JINAV: Journal of Information and Visualization *ISSN: 2746-1440 (online)*
Vol. 3 No. 2 (2022)
https://doi.org/10.35877/454RI.jinav1503

Depression Detection on Twitter Social Media Platform using Bidirectional Long-Short Term Memory

Andre Agasi Simanungkalit*, Warih Maharani, & Prati Hutari Gani

Informatics Study Program, Informatics Faculty, Telkom University Bandung, Jawa Barat, 40257, Indonesia

Abstract

Depression is one of the mental disorders that are often experienced by a person in daily life. Social media platforms is a new thing as an alternative to tell stories and express current feelings by people today. Twitter is one of the social media that is often used to express feelings and opinions through tweets posts, including tweets that contain hate speech which indirectly shows symptoms of depressive disorder through statements uploaded. It also requires modeling that can recognize users with the potential to experience depression so that they can get initial treatment. This can be implemented using the BiLSTM (Bidirectional Long Short-Term Memory) method and the Word2Vec feature. It is also needs a modeling that can recognize the users who have the potential to experience depression so that they can get treatment at the beginning. This can be implemented using the BiLSTM (Bidirectional Long Short-Term Memory) method and the Word2Vec feature.

Keywords: Twitter, mental illness, BiLSTM, depression.

1. Introduction

1.1. Background

Depression is a disorder of a person's emotional characterized by negative thinking about everything (Sulistyorini & Sabarisman, 2017). Quoting from Jaka Arya Pradana (2016) who stated that depression is also referred to as an invisible disorder (Dirgayunita, 2016). Based on WHO data in 1980, almost 20% - 30% of hospital patients in developing countries experience emotional mental illnes, such as depression. This disorder can be experienced by all age groups (Yoeyoen, 2018). Data from *Riskesdas* 2018 results show that depressive disorders have started to occur since the age range of adolescents (15-24 years), with a prevalence of 6,2%. The prevalence pattern of depression is getting increase, the highest at 75+ years old at 8,9%, 65-74 years old at 8,0%, and 55-64 years old at 6,5%.

In the current era of the Industrial Revolution 4.0, technology continues to develop and supports life, one of which is in expressing oneself. There are many ways how people express themselves in this digital era, such as expressing the content of their feelings on social media platforms. The dominant social media platform chosen is Twitter. On this social media, depression person can pour his anxiety on a personal account or base account with a 280-character tweet that can certainly be uploaded every day (Zi'ni, 2020). In addition, another Twitter phenomenon that is currently well known is Autobase (Maulina, 2021). [5] It is an account that allows Twitter users to send messages or called *menfess* (mention and confess) in the form of information or questions anonymously through direct messages (dm) of the account's profile which will automatically be spread to their timeline (Sipahutar et al., 2020). Reporting from the CSSMORA UINSA, various menfess websites that is commonly found on Autobase accounts is information about certain topics, general questions, and confide personal problems. One of the characteristics of a perpetrator who has a depressive mental disorder can be easily seen through his dominant post tweets using negative words, such as feelings of worthless, guilty, and hating himself.

Previously, research related to the detection of a person's depression has been widely carried out. In the study conducted between the Support Vector Machine (SVM) method which was compared with Bidirectional Encoder Representations from Transformers (BERT) and A Lite BERT (ALBERT), the performance of the BERT model has the highest value

* Corresponding author.
E-mail address: andreagasi@students.telkomuniversity.ac.id

with an accuracy value of up to 75%. Another related study between the Naïve Bayes (NB) method and the Support Vector Regression (SVR) which showed that the performance of the SVR method had the highest accuracy value of 79,7% (Isnain et al., 2020). Then, Bidirectional Long Short Term Memory (BiLSTM) is a two-way of LSTM that is able to capture information without neglecting the context and meaning of a sentence. The Bidirectional Long Short Term Memory (BiLSTM) approach is one of the commonly used variants of the LSTM model. In this model variant, the incoming information is only concerned in the last word because it only reads sentences with a one-way version (Isnain et al., 2020). The BiLSTM model variant is a development of the LSTM model variant which technically uses two separate LSTM layers (forward and backward). These two layers read the incoming information from two directions at once and the output is generally combined into one. Through this layer, the model can learn past information (past) and future information for each input sequence. BiLSTM's ability to read incoming information from two directions makes the performance of this method more accurate and higher than all standard baseline models and LSTM because tweets can be understood sequentially every word with the highest accuracy reaching 94,12%. Another study related to the use of the BiLSTM method is in the detection of hate speech which resulted in the highest accuracy of 94.66%. Therefore, this study uses the BiLSTM method since the previous study has a high level of performance and accuracy. This research is expected to detect people who have depressive disorders through tweets of a Twitter user.

1.2. Topics and Limitations

The formulation of the problem is able to detect a person who is indicated to have symptoms of depression on the social media platform Twitter and factors that affect the BiLSTM model in detecting depression and recognize the level of accuracy given by BiLSTM method so that early treatment can be done on the user. Meanwhile, the limitations of the problem in this study involved Twitter tweet posts data used Indonesian language taken from user tweets on Twitter social media. The model used in detecting depression is Depression, Anxiety, and Stress Scales (DASS-42). It did not detect *emoji* and typo problems.

1.3. Objective

The objective of this study is to detect depression in a person on Twitter social media by using the BiLSTM method and Word2vec extraction. Table 1 presents the correlation between objectives, testing, and oconclusions.

Table 1. The Correlation between Objectives, Testing, and Conclusion

No	Objective	Testing	Conclusion
1	Depression Detection using Bidirectional Long Short Term Memory method	Testing performed by inputting a user's tweet dataset	Models can display results in detecting user tweets
2	What factors influence the model in detecting depression on Twitter social media	Through Confusion Matrix Evaluation	The results showed that the testing accuracy point was obtained by 60% with an F-1 Score of 63%
3	How is the level of accuracy in detecting depression using the BiLSTM method	Testing performed with and without using the BiLSTM method	Analysis shows that tuning BiLSTM parameters may affect results. In addition, split data experiments and Word2vec affect the results obtained.

2. Related Studies

2.1. Depression

Depression is a disorder of the general state of emotional tone characterized by a sense of sadness, apathy, pessimism, and loneliness. The term depression was first introduced by Meyer (1905) to describe a mental illness with major symptoms of sadness, which is accompanied by other psychological symptoms, somatic (physical) disorders, and psychomotor disorders over a period of time. It is classified as affective disorder (Mardiya, 2010). There are at least five factors that are known causing a person to experience depression, including psychological factors (where the sufferer loses the object of love), biological factors (due to the lack of neurochemical factors, namely *mono-amine* *neurotransmitters), neoro-immunological* factors (an infection of the central nervous system occurs), genetic factors (caused by hereditary factors, and psychosocial factors (caused by the family environment (broken home, low family economy etc.). Depressive disorders can cause sufferers to experience serious disturbances in social functioning, the quality of life where sufferers feel useless, and intention to death due to suicide. Proper and fast handling needs to be done immediately to determine the next treatment.

2.2. Twitter

Twitter is one of the social media networks that is popularly used by people in Indonesia (Annur, 2022). According to a *Statista* report, there are 18,45 million application users founded by Jack Dorsey in the country as of January 2022 (Annur, 2022). This achievement ranks Indonesia as the 5th most Twitter user country in the world. In the era of current's development, Twitter is one of the places where you pour out the most selected feelings. Twitter users can lock their account or make it open to the public and can also hide their real identity (anonymous). Basically, Twitter only provides a tweet limit with a maximum of 280 characters which makes it difficult for users if they want to discuss a fairly long topic (Maulina, 2021). The thread feature or commonly known as a connecting story can be used by the users to do a collection of continuous tweets that will be very long information. A term that is no less known to many people next is Autobase. It comes from the word of automatic fanbase which is anonymous and becomes a place for followers to ask and convey all problems anonymously. The way autobase works is that followers will send submissions via DM of the account and will be automatically uploaded to the public anonymously. The reason for disclosing the content of feelings into social media Twitter is to release the burden and feelings without expecting interpersonal and social rejection, or intrapersonal difficulties, because it is created in anonymous media.

2.3. Depression Anxiety Stress (DASS-42)

(Kusumadewi & Wahyuningsih, 2020) Depression, Anxiety and Stress Scales (DASS) is a psychological measurement tool consisting of 42 items that measure general psychological distress, such as depression, anxiety, and stress. DASS-42 was developed by Lovibond and Lovibond in 1995 translated by Damanik (2011). There are three scales of 14 items in each test which are further divided into several sub-scales consisting of 2 to 5 items that are estimated to measure the similar thing. The following is a classification of assessments based on DASS-42 in the table:

Table 2. Categorization and Interpretation of DASS-42 Subscla Scores

Interpretation	Subscales		
	Depression	Anxiety	Stress
Normal	0-9	0-7	0-14
Mild	10-13	8-9	15-18
Moderate	14-20	10-14	19-25
Severe	21-27	15-19	26-33
Extremely Sever	28-42	20-42	34-42

2.4. Word2Vec

(Nurdin et al., 2020) Word2vec is one of the word embedding algorithms created by Mikolov *et al.* in 2013 that mapped every word in form of the text into a vector (Firdaus, 2019). For a simple example, the representation vector of the word "*Jeruk*" would be adjacent to the "*Asam*" vector just as the "*Sabun*" vector would be adjacent to the "*Basa*" vector. In other words, the word2vec model will grasp the meaning that "*Lemon*" and "*Asam*" have the same correlation as "*Sabun*" and "*Basa*" i.e. the solution indicator (Firdaus, 2019). Word2vec has 2 architectures, namely:

- Skip-gram

  The skip-gram architecture works by predicting the context (output) around the current word (input)

- Continuous Bag of Word (CBOW)

The CBOW architecture works by (the opposite of skip-gram) which predicts a word (output) when given the context around that word (input).

2.5. LSTM

Rahmadzani (2021) Long short term memory network is one of its original form modification of recurrent neural network or RNN where LSTM is able to predict words based on past information stored for a long time. Thus, LSTM is able to remember a collection of information that has been stored for a long period of time while deleting information that is no longer relevant (Rahmadzani, 2021). Figure 1 shows the structure of LSTM with a single layer.

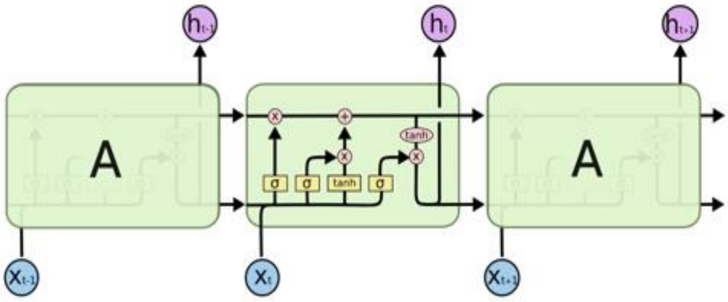

Figure 1. Single Layer LSTM Structure

The following is an explanation of the *gate* and how LSTM works (Prijono, 2018; Rahmadzani, 2021):

1) *Forgot Gate*

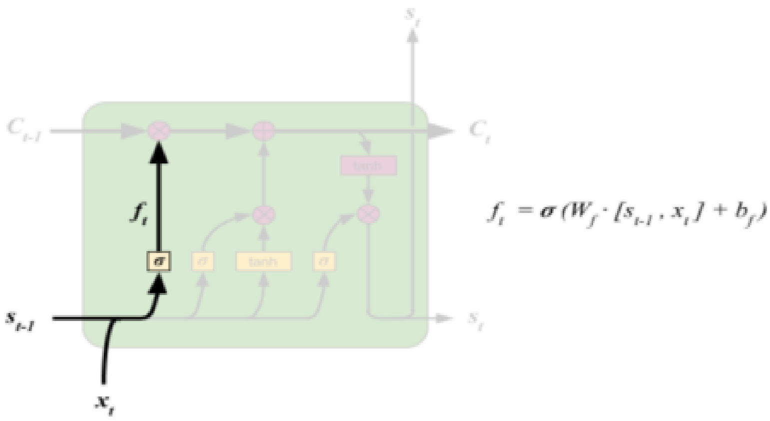

Figure 2. Forgot Gate Architecture (Prijono, 2018)

LSTM decides how much information removed from the context of c_{t-1} through a sigmoid gate called a forgot gate. Then, forgot Gate will read the value of S_{t-1} and will result in the number 0 (element needed) or 1(element discarded) in each element in c_{t-1}. The forget gate value can be calculated by the following equation:

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

Note: the notation [s t-1,x t] is a concatenation operation; meaning we add a line from xt with a line from s $_{t-1}$.

2) *Input Gate*

After that, the input gate will decide on new information that will be used on C_t. This process consists of two parts. First, the sigmoid gate (input gate) will decide which value to be updated. Then, a tanh layer produces a new context vector candidate, C̃t (read: C tilde) and the two are combined to make updates to the context later. The following is the equation of input gate and tanh, namely:

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

$$\tilde{C}_t = tanh\ (W\text{c} \cdot [s_{t-1}, x_t] + b_c) \quad (2.3)$$

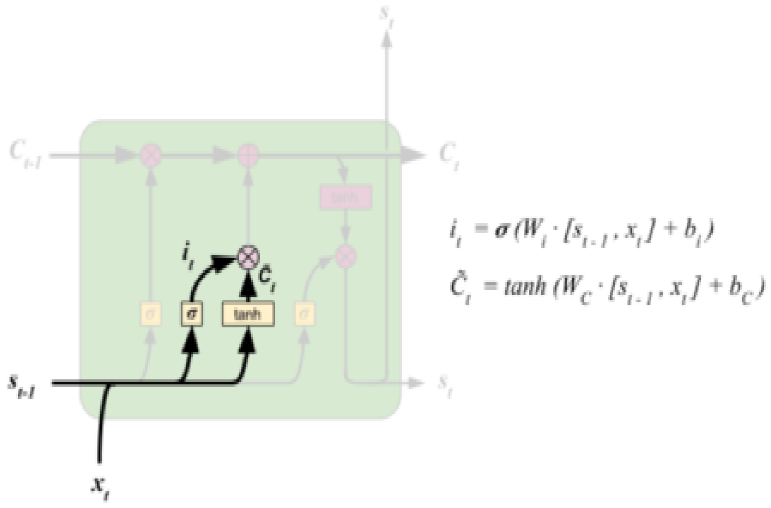

Figure 3. Input Gate Architecture (Prijono, 2018)

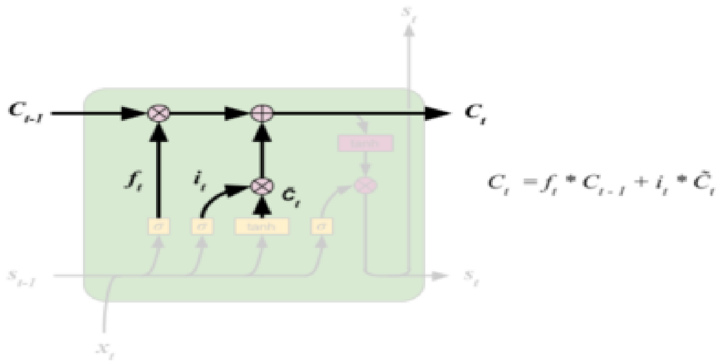

Figure 4. Input Gate Architecture (Prijono, 2018)

Now, it's time to update the old context of C_{t-1} to the new context of C_t by multiplying the old context (C_{t-1}) with the *forget gate* (f_t) to remove unneeded information. After that, the candidate is multiplied by the new context ($\tilde{C}t$) and by the *input gate* (i_t) to decide how much we will include the new context candidate and both are added. The equation used is as follows:

$$C\,t = f\,1^{*}\ C_{t-1} + i_t{}^{*}\ (2.4)\tilde{C}$$

3) *Output Gate*

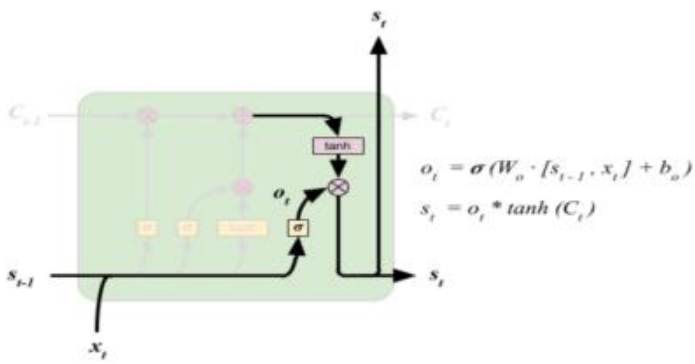

Figure 5. Output Gate Architecture (Prijono, 2018)

The output gate will decide which information to output. This output will be based on a value in our context and passed to a filter. First, we run the sigmoid gate which we call the output gate (output gate, $\mathbf{o}_t$) to decide what parts of the context we are going to generate. Then, we pass the context through tanh to make the value between −1 and 1, and we multiply it by the output of the its sigmoid gate so that we only generate the part we decide. The equation used is as follows:

$$O\,t = \sigma\,(W\,\mathrm{o}\ \cdot [s_{t-1}, x_t] + b_o)$$

$$S\,t\ = O\,\mathrm{t}\ ^{*}\ tanh\ (C_t)$$

2.6. BiLSTM

BiLSTM is a development of the LSTM model where the process is in the opposite direction. Each word in the document is processed sequentially so that the user tweets can be understood sequentially every word. (Isnain et al., 2020;Rizky, 2021) There are two layers in BiLSTM, the layer below moves forward (LSTM *forward*), which is understanding and processing from the first word to the last word while the layer above moves *backwards* (LSTM backward), which is understanding and processing from the last word to the first word. Through these opposing two-way layers, the model can understand and take perspectives from the previous word and the leading word, so that the learning process will deepen and has an impact on the model. It will better understand the context of the tweet. The BiLSTM architecture can be seen in the Figure 6.

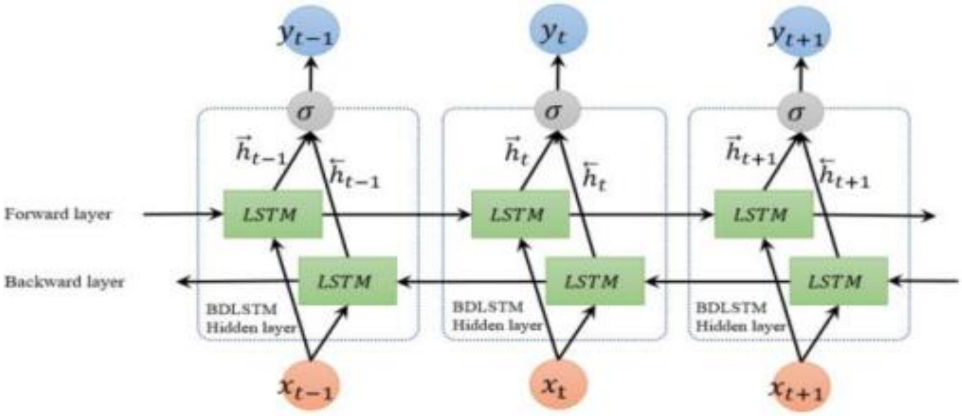

Figure 6. BiLSTM Architecture

2.7. Confusion Matrix

Confusion matrix is a table used to measure the performance of the classification model used. (Rizky, 2021) In the confusion matrix, there are four values that present the results of the classification process (Rizky, 2021). These values can be seen in the table 3.

Table 3. Confusion Matrix

Confusion Matrix		**Real Value**	
		Positive	Negative
Predicted Value	Positive	TP	FN
	Negative	FP	MR

where:

- TP (True positive): Predictions that are positive and correct on target.
- TN (True negative): Prediction that is negative and correct on target.
- FP (False positive): Predictions that are positive and false are not on target.
- FN (False negative): Predictions that are negative and false are not on target.

(Arthana, 2019; Nugroho, 2019) Performance metrics and confusion matrices above can be used to measure the performance of the model we create. Some models are Accuracy, Precision, Recall, and F-1Score. Here's an explanation and formula to find out (Hidayatullah, 2021; Rizky, 2021):

1) Accuracy

Accuracy is a model that describes how accurately a model classifies correctly.

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

2) Precision

Precision is a model that describes the accuracy between the requested data and the predicted results given by the model.

$$precision = \frac{TP}{TP+FP}$$

3) Recall

Recall is a model that describes the success of a model in reinventing information.

$$recall = \frac{TP}{TP+FN}$$

4) F-1 Score

The F 1-Score is a model that describes the comparison of mean precision and recall.

$$F\text{-}1\ Score = \frac{2\ x\ Recall\ x\ Precision}{Recall+Precision}\ x\ 100\%$$

3. Built System

3.1. General Description of the System

In the system design stage, what has done is to be able to detect depression in tweets on Twitter using the BiLSTM method and extraction using the *word2vec* feature. The design of the system can be seen in Figure 7.

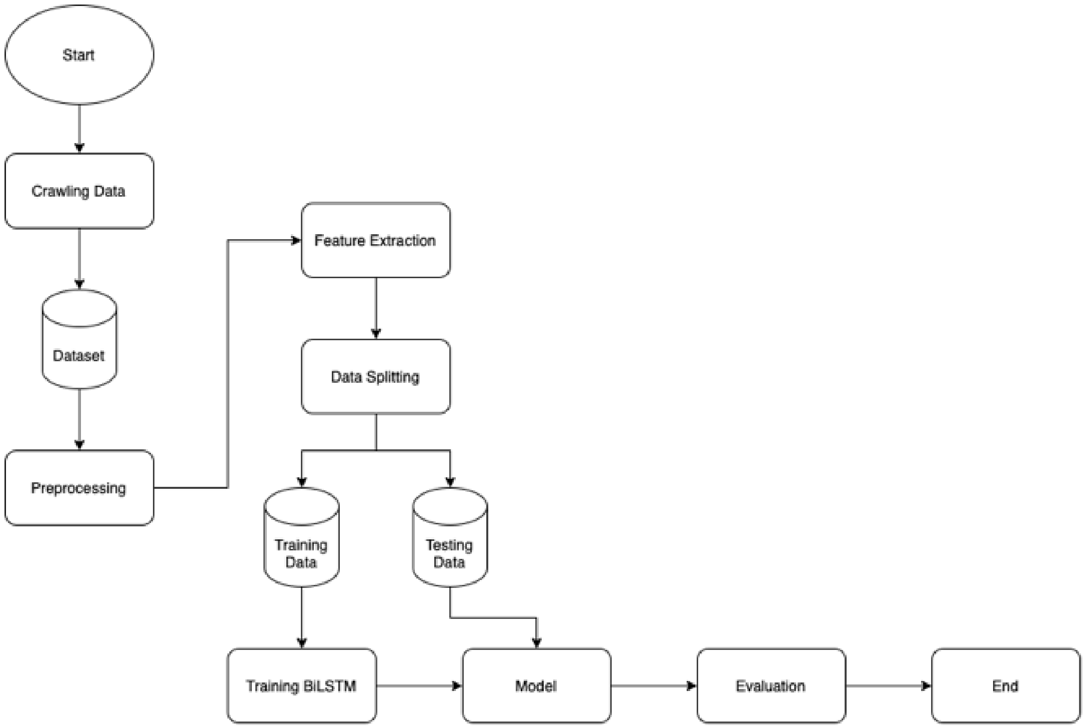

Figure 7. Flowchart System

Based on Figure 7, the system flow begins by collecting Twitter data. The data are collected and stored in the form of csv. The next process data will be processed at the preprocessing stage which consists of case folding, normalization, stopword removal, stemming and tokenizing data. Furthermore, move to the feature extraction process using the word2vec feature aims to convert words into a vector and after becoming a vector, the data is divided into 2 parts, namely train data and test data. The next process, the train dataset and testing datasets will be modeled testing. The data that has been tested by the model will produce results, namely detection and system evaluation.

3.2. Data Retrieval

1) DASS-42 Questionnaire Deployment

In this study, respondents were given a questionnaire form based on DASS-42 to find out their state and mentality. There were 42 questions that had been divided for each item number for symptoms of depression, anxiety, and stress. The obtained data was then saved in a csv with the file format. As a result of filling out the questionnaire form, there were 162 accounts that had filled out the questionnaire according to 42 questions based on DASS-42. According to Table 2 of the division of each item number for symptoms of depression, anxiety, and stress, users with a total score between 0-9 were indicated to have no depressive symptoms or normal, while users with a score of 10+ were indicated to have depressive symptoms. We used the depression category according to the results of the study in detecting depression.

2) Labeling Data

After the data was successfully collected, the researcher used two information namely tweets and usernames with the total data collected was 162 users provided in csv format. In this study, the dataset was divided into two, namely negative and positive. The negative label indicated that the user was not indicated depressed, while the positive label indicated that the user was indicated depressed. Labeling by dividing positive and negative was carried out according to the seriousness of the depressive condition scores in Table 2 of DASS-42, 0-9 for negative and 10+ for positive.

3.3. Preprocessing Data

Preprocessing data is a technique of transforming raw data in a more useful format. This process is necessary to correct errors in raw data that are often incomplete and have an irregular format. At this stage, *case folding* was carried out, namely changing the *character* with capital letters to lowercase. *Normalization* is the process of changing the word format because some *tweets* from users often contain words that are difficult to define and different from the original meaning. This process was also carried out a cleaning process such as symbols, *re-tweeting, username, url, extra space*, and *non-aphanumeric*. Then, *stopword removal* was used to remove unimportant words by using the *libraries nltk*, *stemming* reverts affixes into original words and *tokenization* in separating sentences into a series between words. The process and results of *tweet preprocessing* can be seen in Table 4.

Table 4. The Preprocessing Process

User	Input	Preprocessing	Output
User 42	@koalakumal123 *Keknya related sih Di film "Susah Sinyal" kalo mau nenangin diri larinya ke Lombok. Emang bener sih.* https://t.co/w2YzzENkIQ @nisalvna Temenku ulang tahun ucapin apa ya @Faregi99 Mana ga pake semvak lagi hahaha @mentallycracked @lutfisirajs @radenrauf Ada gengsi yg harus dijaga Â	***Case Folding***	@koalakumal123 *keknya related sih di film susah sinyal kalo mau nenangin diri larinya ke lombok. emang bener sih* https://t.co/w2yzzenkiq @nisalvna t*emenku ulang tahun ucapin apa ya* @faregi99 *mana ga pake semvak lagi hahaha* @mentallycracked @lutfisirajs @radenrauf *ada gengsi yg harus dijaga â*

User	Input	Preprocessing	Output
	@koalakumal123 *keknya related sih di film susah sinyal kalo mau nenangin diri larinya ke lombok. emang bener sih* https://t.co/w2yzzenkiq @nisalvna *temenku ulang tahun ucapin apa ya* @faregi99 *mana ga pake semvak lagi hahaha* @mentallycracked @lutfisirajs @radenrauf *ada gengsi yg harus dijaga â*	***Normalization***	*sepertinya related sih di film susah sinyal kalau mau menenangkan diri larinya ke lombok memang benar sih temanku ulang tahun mengucapkan apa ya mana tidak pakai semvak lagi haha ada gengsi yang harus dijaga*
	sepertinya related sih di film susah sinyal kalau mau menenangkan diri larinya ke lombok memang benar sih temanku ulang tahun mengucapkan apa ya mana tidak pakai semvak lagi haha ada gengsi yang harus dijaga	***Stopword Removal***	*sepertinya related film susah sinyal kalau mau menenangkan diri larinya ke lombok memang benar temanku ulang tahun mengucapkan apa mana tidak pakai semvak lagi ada gengsi yang harus dijaga*
	sepertinya related film susah sinyal kalau mau menenangkan diri larinya ke Lombok memang benar temanku ulang tahun mengucapkan apa mana tidak pakai semvak lagi ada gengsi yang harus dijaga	***Stemming***	*seperti related film susah sinyal kalau mau tenang diri lari ke lombok memang benar teman ulang tahun ucap apa mana tidak pakai semvak lagi ada gengsi yang harus jaga*
	seperti related film susah sinyal kalau mau tenang diri lari ke lombok memang benar teman ulang tahun ucap apa mana tidak pakai semvak lagi ada gengsi yang harus jaga	***Tokenization***	[*'seperti', 'related', film, 'susah', 'siynal', 'kalau', 'mau', 'tenang', 'diri', 'lari', 'ke', 'lombok', 'memang', 'benar', 'teman', 'ulang', 'tahun', 'ucap', 'apa', 'mana', 'tidak', 'pakai', 'semvak', 'lagi', 'ada', 'gengsi', 'yang', 'harus', 'jaga'*]

3.4. Implementation of Word2Vec Features

After the process of *data preprocessing*, the next process is feature extraction using *Word2vec*. It is a useful word embedding method for representing words as vectors. *Word2vec* consists of three layers namely *input*, *projection*, and *output* and also two types of *neural network* architecture from *word2vec* namely "*skip-gram*" (SG) and "*continous bag of word* " (CBOW). Here, the researcher have set the *word2vec* model settings with vector size of 300, 5 window size, 32 *epoch*, and 10 *min-count*. Table 5 shows a vector representation of the word "*sedih*" with a vector length of 300. The results of feature extraction can be seen in the Table 5.

Table 5. *Word2vec* Vector Result

Word	***Word2Vec* Vector**
Sedih	-1.03839617e-02, 4.05715495e 01,....,3.19459558e-01

In representing a word, *word2vec* implements a neural network to calculate the contextual and semantic similarity of each word (input) in the form of a *satu-panas* encoded vector. The result of this contextual and semantic similarity can represent the correlation of one word to another. In Table 6, you can see the vector value of the word "*sedih*" with several other terms related to the word "*sedih*".

Table 6. *Word2vec* Semantic Similarity

Words	Semantic Similarity
Sedih	- '*pikir*' , 0.955095648765564 - '*parah*' , 0.9428145289421082 - '*hidup*' , 0.932928204536438 - '*pergi*' , 0.9288321137428284 - '*sesal*' , 0.9286637306213379 - '*bosan*' , 0.9271129369735718 - '*sering*', 0.9238063097000122 - '*stres*' , 0.9219777584075928 - '*jujur*' , 0.9170176982879639 - '*akan*' , 0.9155300855636597

3.5. BiLSTM Classification

1) Embedding Layer

Embedding Layer is the layer responsible for converting a token into its vector representation generated by the *word2vec* model. Embedding matrices are the matrix of all words and embeddings accordingly. The researcher used the embedding matrix in the embedding layer of our model to attach ta token into its vector representation, which contains information about the token or words. The embedding vocabulary is obtained from the *tokenizer* and the corresponding vector of the embedding model, which is the *word2vec* model. The data that goes into the *embedding layer* process has a size of 300 dimensions after becoming a vector due to the extraction of features from *word2vec*.

2) Bidirectional LSTM

The LSTM Bidirectional classification process was started with *Input layer*, *LSTM layer*, *Conv1D layer*, and *Output layer*. The *Input layer* then received data consisting of 636 unique vocabulary words and a maximum of 2000 words for all text documents. *The input layer* entered LSTM *layer* which the result will use the *Bidirectional* LSTM process with a dimension size of 100 *neurons* and *dropout* 0.2. Next, entered to the *Conv1D layer* containing the *Conv1D* process to increase or decrease the intensity of the value with a dimension size of 100 neurons. The next process was *GlobalMaxPool1D* which is a polling method where the *global magic polling layer* reduced the *sample* of all *feature maps* to one value. This would be the same as setting the join size to the input feature map size. The result will go to the *output layer* where this layer has 2 *dense layers*, namely *dense* with the *ReLU* activation function which has 16 dimensions and *dense* with the *Sigmoid* activation function which has 1 dimension. The model must be compiled to determine *loss*, *metrics*, and *optimizers*. Determining *the right loss* and *metrics* was critical when training a model. *Binary Cross-entropy* is a *loss function* that the researcher used. For *metric*, the researcher have accuracy and *adam optimizer* as an optimization algorithm for *gradient descent*.

4. Evaluation

4.1 Test Results

The baseline parameters used can be seen in the Table 7.

Table 7. Baseline parameter model

Large Feature W2V	LSTM Layer	Conv1D Layer	Dropout	ReLU	Sigmoid
100	100	100	0.2	16	1

Figure 8 shows the data measurement of data sharing used in testing the baseline model.

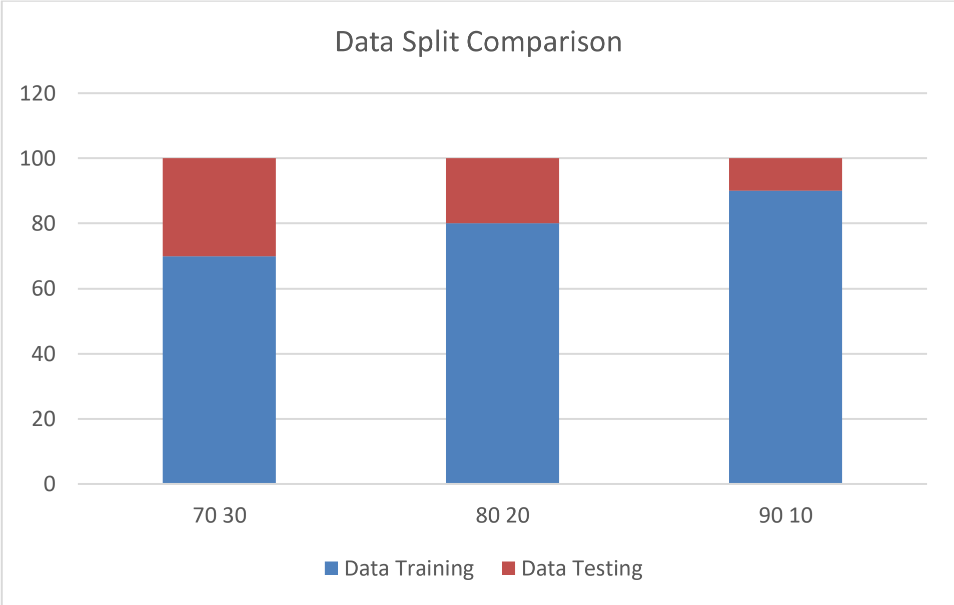

Figure 8. Data sharing size

The result of comparison between data split that can be seen in Table 8.

Table 8. Model Result

Large Featured W2V	Data Split Comparison	Precision	Recall	F-1 Score	Training Accuracy	Testing Accuracy
200	70:30	49%	88%	63%	62. 50%	46. 94%
200	80:20	59%	72%	65%	67.97%	57.58%
200	90:10	70%	64%	**67%**	52.78	**58.82**

In Table 8, it can be seen that the best produced dataset size accuracy is 90:10 with an accuracy testing of 58.82%. Next, the researcher conducted an experiment by comparing between 2 large *feature* dimensions of *word2vec*. *The size parameters of the word2vec feature* that the researcher selected the experiment were dimensions100 and 300 using the best *testing* accuracy in Table 8, namely split comparison data of 90:10. Differences of size feature *Word2vec* used and the spliting data will produce different results. It can be seen in Table 9.

Table 9. Experiments on Word2vec's Large Featured Dimensions

Large Featured W2V	Data Split Comparison	Precision	Recall	F-1 Score	Training Accuracy	Testing Accuracy
100	90:10	57%	73%	64%	53. 47%	47. 06%
300	90:10	65%	64%%	**70%**	59.94%	**64.71%**

Based on the Table 9, the highest accuracy testing results are at dimension 300 with an accuracy of 64.71%. Furthermore, with dimension parameters and *split comparison data* from the highest testing accuracy results, the researcher continued the experiment by testing the model through setting the *hyperparameter* category on the *BiLSTM* model dimensions. Here, the researcher set the parameters on the *LSTM* and *ConV1D* layers with experimental dimensions of 64,128 and 246 with batch sizes of 16 and 32 and with *epoch* 10 as the subject of the experiment as well as the size of the *split data* and *word2vec* dimensions according to the highest result of Table 8 which is 300 dimensions. The results and comparison between *hyperparameters* can be seen in the Table 10.

Table 10. Experiments on Batch Size and Neurons

Batch Size	Neurons	Precision	Recall	F1-Score	Training Accuracy	Testing Accuracy
16	64	60%	82%	69%	67.36%	52.94%
	128	62%	73%	67%	73. 61%	52. 94%
	246	71%	91%	80%	52. 08%	70. 59%
32	64	86%	55%	67%	54. 86%	64. 71%
	128	65%	61%	79%	54.17%	64.71%
	246	54%	36%	53%	55.56%	58. 82%

Based on Table 10, it is found that *batch size* 16 with neuron 246 produced *a testing* value with the best accuracy of 70.59% with an F1 score of 80%. As a result, the researcher concluded that the best accuracy in testing model was in *batch size*16 with dimensions of 246 testing accuracy of 70.59% and F1-score of 80%. Table 11 represents a confusion matrix of the best testing accuracy.

Table 11. Confusion Matrix from Experiment 3

Confusion Matrix		**Its True Value**	
		Positive	Negative
Predicted Value	Positive	58. 82%	5. 88%
	Negative	23. 53%	11. 76%

Based on Table 10, confusion matrix with batch size 32 and neuron 128 shows as many as 58. 82% of users are detected depressed and completely depressed (True Positive), and as many as 11.76% of users are not depressed (True Negative), 5.88% of users were declared depressed but the diagnosis was not depressed (False Positive), and as many as 29.41% of users identified not depressed but they were actually depressed (False Negative). Next, the researcher prediction whether the user is depressed or not can be seen in Table 12.

Table 12. Model predictions

Username	Tweets	Predict
User 42	*'seperti', 'related', 'film', 'susah', 'sinyal', 'kalau', 'mau', 'tenang', 'diri', 'lari', 'ke', 'lombok', 'memang', 'benar', 'teman', 'ulang', 'tahun', 'ucap', 'apa', 'mana', 'tidak', 'pakai', 'semvak', 'lagi', 'ada', 'gengsi', 'yang', 'harus', 'jaga'*	Positive

Table 12 shows that testing was successful in detecting depression in tweets using the best testing accuracy from the experimental results. The researcher also compared the tests carried out on this modeling with the DASS-42 label where it was stated that User 42 was declared depressed and in accordance with the model the researcher conducted.

4.2 Analysis Test Results

Initial testing of the model was carried out using the following parameters, word2vec large feature 200, batch size model 64, LSTM layer 100, Conv1D layer Day 100, and dropout 0.2 with 3 different dataset sizes of 70:30, 80:20, and 90:10. First, from the results of testing the base line model, it was found that the size of split 90: 10 data produced the best testing accuracy with an accuracy of 58.82% and f1-score of 67%. So that the researcher used the best accuracy results to conduct further experiments. Data split size of 90:10 produced the best accuracy because the more amount of training data, the accuracy also increases because of deep learning models, such as BiLSTM that will keep to learn and improve accuracy. Second, the experiment was to compare 2 dimensional measurements of large feature on word2vec where the vector dimensions used as a comparison were 100 and 300. Previous research by (Mikolov et al., 2013) stated that a vector dimension of 300 has an optimal result. Each word could have more than one vector because it has a different context in a sentence. Therefore, the researcher wanted to use a variety of vector dimension number of 100 and kept using the vector dimension 300 to prove the correctness of the optimal results. Our experiment aims to determine the best number of vector dimensions so it can improve testing model accuracy. Table 8 shows that dimension 300 is the most optimal dimension and proves the research by (Mikolov et al., 2013) in its effect on model accuracy. This is also because the longer the vector and the more data was trained, the better model in processing the data. Meanwhile, the smaller dimensions caused the performance of the model decreased. It can be seen in the accuracy with a dimension of 100 testing accuracy of 47.06%, and dimension 200 performed in the initial test with an accuracy testing of 58.82%. Third, the next experiment was the configure hyperparameters category in the hidden layer model of BiLSTM. In determining the model batch size and experiment dimensions at each layer, there were no rules governing the number of sizes used, so the model batch size and neuron dimensions that the researcher selected in the experiment could be used to find the most optimal results. Here, the researcher set the parameters on the LSTM and Conv1Dlayers with the size of the experimental dimensions of 64,128 and 246 on each model batch size of 16 and 32. The size of split data and word2vec dimensions based on the best results of Table 8 is 300 dimensions with data split of 90:10. The best testing accuracy is on the hyperparameter model of batch size 16 with neuron 246 resulting the best accuracy of 70.59% and an f1 score of 80%. This shows an increase in performance compared to the baseline model results in Table 8. The size of the hidden layer neurons affects the accuracy of the model. This is because the hidden layer processes the input neurons and connects them to the output neurons. Thus, the number of hidden neurons will determine the value of the output produced by the LSTM unit.

5. Conclusion

According to the test results using the word2vec large feature 200 parameters, batch model of LSTM 16, dropout 0.2, LSTM layer of 246, Conv1D layer of 246, these produced accuracy and F1-score on the testing data of 70.69% and 80%. It can be concluded that the dimensional size of the large feature word2vec, LSTM, and Conv1d layers influenced the model in detecting depression which can be seen in the testing accuracy and F-1 score according to the split data used. The numbers in this research did necessarily apply to other datasets because it could be the experimental parameters that the researcher use could produce different outputs on other datasets.

The suggestion for further research is the results of this research can be used as comparison material and research references. Besides, it can be done to adjust other parameters that can be used as a comparison for better work evaluation.

References

Annur, C. M. (2022). *Indonesian Twitter Users Enter the Most List in the World, In What Order?* Databoks.Com.

Arthana, R. (2019). *Knowing Accuracy, Precision, Recall and Specificity and what is prioritized in Machine Learning*. Medium.Com.

Dirgayunita, A. (2016). Depression: Its Characteristics, Causes and Handlers. *Journal An-Nafs: Psychological Studies and Research*, *1*, 1–14.

Firdaus, A. (2019). *How Word2 Vec Works*. Medium.Com.

Hidayatullah, H. F. & A. (2021). *Identification of Cyberbullying on Twitter Social Media Using LSTM and BiLSTM Methods*.

Isnain, A. R., Sihabuddin, A., & Suyanto, Y. (2020). Bidirectional Long Short Term Memory Method and Word2vec Extraction Approach for Hate Speech Detection. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, *14*(2), 169. https://doi.org/10.22146/ijccs.51743

Kusumadewi, S., & Wahyuningsih, H. (2020). Model Sistem Pendukung Keputusan Kelompok untuk Penilaian Gangguan Depresii, Kecemasan dan Stress Berdasarkan DASS-42. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, *7*(2), 219. https://doi.org/10.25126/jtiik.2020721052

Mardiya. (2010). *Knowing Depressive Disorders in Adolescents*. Dinas Pemberdayaan Masyarakat Desa Dan Kontrol Penduduk Dan Keluarga (DPMDPPKB).

Maulina, R. (2021). *Vent Anonymously on Twitter : Come on, Be a Good Listener*.

Mikolov, T., 13MIDI, Chen, K., Corrado, G., & Dean, J. (2013). *Distributed representations of words and phrases*.

Nugroho, K. S. (2019). *Confusion Matrix for Model Evaluation in Supervised Learning*. Medium.Com.

Nurdin, A., Anggo Seno Aji, B., Bustamin, A., & Abidin, Z. (2020). PERBANDINGAN KINERJA WORD EMBEDDING WORD2VEC, GLOVE, DAN FASTTEXT PADA KLASIFIKASI TEKS. *Jurnal Tekno Kompak*, *14*(2), 74. https://doi.org/10.33365/jtk.v14i2.732

Prijono, B. (2018). *Introduction to Long Short Term Memory (LSTM) and Gated Recurrent Unit (GRU) – RNN Part 2*.

Rahmadzani, R. F. (2021). *How Long Short-Term Memory (LSTM) Works*. Rifqifai.Com.

Rizky, M. (2021). *Comparative Analysis of LSTM and BiLSTM Methods for Phonocardiogram Cardiac Signal Classification*. Final Project, Universitas Dinamika, Surabaya.

Sipahutar, C., Poerna, A., & Nurkinan. (2020). ANONYMOUS CONFIDANT COMMUNICATION EXPERIENCES FOR @18AUTOBASE FOLLOWERS ON TWITTER. *Communication Pearl Lens*, *4*, 56–74.

Sulistyorini, W., & Sabarisman, M. (2017). DEPRESI : SUATU TINJAUAN PSIKOLOGIS. *Sosio Informa*, *3*(2), 153–164. https://doi.org/10.33007/inf.v3i2.939

Yoeyoen, W. (2018). Data and Information Center of the Ministry of Health of the Republic of Indonesia. *InfoDatin-Health-Jiwa*, 1–12.

Zi'ni, L. M. and A. F. (2020). SELF-DISCLOSURE OF AUTOBASE ACCOUNT USERS TWITTER @SUBTANYARL. *Communication Science Audience*, *3*, 34–54.