocor', 'mbak', 'sai', 'reu', 'bplao', 'ya', 'bunda', 'padu', 'tarik', 'kelapa', 'tangan', 'krim', 'kelapa', 'badan', 'losion'	Positive
Username2	'fyp', 'makan', 'fase', 'bingung', 'thv', 'foto', 'sayang', 'sumpah', 'marah', 'ya', 'sudah', 'efek', 'pms', 'bawa', 'asa', 'banding', 'coba', 'lihat', 'iya', 'haha', 'jari', 'tangan', 'kiri', 'darah', 'tinggal', 'sebal', 'tutup', 'mangkel', 'ambek', 'adem', 'banget', 'bilang', 'terima', 'kasih', 'bikin', 'harga', 'sayang', 'sembuh', 'baik', 'gin', 'ubah', 'suasana', 'hati', 'malam', 'mimpi', 'om', 'kaya', 'om', 'kaya', 'jb', 'mimpi', 'nama', 'auto', 'bieber', 'hailey', 'ter', 'rontok', 'banget', 'drama', 'laptop', 'mati', 'ulang', 'mcflury', 'malas', 'gerak', 'beli', 'leyot', 'anak', 'ta', 'ngueng', 'ngueng', 'pusing', 'seblak', 'gas', 'dasar', 'bocah', 'aneh', 'sakit', 'pinggang', 'badan', 'ikut', 'meriang', 'main', 'main', 'kelam', 'duduk', 'sakit', 'main', 'pakai', 'makan', 'benar', 'napas', 'bahasa', 'ta', 'ledak', 'otak', 'thor', 'amuk', 'lemah', 'banget', 'perut', 'heran', 'banyak', 'bicara', 'senang', 'ulang', 'sayang', 'cape', 'kali', 'kenal', 'materai', 'pikir', 'ya', 'beli', 'jajan', 'pakai', 'materai', 'materai', 'orang', 'malu', 'enak', 'tawa', 'hujan', 'seblak', 'parah', 'sebal', 'senang', 'dasar', 'wanita', 'lewat', 'bandung', 'sakit', 'selamat', 'ulang', 'sayang', 'kali', 'mimpi', 'cowok', 'muka', 'tolong', 'makan', 'canda', 'canda', 'sayang', 'bangun', 'pagi', 'pusing', 'lanjut', 'tidur', 'pusing', 'praktis', 'oke', 'oke', 'oke', 'tatap', 'mata', 'tarik', 'hati', 'asik', 'asik', 'gila', 'orang', 'benar', 'cape', 'pusing', 'cape', 'salah', 'marah', 'pintar', 'serba', 'salah', 'sial', 'bingung', 'banget', 'langit', 'tolong', 'cuek', 'ya', 'cape', 'tanggap', 'orang', 'lambat', 'gas', 'langit', 'kasih', 'lihat', 'beda', 'paket', 'seminar', 'proposal', 'buru', 'tonton', 'cowok', 'lucu', 'pai', 'ucap', 'uang', 'habis', 'revisi', 'niat', 'banget', 'kece', 'parah', 'cinta', 'pokok', 'suka', 'sahabat', 'vmin', 'jual', 'lihat', 'tunggu', 'biar', 'kelelawar', 'iya', 'juta', 'umat', 'rasa', 'bangun', 'kangen', 'nchim', 'ulang', 'ketua', 'cinta', 'hati', 'sanggup', 'jam', 'lambat', 'ulang', 'sayang', 'rusa', 'sakit', 'kenya', 'makaroni', 'pintar', 'tulang', 'jikok', 'marah', 'galau', 'dasar', 'jidat', 'paripurna', 'ya', 'ampun', 'senang', 'terimakasih', 'semesta', 'hadir', 'sulit', 'banget', 'budak', 'cinta', 'lucu', 'banget', 'lari', 'gemas', 'capek', 'kuat', 'lihat', 'mentega', 'sankyu', 'selesai', 'tugas', 'nyaman', 'rumah', 'hinga', 'lupa', 'kenyat', 'tonton', 'drama', 'korea', 'kenya', 'tenang', 'pikir', 'gara', 'vincenzo', 'cari', 'jiwa', 'mafia', 'muka', 'imut', 'senang', 'yongi', 'sore', 'bahagia', 'kapal', 'senang', 'banget', 'enak', 'banget', 'tahan'	Negative

Table 10 shows that the model's predictions can accurately and successfully predict a sample of Twitter users. Tweets are explicit that tweets from username1 are negatively labeled while tweets from username2 are favorably labeled. Based on table 10, username1 with a positive label is depressed, while username2 with a negative label is not depressed.

4. CONCLUSION

Based on this research there are several conclusions that can be drawn from this research based on the results and discussion. First, the BiLSTM method with wor2vec extraction can be used to identify Twitter users' depression. In this study, we used hyperparameters with batch size 32, epoch 8, and 256 neurons in the hidden layer, which resulted in accuracy and f1-score on training data of 85.82% and 88.75%, and accuracy and f1-score on test data of 53.12% and 68.08%, respectively. Second, It can be concluded that the size of the number of neurons and batch size affects the accuracy and f1-score adjusted to the number of datasets. Third, an accuracy of 90% proves that the labeling results using DASS-42 on the questionnaire and the tweet classification results using the BiLSTM method correlate. According to the evaluation findings, this model is more likely to categorize data with positive labels than negative

Building of Informatics, Technology and Science (BITS)
Volume 4, No 2, September 2022 Page: 369−376
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i2.1850

labels due to the unbalanced labels between the two. For future work, we will try to expand the dataset and use a balanced dataset to get better accuracy and improve the quality of the model. We added a library of slang and class characteristics. Classes can be more triggering for stress and anxiety than depression or non-depression.

REFERENCES

[1] S. Mutmainah, "KEMUNGKINAN DEPRESI DARI POSTINGAN PADA SOSIAL MEDIA," *J. Sains, Nalar, dan Apl. Teknol. Inf.*, vol. 1, no. 2, 2022.

[2] WHO, "Mental health," *WHO*, 2021. https://www.who.int/health-topics/mental-health (accessed Oct. 16, 2021).

[3] A. A. Rachmawati, "Darurat Kesehatan Mental bagi Remaja - Environmental Geography Student Association," *egsaugm*, 2020. https://egsa.geo.ugm.ac.id/2020/11/27/darurat-kesehatan-mental-bagi-remaja/ (accessed Oct. 16, 2021).

[4] D. S. Oktarini, "Curhat di Twitter Saat Depresi, Pria Ini Ditemukan Bunuh Diri," *suara.com*, 2019. https://www.suara.com/tekno/2019/01/28/103735/curhat-di-twitter-saat-depresi-pria-ini-ditemukan-bunuh-diri?page=all (accessed Dec. 07, 2022).

[5] S. R. Kamite and V. B. Kamble, "Detection of Depression in Social Media via Twitter Using Machine learning Approach," 2020. doi: 10.1109/ICSIDEMPC49020.2020.9299641.

[6] F. M. Shah *et al.*, "Early Depression Detection from Social Network Using Deep Learning Techniques," 2020. doi: 10.1109/TENSYMP50017.2020.9231008.

[7] A. R. Isnain, A. Sihabuddin, and Y. Suyanto, "Bidirectional Long Short Term Memory Method and Word2vec Extraction Approach for Hate Speech Detection," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 14, no. 2, 2020, doi: 10.22146/ijccs.51743.

[8] E. Kang, "Long Short-Term Memory (LSTM): Concept," *Medium.com*, 2017. https://medium.com/@kangeugine/long-short-termmemory-%0Alstm-concept-cb3283934359 (accessed Oct. 14, 2021).

[9] M. Soni, "Understanding architecture of LSTM cell from scratch with code.," *Medium.com*, 2018. https://medium.com/hackernoon/understanding-architecture-of-lstm-cell-fromscratch-%0Awith-code-8da40f0b71f4 (accessed Nov. 18, 2021).

[10] S. Kusumadewi and H. Wahyuningsih, "Model Sistem Pendukung Keputusan Kelompok untuk Penilaian Gangguan Depresii, Kecemasan dan Stress Berdasarkan DASS-42," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 2, 2020, doi: 10.25126/jtiik.2020721052.

[11] K. S. Nugroho, "Dasar Text Preprocessing dengan Python," *Medium.com*, 2019. https://ksnugroho.medium.com/dasar-text-preprocessing-dengan-python-a4fa52608ffe (accessed Nov. 27, 2021).

[12] C. Khanna, "Text pre-processing: Stop words removal using different libraries," *Medium.com*, 2021. https://towardsdatascience.com/text-pre-processing-stop-wordsremoval-%0Ausing-different-libraries-f20bac19929a (accessed Nov. 28, 2021).

[13] A. Firdaus, "Cara Kerja Word2vec," *Medium.com*, 2019. https://medium.com/@afrizalfir/mengenal-word2vec-af4758da6b5d (accessed Nov. 29, 2021).

[14] A. M. Ertugrul and P. Karagoz, "Movie Genre Classification from Plot Summaries Using Bidirectional LSTM," in *Proceedings - 12th IEEE International Conference on Semantic Computing, ICSC 2018*, 2018, vol. 2018-January. doi: 10.1109/ICSC.2018.00043.

[15] P. Bahad, P. Saxena, and R. Kamal, "Fake News Detection using Bi-directional LSTM-Recurrent Neural Network," in *Procedia Computer Science*, 2019, vol. 165. doi: 10.1016/j.procs.2020.01.072.

[16] S. Narkhede, "Understanding Confusion Matrix," *Medium.com*, 2018. https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62 (accessed Nov. 27, 2021).

[17] H. N B, "Confusion Matrix, Accuracy, Precision, Recall, F1 Score," *Medium.com*, 2019. https://medium.com/analytics-vidhya/confusion-matrix-accuracyprecision-%0Arecall-f1-score-ade299cf63cd (accessed Nov. 27, 2021).

[18] N. L. P. P. Waisnawa, D. Nurjanah, and H. Nurrahmi, "Cyberbullying Detection on Twitter using Support Vector Machine Classification Method," *Build. Informatics, Technol. Sci.*, vol. 3, no. 4 SE-Articles, Mar. 2022, doi: 10.47065/bits.v3i4.1435.