Quest Journals
Journal of Electronics and Communication Engineering Research
Volume 9 ~ Issue 8 (2023) pp: 06-14
ISSN(Online) : 2321-5941
www.questjournals.org

Research Paper

Sentiment Analysis of 3 Periods of Presidential Topic Using Long Short-Term Memory (LSTM)

Yohana Karuniawati P.[1], M. Ulil Albab[2], Mohammad Nur Fawaiq[3]
[1,2,3] *Informatics Engineering Department Postgraduate Program,*
[1,2,3] *AMIKOM Yogyakarta University, Yogyakarta, Indonesia*

ABSTRACT: *Along with the development of technology, social media is also growing rapidly, and among the most popular now is Twitter. Various hot topics are often discussed on Twitter, ranging from social, cultural, legal, economic, political, and so on. A topic that was quite busy being discussed on Twitter some time ago was the issue of a 3-term presidential term, which of course invites debate from the pros and cons because each has a different opinion. Sentiment analysis is a process of processing raw text data in order to obtain information from opinions or opinion sentences in the form of sentiments. In the application of sentiment analysis, there are several methods that are often used. This study uses Long Short-Term Memory (LSTM), which is an algorithm that can be said to complement the shortcomings of RNN which cannot yet predict words from past information stored for a long time. Technical classification is divided into 3 classes, namely positive, neutral, and negative. Based on the results of testing presidential sentiment data for 3 periods from Twitter with a total of 638 clean tweets (343 positive, 223 negative, and 72 neutral) an accuracy = 75% was obtained.*
KEYWORD: *sentiment analysis, Twitter, president, LSTM.*

R*eceived 15 August, 2023;* ***R****evised 28 August, 2023;* ***A****ccepted 31 August, 2023 © The author(s) 2023.* ***P****ublished with open access at* ***www.questjournals.org***

I. INTRODUCTION

One thing that has been widely discussed by the public is the issue of a 3-term presidential term. The discourse for President Joko Widodo (Jokowi) to serve 3 terms has not ended [1]. This discourse continues to roll after being voiced by some party elites and ministers in the cabinet. Towards the end of President Jokowi's term of office his second term, issues arose regarding the president who would be re-elected, causing controversy on various social media, such as Twitter, Instagram, and so on. Of the various public opinions that emerged, many were sentimental, so it is interesting to carry out further research with an emphasis on public opinion on Twitter. The research team began collecting data and a series of text preprocessing processes on the data that had been collected. Furthermore, the results of the text preprocessing are carried out to determine the polarity of tweet sentiment using the Indonesian sentiment dictionary, then sentiment analysis is carried out with data.

Similar research has been conducted, including [2] researching movie reviews which consist of 25,000 documents that have an average length per review of 233 words, using the *Continuous Bag of Words Model* (CBOW) and Skip-Gram methods on word2vec to form a vector representation of each word (*word vector*) in the corpus. The result is the best accuracy of the 100-word *vector dimension* of 88.17% and the lowest accuracy of 85.86% for the 500-word *vector dimension*[2] . Another study was conducted by [3] on air quality and temperature data in the city of Bandung using LSTM modeling with 4 hidden layers, batch size 32, optimizer adam and epoch 1000 by showing the predicted RMSE value is smaller than the standard deviation test value dataset[3].

In addition, there is also research conducted by [4] on reviews on *TripAdvisor* about the influence of COVID-19 on tourist attractions in Bali from *TripAdvisor* using the *Long Short-Term Memory (LSTM) method*. Before processing with LSTM, every text in the review will be vectorized with *word2vec*. The test results on the built model obtained an accuracy value of 71.67% [4]. Another research conducted was by [5] which produced accuracy = 96.68%, *recall* = 94.04%, *precision* = 95.82%, and AUC = 0.979 on *Twitter* data discourse on relocating the Indonesian capital using SVM [5]. Unlike [6], who also researched *Twitter data* regarding PPKM policies, Aldiansyah stated that his research results only reached 64% with the *Linear kernel SVM* [6]. It's

*Corresponding Author: Yohana Karuniawati
MODES Department Atilim University Turkey

different from [7] who got 71.88% accuracy with 51% neutral polarity data, 39% positive polarity, and 10% negative polarity by researching *Twitter data* regarding Xiaomi Indonesia using the *Naive Bayes Classifier* [7].

II. METHOD

The research method used includes literature review, data collection, text preprocessing, dataset distribution, implementation and testing, discussion of results, and finally reporting. Visually the research method is shown in Figure 1.

2.1. Datasets

From Twitter's data of 797 data, it is certain that the data does not display retweet results. This data was obtained in stages from Twitter crawling from April 15, 2022 to April 20, 2022. The results of Twitter crawling data are stored in a .csv file so that it is easy to process in the text preprocessing process using python. Then the data is divided into 80% training data and 20% testing data.

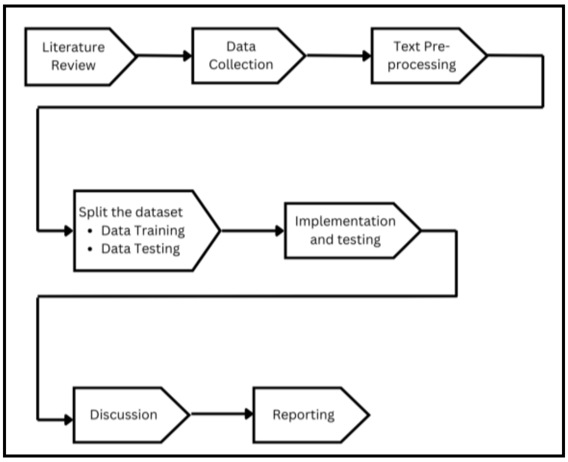

Figure 1: Research Method

2.2. Stages of Text Preprocessing

According to Oueslati et al [8], text preprocessing is a very important step before starting a research. Because a research is said to be successful and smooth if in the text preprocessing stage there are very few errors. Text preprocessing is done with the aim that the initial data is processed by going through several stages until the data is completely ready for use. The stages of Text Preprocessing are shown in Figure 2.

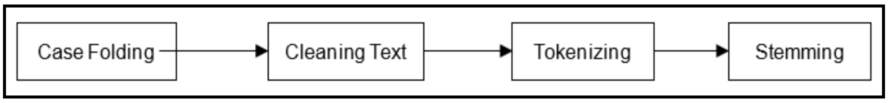

Figure 2: Stages of Text Preprocessing

2.2.1.Case Folding

Case folding is a step in the process of changing words into the same form using the Python string lower method [9]. The purpose of case folding is to return all words to all lowercase so that the text data processed is all in the same condition.

2.2.2. Cleaning Text

Cleaning text is the process of cleaning the review of words that are not needed to reduce noise in the classification process. The omitted words are characters [10].

2.2.3. Tokenizing

Tokenization broadly breaks down a set of characters in a text into units of words, how to distinguish certain characters that can be treated as word separators or not. For example whitespace characters, such as enter, tabulation, space are considered as word separators. But for single quote characters ('), periods (.), semicolon (;), colons (:) or others, they can have quite a lot of roles as word separators. In treating the characters in the text is very dependent on the context of the application being developed. This tokenization work will be even more difficult if you also have to pay attention to the structure of the language (grammatical) [11].

2.2.4. Stemming

Stemming is the process of mapping and decomposing various forms of words into their basic forms. The process of mapping and parsing is used to find the root word of a word that has affixes by removing or deleting these affixes [12].

2.3. Long Short-Term Memory (*LSTM)*

The LSTM *deep learning* method was chosen because it has good accuracy in text data and also advantages in processing relatively long data (*long-term dependencies*) [2]. LSTM as shown in Figure 4 consists of *an input gate, output gate, cell state*, and *forget gate* which will calculate the output value as a *hidden layer* for the next network. LSTM *cell* takes input and stores it for some time. Intuitively, the *input gate* controls how far the new value goes into the *cell*, the *forget gate* controls how far the value stays in the *cell*, and the *output gate* controls how far the value in the *cell* is used to calculate the activation output of the LSTM unit [3]. To determine the *forget gate*, that is by using the formula (1). Formulas (2) and (3) are used to calculate the gate input. Meanwhile, formula (4) is used to calculate the *gate output*. The LSTM architecture is shown in Figure 3.

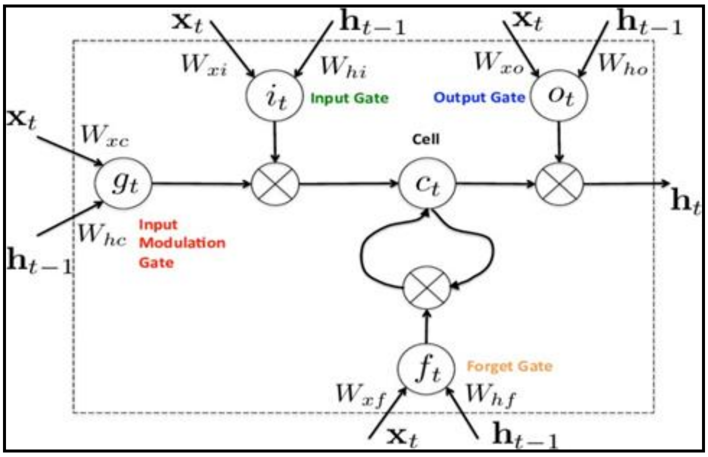

Figure 3: LSTM architecture

The forget gate formula

$ft = \sigma(WfxXt+Wfhht\text{-}1+WfcCt\text{-}1+bf)$ → (1)

Gate inputs

$it = \sigma\ (WixXt+Wih\ h\ t - 1+WicCt - 1+bi) * tanh\ \ (WcxXt+Wch\ h\ t - 1+bi)$ → (2)

$Ct = ft*Ct{-}1+it$ → (3)

Output gates

$Ot = \sigma(WoxXt + Wohht - 1 + WocCt - 1 + bo)Ot = \sigma(WoxXt + Wohht - 1 + WocCt - 1 + bo)$ → (4)

Information

it = Input Gates
ft = Forget Gate
Ot = Output Gates
Ct = Update Cell State
Xt = Input Value
ht = Output Value
ht-1 = Previous Cell Output Value
$Ct{-}1+it$ = Previous Cell State
bb = Biased
W = Weight
$\sigma\sigma$ = Sigmoid
tanh = Tanh

2.4. Long Short-Term Memory (*LSTM)*

Data that is ready to be analyzed is data that has gone through the stages of *text preprocessing*. The initial data is 787 data after being filtered through the *text processing process* so that the data becomes 638 data. Then the data is divided into 20% testing data and 80% training data. Then do the *tweet data labeling* into positive, negative, and neutral. At the classification stage, the researcher used an Indonesian lexicon dictionary. This stage is very vital because the data to be classified is in the form of words in Indonesian (*word level*). After the classification is complete, data visualization can be performed in the form of a graph, or matrix, or *word cloud*.

III. RESULTS AND DISCUSSION

3.1. Case Folding and Cleaning Text

The raw data that has been obtained from the results of direct crawling on the Twitter web is then processed with *python* to make all text lowercase (*case folding*) and also removes all symbols, URLs, and special characters (*cleaning*). Some data that has gone through the case folding and cleaning stages are shown in Table 1.

Table 1 Example of Case Folding and Cleaning Process Results

No.	Tweet	Cleaning
1.	b'Memperpanjang masa jabatan presiden menjadi 3 periode, atau lebih, sama dengan pengkhianatan terhadap reformasi, oleh karena itu harus ditolak keras....!!!'	memperpanjang masa jabatan presiden menjadi periode atau lebih sama dengan pengkhianatan terhadap reformasi oleh karena itu harus ditolak keras
2.	b'Tunduk terhadap konstitusi, Presiden @jokowi tidak ada NIAT untuk lanjutkan 3 Periode.. https://t.co/OTktUswdnP'	tunduk terhadap konstitusi presiden tidak ada niat untuk lanjutkan periode
3.	b'@PartaiSocmed @jejaflo Dengan bahasa ambigu, Presiden Jokowi tidak pernah secara tegas menolak 3 periode, tapi hanya akan patuh dan taat pada konstitusi.\nKlo konstitusi yang diamandemen membolehkan 3 periode berarti......'	dengan bahasa ambigu presiden jokowi tidak pernah secara tegas menolak periode tapi hanya akan patuh dan taat pada konstitusinklo konstitusi yang diamandemen membolehkan periode berarti
4.	b'@PartaiSocmed Apa kira2 yg akan terjadi kalo presiden Jokowi bisa 3 periode?'	apa kira yg akan terjadi kalo presiden jokowi bisa periode
5.	b"Amien Rais 'Wanti-wanti' Rencana Presiden 3 Periode Masih Bisa Terealisasi\nhttps://t.co/2A646kHrFd"	amien rais wantiwanti rencana presiden periode masih bisa terealisasi

Table Source: Author Research Results

3.2. Tokenizing and Stemming

After cleaning unnecessary characters (*cleaning text*), the data is then divided into syllables (*tokenizing*). After that, words containing Indonesian *stopwords* are deleted, and the final text *pre-processing step* is removing affixes and processing them into basic words (*stemming*). In this process using the Literature library where the *stemming algorithm* used is Nazief-Adriani with a little customization to the *python script* to optimize the *stemming results* so that it is more optimal. Figure 4 is the result of this series of processes.

	tweet	tokenizing	stemming
0	b'Memperpanjang masa jabatan presiden menjadi ...	[memperpanjang, masa, jabatan, presiden, menja...	[panjang, jabat, presiden, periode, khianat, r...
1	b'@Sutikno083 @Rahmatmikhaela @DokterTifa Kok ...	[kok, lari, nya, ke, periodeanggaplah, anis, j...	[lari, periodeanggaplah, anis, presidenemang, ...
2	b'@lyan_concept @SBYudhoyono @jokowi @PartaiSo...	[kemungkinan, besar, bisa, berdarah, masmengin...	[darah, masmengingat, khianat, amanah, revorma...
3	b'@Mythicalforest @budaksenar Presiden Uganda ...	[presiden, uganda, yang, minta, periode, beser...	[presiden, uganda, periode, serta, menteri, ga...
4	b'Tunduk terhadap konstitusi, Presiden @jokowi...	[tunduk, terhadap, konstitusi, presiden, tidak...	[tunduk, konstitusi, presiden, niat, lanjut, p...
...	...	...	...
792	b'@KangUtang04 @MatahariTimur17 @l205mdr____ @...	[njokowi, luhut, tito, teroris, konstitusinnpa...	[njokowi, luhut, tito, teroris, konstitusinnpa...
793	b'@MustofaNahra_ID \xc2\xb0\xc2\xb0\nDi Sri La...	[xcxbxcxbndi, sri, langka, setelah, presiden, ...	[sri, langka, presiden, tumpuk, utang, kepincu...
794	b'@EdiMahaMG \xc2\xb0\xc2\xb0\nDi Sri Langka s...	[xcxbxcxbndi, sri, langka, setelah, presiden, ...	[sri, langka, presiden, tumpuk, utang, kepincu...
795	b'@KontraS Dia yg bikin isu penundaan & 3 ...	[dia, yg, bikin, isu, penundaan, amp, periode,...	[bikin, isu, tunda, periode, presiden, demo, n...
796	b'Sekali lagi Presiden @jokowi TEGASKAN, tidak...	[sekali, lagi, presiden, tegaskan, tidak, niat...	[presiden, tegas, niat, perintah, tunda, pemil...

797 rows × 3 columns

Figure 4: Example of Tokenizing and Stemming Results

3.3. Tweet Data Labelling

The labeling used in this study uses the Indonesian language lexicon database. There are 638 *clean data* ready for the sentiment labeling stage using the Lexicon dictionary so that the *polarity score can be calculated*. From the results of the *polarity score*, it can be determined which tweets belong to the *positive, negative,* and *neutral classes.* An example of tweet data *labeling results with polarity positive* is shown in Table 2, where the *polarity score* shows a positive number above the value 0.

Table 2 Example of Positive Tweet Labeling

No.	Text	Polarity Score	Polarity
1.	Dude at the moment what is being done is not a figure with a strong character but a great portfolio leading with integrity understanding ethics and morals but those who give money to the people are the ones who are chosen from a provision perspective, Mr. Tbh is very qualified once in the palace of the presidential adjutant of RI's period	17	Positive
2.	Do you think that only the president's proposal will be approved by the president anyway, right?	12	Positive
3.	all the people love Mr. President Jokowi, I hope Mr. President Jokowi can continue the period, aamiin	11	Positive
4.	As a Muslim, how come you are so stupid that the president's sister can marry the head of the MK?	10	Positive
5.	he's a minister, he doesn't have a vision and mission of a minister who has a vision and mission of the president, meaning if there's a minister talking about a period or postponing an election, that means it's from the president, right or not?	10	Positive

Table Source: Author Research Results

An example of the results of *labeling* tweet data with *negative polarity is* shown in Table 3, where you can see the negative *polarity score* or minus number. Meanwhile, an example of neutral labeling is shown in Figure 5.

Table 3 Example of Negative Tweet Labeling

No.	Text	Polarity Score	Polarity
1.	those radicals who want to extend the term of office of the president, those who want to amend the constitution so that the president can serve a term, those radicals who say they have big data but don't want to open it to the public have deceived the public by spreading false news that is subject to articles	-29	Negative
2.	because of the slander about the president during the Felis period & Silalahi was paid for by the Suharto clan, it was compromised & the greedy bastards masyutk a constitutional coup even though it was a slander that was repeated by Suharto during the mingkem period all asxexxbcxefxbxf	-25	Negative
3.	it turned out that the demonstration still had an effect, there was also the suspect of the cpo mafia, although only his cronies appeared before the presidential demonstration said it would stop.	-22	Negative
4.	so it's useless to study at school study ppkn about the election of the president and vice president and his term of office is only the year of the period in ppkn there is no study of the president's period considerations and full of consideration regrets bjr uud there is also no vote from fb	-21	Negative
5.	the threshold is maintained to smooth Jokowi periodnnpt is the politics of the pigsty to continue gripping Indonesiann these thieving pigs who robbed Indonesia must be expelled from Indonesia	-19	Negative

Source Table of Author Research Results

index	text_clean	polarity_score	polarity
1	dengan bahasa ambigu presiden jokowi tidak pernah secara tegas menolak periode tapi hanya akan patuh dan taat pada konstitusinklo konstitusi yang diamandemen membolehkan periode berarti	0	neutral
2	ini biangnya tiga periode nn	0	neutral
3	akhiri polemik wacana penundaan pemilu dan masa jabatan presiden menjadi perioden	0	neutral
4	warawiri tunda sambil putusan periode kan bisa harapan cak imin apabila nantinya menang jadi presiden xfxfxx ngebet jadi presidenxfxfxxd	0	neutral
5	presiden jg kagak minta periode	0	neutral

Figure 5: Labeling Neutral Tweets

3.4. Parameter tuning of the LSTM Model

A comparison of the results of the tests carried out is shown in Table 4. For the number of LSTM 1 units, the epoch used was 10, with a dropout rate of 0.2, an embed dim of 32, resulting in a learning rate of 0.01. For the number of LSTM 215 units, the epochs were carried out 5 times, with the same *dropout rate* of 0.2 and embed dim 64, with a smaller learning rate of 0.0001, as well as other test data. In this study the researchers used several parameters which are shown in more detail in Figure 6.

Table 4 Example of LSTM Model Parameter *Tuning*

LSTM Units	Epoch	Dropout Rate	Embed Dim	Learning Rate
1	10	0.2	32	0.01
22	5	0.2	64	0.01
215	5	0.2	64	0.0001

Best: 0.694118 using {'batch_size': 128, 'dropout_rate': 0.2, 'embed_dim': 32, 'epochs': 10, 'hidden_unit': 16, 'learning_rate': 0.01, 'optimizers': <class 'keras.optim

1 to 25 of 21

index	means	stds	params
0	0.6941176454226176	0.0220097374020754	{'batch_size': 128, 'dropout_rate': 0.2, 'embed_dim': 32, 'epochs': 10, 'hidden_unit': 16, 'learning_rate': 0.01, 'optimizers': <class 'keras.optimizer_v2.adam.Adam'>}
1	0.6901960770289103	0.012087098452086373	{'batch_size': 128, 'dropout_rate': 0.2, 'embed_dim': 64, 'epochs': 10, 'hidden_unit': 32, 'learning_rate': 0.01, 'optimizers': <class 'keras.optimizer_v2.adam.Adam'>}
3	0.6843137343724569	0.01941077122155295	{'batch_size': 128, 'dropout_rate': 0.2, 'embed_dim': 32, 'epochs': 5, 'hidden_unit': 32, 'learning_rate': 0.01, 'optimizers': <class 'keras.optimizer_v2.rmsprop.RMSprop'>}
2	0.6843137343724569	0.012087072667685888	{'batch_size': 128, 'dropout_rate': 0.2, 'embed_dim': 32, 'epochs': 5, 'hidden_unit': 16, 'learning_rate': 0.01, 'optimizers': <class 'keras.optimizer_v2.rmsprop.RMSprop'>}
4	0.6803921461105347	0.005545935208154502	{'batch_size': 128, 'dropout_rate': 0.2, 'embed_dim': 32, 'epochs': 20, 'hidden_unit': 32, 'learning_rate': 0.01, 'optimizers': <class 'keras.optimizer_v2.adam.Adam'>}
6	0.6745098233222961	0.01999615376340211	{'batch_size': 256, 'dropout_rate': 0.2, 'embed_dim': 32, 'epochs': 10, 'hidden_unit': 32, 'learning_rate': 0.01, 'optimizers': <class 'keras.optimizer_v2.rmsprop.RMSprop'>}
5	0.6745098233222961	0.014673165348054048	{'batch_size': 256, 'dropout_rate': 0.2, 'embed_dim': 64, 'epochs': 10, 'hidden_unit': 64, 'learning_rate': 0.01, 'optimizers': <class 'keras.optimizer_v2.adam.Adam'>}
7	0.6745098034540812	0.021657556739559273	{'batch_size': 128, 'dropout_rate': 0.2, 'embed_dim': 64, 'epochs': 10, 'hidden_unit': 64, 'learning_rate': 0.01, 'optimizers': <class 'keras.optimizer_v2.rmsprop.RMSprop'>}
8	0.6725490291913351	0.022697731795165266	{'batch_size': 128, 'dropout_rate': 0.2, 'embed_dim': 32, 'epochs': 10, 'hidden_unit': 32, 'learning_rate': 0.01, 'optimizers': <class 'keras.optimizer_v2.rmsprop.RMSprop'>}
9	0.6686274607976278	0.026452413236045656	{'batch_size': 128, 'dropout_rate': 0.2, 'embed_dim': 64, 'epochs': 10, 'hidden_unit': 16, 'learning_rate': 0.01, 'optimizers': <class 'keras.optimizer_v2.rmsprop.RMSprop'>}
10	0.6686274607976278	0.012087098452086373	{'batch_size': 128, 'dropout_rate': 0.2, 'embed_dim': 32, 'epochs': 10, 'hidden_unit': 64, 'learning_rate': 0.01, 'optimizers': <class 'keras.optimizer_v2.rmsprop.RMSprop'>}
11	0.6686274607976278	0.007336572054025197	{'batch_size': 256, 'dropout_rate': 0.2, 'embed_dim': 64, 'epochs': 20, 'hidden_unit': 16, 'learning_rate': 0.01, 'optimizers': <class 'keras.optimizer_v2.adam.Adam'>}
12	0.6686274409294128	0.014673165348054048	{'batch_size': 128, 'dropout_rate': 0.2, 'embed_dim': 32, 'epochs': 10, 'hidden_unit': 64, 'learning_rate': 0.01, 'optimizers': <class 'keras.optimizer_v2.adam.Adam'>}
13	0.6686274409294128	0.011091870416309005	{'batch_size': 256, 'dropout_rate': 0.2, 'embed_dim': 32, 'epochs': 20, 'hidden_unit': 16, 'learning_rate': 0.01, 'optimizers': <class 'keras.optimizer_v2.adam.Adam'>}
14	0.6666666666666666	0.012087098452086373	{'batch_size': 128, 'dropout_rate': 0.2, 'embed_dim': 32, 'epochs': 20, 'hidden_unit': 64, 'learning_rate': 0.01, 'optimizers': <class 'keras.optimizer_v2.adam.Adam'>}
15	0.6666666666666666	0.007336572054025198	{'batch_size': 128, 'dropout_rate': 0.2, 'embed_dim': 32, 'epochs': 10, 'hidden_unit': 32, 'learning_rate': 0.01, 'optimizers': <class 'keras.optimizer_v2.adam.Adam'>}
16	0.6647058924039205	0.00831888876328225	{'batch_size': 256, 'dropout_rate': 0.2, 'embed_dim': 32, 'epochs': 10, 'hidden_unit': 16, 'learning_rate': 0.01, 'optimizers': <class 'keras.optimizer_v2.rmsprop.RMSprop'>}
18	0.6647058725357056	0.033275611248927014	{'batch_size': 128, 'dropout_rate': 0.2, 'embed_dim': 64, 'epochs': 20, 'hidden_unit': 64, 'learning_rate': 0.001, 'optimizers': <class 'keras.optimizer_v2.rmsprop.RMSprop'>
17	0.6647058725357056	0.016637805624463507	{'batch_size': 256, 'dropout_rate': 0.2, 'embed_dim': 64, 'epochs': 10, 'hidden_unit': 16, 'learning_rate': 0.01, 'optimizers': <class 'keras.optimizer_v2.rmsprop.RMSprop'>}
20	0.6647058725357056	0.009605841556008675	{'batch_size': 128, 'dropout_rate': 0.2, 'embed_dim': 64, 'epochs': 20, 'hidden_unit': 32, 'learning_rate': 0.01, 'optimizers': <class 'keras.optimizer_v2.adam.Adam'>}
19	0.6647058725357056	0.009605841556008675	{'batch_size': 128, 'dropout_rate': 0.2, 'embed_dim': 64, 'epochs': 10, 'hidden_unit': 16, 'learning_rate': 0.01, 'optimizers': <class 'keras.optimizer_v2.adam.Adam'>}
21	0.6627451181411743	0.02772967604077251	{'batch_size': 256, 'dropout_rate': 0.2, 'embed_dim': 64, 'epochs': 10, 'hidden_unit': 32, 'learning_rate': 0.01, 'optimizers': <class 'keras.optimizer_v2.adam.Adam'>}
23	0.6627450982729594	0.04196262449343916	{'batch_size': 256, 'dropout_rate': 0.2, 'embed_dim': 32, 'epochs': 5, 'hidden_unit': 16, 'learning_rate': 0.01, 'optimizers': <class 'keras.optimizer_v2.rmsprop.RMSprop'>}
22	0.6627450982729594	0.009998072985232642	{'batch_size': 128, 'dropout_rate': 0.2, 'embed_dim': 64, 'epochs': 5, 'hidden_unit': 16, 'learning_rate': 0.01, 'optimizers': <class 'keras.optimizer_v2.adam.Adam'>}
27	0.6607843240102133	0.03197943586719807	{'batch_size': 128, 'dropout_rate': 0.2, 'embed_dim': 64, 'epochs': 10, 'hidden_unit': 64, 'learning_rate': 0.01, 'optimizers': <class 'keras.optimizer_v2.adam.Adam'>}

Figure 6: Tuning the LSTM parameters

3.5 Output and Visualization

The results of the classification of tweet sentiments are shown in Figure 7. Out of a total of 638 *clean* tweet data, there were 53.8% tweets with negative labels, 35.0% with positive labels, and the remaining 11.3% were labeled neutral.

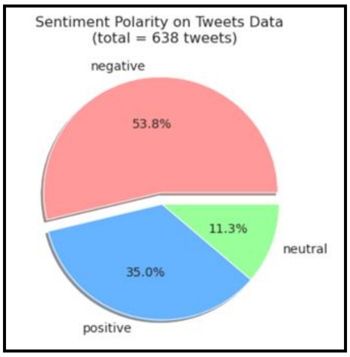

Figure 7: Sentiment Results Percentage Graph

After obtaining tweets with each label grouping, the *word cloud* tweet can be displayed as a whole data as shown in Figure 8. The words that have been labeled positive and negative are also shown in the *word cloud visualization* in Figure 9 and for the positive class and in Figure 10 for the negative class.

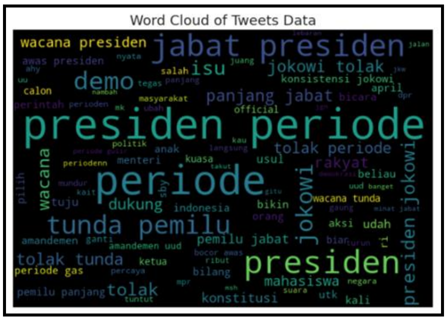

Figure 8: Word cloud Tweet Data

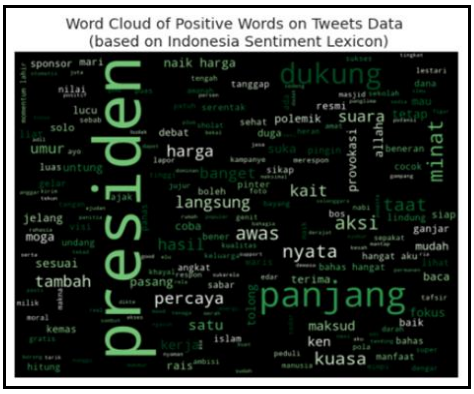

Figure 9: Positive Class word cloud

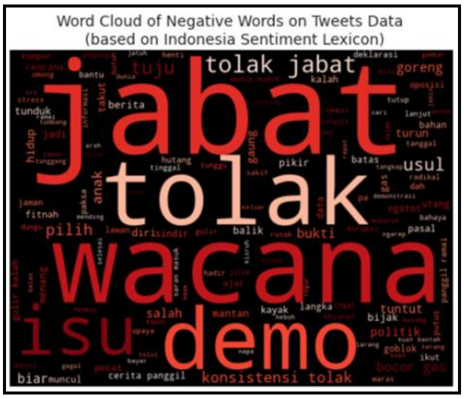

Figure 10: Negative Class word cloud

3.5 Testing and Evaluation

The last stage in this research is to carry out the testing and evaluation process by measuring the accuracy of the model that has been produced in the previous stage. Figure 11 shows the results of model accuracy on the train data and test data. Figure 12 shows *the confusion matrix* for testing the Twitter sentiment analysis data on the topic of Presidents for 3 periods. The test results show that the accuracy obtained is 75%. This result is higher than the results of studies [13], [14], [15], and [16].

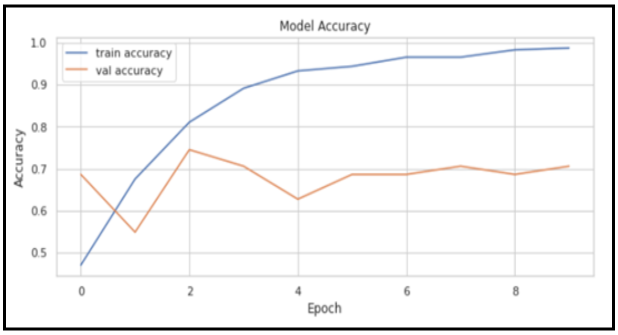

Figure 11: Model Accuracy

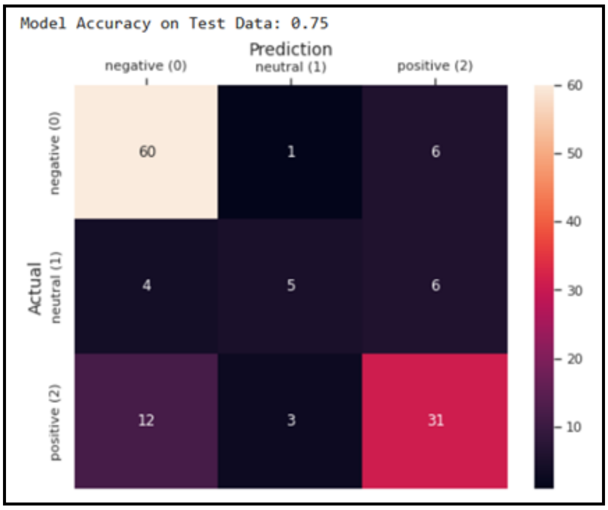

Figure 12: Confusion Matrix

IV. CONCLUSION

Based on the results of testing that has been carried out on *tweets* on the topic sentiment of the president for 3 periods from social media Twitter, there are 638 tweets (343 positive, 72 neutral, and 223 negative) using the *Long Short Term Memory* (LSTM) algorithm obtained quite good results with an accuracy of = 75%. For future research, the results of this study have the opportunity to be developed further by increasing the amount *of train data* and *test data* from the dataset taken to improve accuracy. Other research can also make improvements to the dataset by changing search keywords on Twitter, by utilizing specific account filters according to the topic under study in order to get maximum results. Other studies may also use different methods for this topic, for example *Convolutional Neural Network* (CNN), *Support Vector Machine* (SVM), or by other methods.

REFERENCES

[1] Putranto, SA (2022, 04 01). Jokowi's Discourse on 3 Periods, Claims for Democracy and Symptoms of Authoritarianism. Retrieved fromnasional.kompas.com/: https://nasional.kompas.com/read/2022/04/01/09010031/wacana-jokowi-3-period-klaim-demokras-dan-symptoms-otoritarianism/

[2] Widayat, W. (2021). Movie Review Sentiment Analysis using Word2Vec and the LSTM Deep Learning method. Journal of Media Informatics Budidarma , 5 (3), 1018. https://doi.org/10.30865/mib.v5i3.3111

[3] Khumaidi, A., Raafi'udin, R., & Solihin, IP (2020). Testing the Long Short Term Memory Algorithm for Predicting Air Quality and Temperature in Bandung City. Journal of Telematics, 15 (1), 13–18. https://journal.ithb.ac.id/telematika/article/view/340

[4] Nurvania, J., & Lhaksamana, KM (2021). Sentiment Analysis on TripAdvisor Reviews Using the Long Short-Term Memory (LSTM) Method. E-Proceedings of Engineering , 8 (4), 4124–4135.

[5] Putra, A., Haeirudin, D., Khairunnisa, H., & Latifah, R. (2021). Community Sentiment Analysis of PPKM Policy on Twitter Social Media Using the Svm Algorithm. National Seminar on Science and Technology 2021 , November , 1–6.

[6] Arsi, P., & Waluyo, R. (2021). Sentiment Analysis on Moving Indonesia's Capital City Discourse Using the Support Vector Machine (SVM) Algorithm. Journal of Information Technology and Computer Science , 8 (1), 147. https://doi.org/10.25126/jtiik.0813944

[7] Utami, RW, Jazuli, A., & Khotimah, T. (2021). Sentiment Analysis of Xiaomi Indonesia Using the Naïve Bayes Method. Indonesian Journal of Technology, Informatics and Science (IJTIS) , 3 (1), 21–30. https://doi.org/10.24176/ijtis.v3i1.7514

[8] Oumaima Oueslati, et al. (2020). A review of sentiment analysis research in Arabic language, Future Generat. Comput. Syst.

[9] Ronal Watrianthos, Samsir, Basyarul Ulya, Junaidi Mustapa Harahap, Deci Irmayani, Firman Edi, Jupriaman, Rizki Kurniawan Rangkuti (2021), Naives Bayes Algorithm for Twitter Sentiment Analysis, https://iopscience.iop.org/article/10.1088/1742-6596/1933/1/012019/pdf

[10] Widodo, Fani. (2021). NLP Sederhana Dengan Python. <https://sites.unpad.ac.id/widodo/2021/03/09/nlp-dengan-python/> (Accessed on 11 Mei 2022)

[11] Hendrawan, Aria. (2020). PENGEMBANGAN STEMMING UNTUK ARTIKEL BERBAHASA INDONESIA. https://repository.usm.ac.id/files/research/G071/20200410024415-PENGEMBANGAN-STEMMING-UNTUK-ARTIKEL-BERBAHASA-INDONESIA.pdf

[12] Hapsari, R. K., & Santoso, Y. J. (2015, January). Stemming Artikel Berbahasa Indonesia Dengan Pendekatan Confix-Stripping. In Prosiding Seminar Nasional Manajemen Teknologi XXII (pp. 1-8). https://www.academia.edu/download/54169782/Stemming_Artikel_Berbahasa_Indonesia_Dengan_Pendekatan_Confix_stripping.pdf

[13] Mollah, Md. (2022). An LSTM model for Twitter Sentiment Analysis. 10.48550/arXiv.2212.01791.

[14] Ihsan, Miftahul & Negara, Benny & Agustian, Surya. (2022). LSTM (Long Short Term Memory) for Sentiment COVID-19 Vaccine Classification on Twitter. Digital Zone: Jurnal Teknologi Informasi dan Komunikasi. 13. 79-89. 10.31849/digitalzone.v13i1.9950.

[15] Y. Astari and S. W. Rozaqi, "Analisis Sentimen Multi-Class pada Sosial Mediamenggunakan metode Long Short-Term Memory (LSTM)," vol. 4, no. 1, pp. 8–12,2021

[16] P. F. Muhammad, R. Kusumaningrum, and A. Wibowo, "Sentiment Analysis UsingWord2vec and Long Short-Term Memory (LSTM) for Indonesian Hotel Reviews,"Procedia Comput. Sci., vol. 179, pp. 728–735, 2021