

TIC-322144 - Sentiment analysis on Twitter reviews data using ...

Sources Overview

26%

OVERALL SIMILARITY

1	journal.ugm.ac.id	16%
	INTERNET	
2	D F Anas, I Jaya, Nurjanah. "Design and implementation of fish freshness detection algorithm using deep learning", IOP Conference Se...	<1%
	CROSSREF	
3	Yunian Ru, Bo Li, Jianping Chai, Jianbo Liu. "A Daily Box Office Prediction Model with LSTM", 2019 International Conference on Artifici...	<1%
	CROSSREF	
4	www.hindawi.com	<1%
	INTERNET	
5	University of Oxford on 2016-08-30	<1%
	SUBMITTED WORKS	
6	adoc.pub	<1%
	INTERNET	
7	eprints.uad.ac.id	<1%
	INTERNET	
8	Tong Deng, Feng-hai Liu, Guo-zhu Jia. "Prediction carbon dioxide solubility in ionic liquids based on deep learning", Molecular Physics,...	<1%
	CROSSREF	
9	Universitas Atma Jaya Yogyakarta on 2019-02-23	<1%
	SUBMITTED WORKS	
10	garuda.ristekbrin.go.id	<1%
	INTERNET	
11	CSU, San Jose State University on 2018-11-20	<1%
	SUBMITTED WORKS	
12	University College London on 2018-09-07	<1%
	SUBMITTED WORKS	
13	Xinfang Wu, Yong Xiang, Gang Mao, Mingqian Du, Xiuqing Yang, Xinzhi Zhou. "Forecasting air passenger traffic flow based on the two-...	<1%
	CROSSREF	
14	Zhong-shi Gao, Yang Su, Yi Ding, Yu-dong Liu, Xu-an Wang, Jun-wei Shen. "Chapter 24 Key Technologies of Anomaly Detection Using P...	<1%
	CROSSREF	
15	National College of Ireland on 2014-05-28	<1%
	SUBMITTED WORKS	
16	Tarumanagara University on 2021-01-27	<1%
	SUBMITTED WORKS	

17

Universitas Mercu Buana on 2020-05-08

SUBMITTED WORKS

<1%

18

jurnal.ugm.ac.id

INTERNET

<1%

19

Ramadhan Renaldy, Muljono, Guruh Fajar Shidik, Catur Supriyanto, M.A. Soeleman, Ahmad Zainul Fanani, Aditya Dwi Pratama. "Hybrid ...

CROSSREF

<1%

20

www.edureka.co

INTERNET

<1%

Excluded search repositories:

None

Excluded from document:

Bibliography

Quotes

Excluded sources:

None

Sentiment analysis on Twitter reviews data using the bidirectional long short-term memory (BiLSTM)

Riszki Wijayatun Pratiwi*, Sharfina Febbi Handayani, Dairoh, Dwi Intan Afidah

Department of Informatics Engineering, Politeknik Harapan Bersama, Indonesia

*Corresponding author: riszkipratiwi@poltektegal.ac.id

Abstract. Twitter is one of the social media that can see written expressions or tweets from users, both praise tweets or tweets that contain hate speech. From the tweet review, the dataset can be used as research, namely sentiment analysis. In this study, the dataset used was a dataset from previous studies. Currently, the development of Deep Learning is very rapid. One method of Deep Learning is Bidirectional Long Short Term Memory (BiLSTM). BiLSTM is a development method from Long Short Term Memory, called “Bidirectional” because the structure processes data sequences back and forth. This bidirectional is in the RNN and LSTM methods. This study aims to carry out a classification process on Twitter review data whose sentiment classes consist of positive, neutral, and negative using the BiLSTM method. After doing the research, the BiLSTM method obtained an accuracy of 58.56%. The accuracy is obtained from the hyperparameter tuning value. Several parameters produce optimal accuracy, namely word2vec CBOW, number of neurons 200, learning rate 0.00001, number of epochs 25, and softmax activation function. Then, with this parameter value, the accuracy is 58.56%.

1. Introduction

Social media users are currently increasing because through social media people can interact without limits. Twitter is a social media that can see written expressions or tweets from users, both praising tweets or tweets containing hate speech [1]. However, if reading all these tweets can take a long time, and if few tweets are read, the evaluation will be biased [2]. To understand and handle the opinions of these tweets, algorithms and programs are needed so that they can process information and opinion data, and can analyze the opinions of social media users, which is called sentiment analysis [3].

Sentiment analysis is a field of study that analyzes people's opinions, sentiments, evaluations, attitudes, and emotions from written language[4]. Sentiment analysis is also used as a way to evaluate public opinion through writing or from other related subjects on several topics [5].

Several studies apply Twitter reviews to sentiment classification. The methods used are varied. One of them uses machine learning, namely the naive Bayes method and support vector machine (SVM) which is used to find models or features that match the target. However, there are still shortcomings in machine learning in the complex feature extraction process and getting better types of features. The feature extraction methods used in machine learning are single words, single-character N-gram, multi-word N-gram, and lexical syntactic [2].

Deep Learning is a branch of Machine Learning that is inspired by the human cortex by applying an artificial neural network that has many hidden layers [6]. There have been many previous studies using deep learning with the Bidirectional Long Short-Term Memory (BiLSTM) method. The BiLSTM

method can be applied in NLP fields such as speech recognition, text translation, text summarization, and sentiment analysis.

Based on the explanation above, this study discusses sentiment analysis on Twitter review data using the deep learning method, namely BiLSTM. BiLSTM is an extension of Long Short Term Memory.

2. Methods

The data used in this study were sourced from [7]. Overall data is 10806 tweet reviews. Table 1 shows the positive, neutral and negative classes. The flow used in this study is presented in Figure 1.

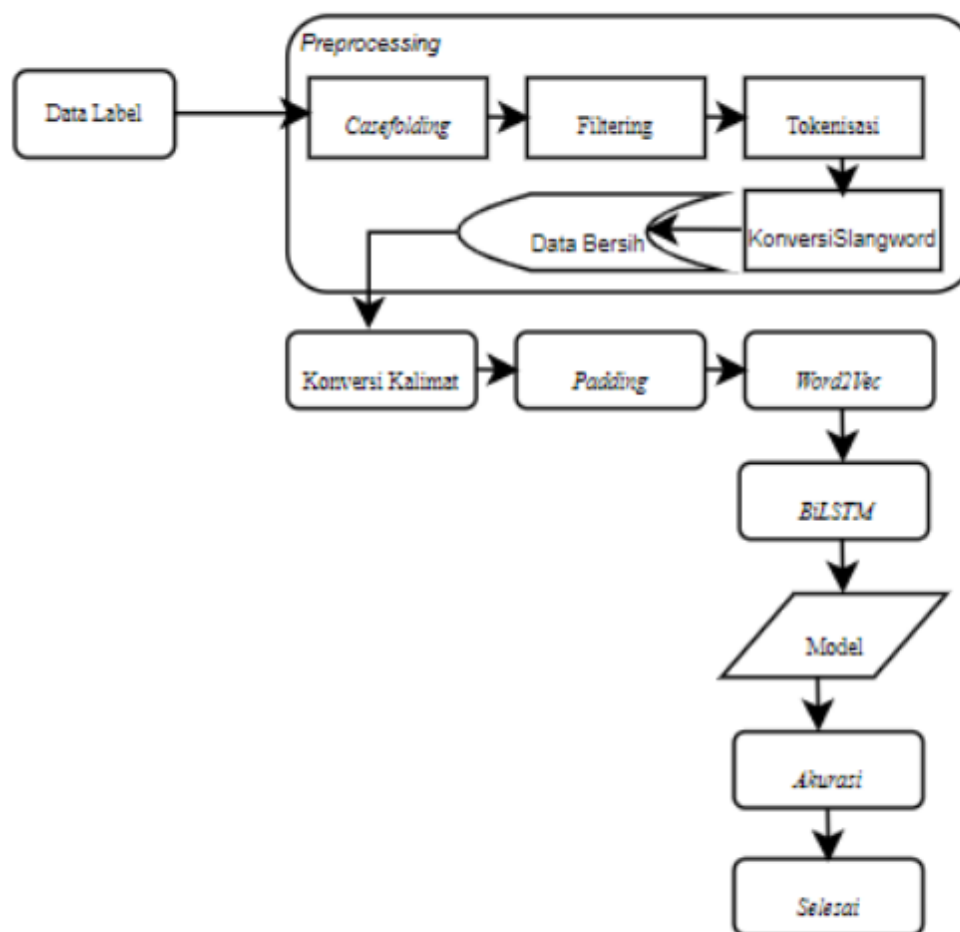


Figure 1. Sentiment Analysis Process using BiLSTM

While the Twitter Review Sentiment Class in Table 1.

Table 1. Twitter Review Sentiment Class

Twitter Review	Class
Whatever you have done, whatever your mistakes, you will always find the word sorry in a mother's heart	positive
Life is not counted by how much you loot indomie but by how much you loot television	Neutral
In the end, we will know failure without preparation	Negative

2.1. Preprocessing

Preprocessing is data cleaning, which is carried out as an optimization of the classification process later. Preprocessing consists of several stages, namely case-folding, filtering, tokenization, and slang words conversion. The Case-folding stage aims to change all the letters in the tweet review to all

lowercase characters. The filtering process removes special characters (\$, %, *, etc.), and removes words that do not match the parsed results, such as usernames starting with the symbol "@", hashtags, "#", Uniform Resource Locator (URL), and emoticons. Symbols, signs, or numbers are omitted because they do not have much influence to determine the label.

The function of tokenization is to divide tweet reviews from sentences into word units. This process is done by looking at the spaces in the review sentence, from these spaces the sentences can be separated into words. The slang word conversion is done to change non-standard words into standard words. At this stage, there is help from a slang word dictionary and standard word equivalents. This stage will check whether the word is present or not in the slang word dictionary, if there is a non-standard word it will be converted into a standard word according to the slang word dictionary.

2.2. Sentence Conversion

There are several stages in sentence conversion, namely making a word dictionary, converting sentences into numbers, and finally padding. Then the results of the sentence conversion will be used for the input process in BiLSTM.

2.3. BiLSTM Architecture

According to [8] the main step in the LSTM is to ensure that the information on the input X_{t-1} and X_t can pass or not from the cell state. This provision is obtained from the sigmoid layer, namely the "forget gate". If the output is 1 then "may go through that path" but if the output is 0 then "forget the information". The forget gate value can be calculated by equation 1.

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

Next, determine new facts, namely in the form of information that will be stored in the cell state. The first sigmoid layer is the input gate which determines which part will be updated later. The tan h layer creates a vector value for the new candidate, \tilde{c}_t can be entered in the cell state. The next step is to determine what new information will be stored in the cell state. The first is a sigmoid layer called the input gate which determines which part to update. Then the tanh layer which creates a new candidate value vector, \tilde{c}_t can be added to the cell state. Then the two are combined to get an update to the state. The calculation of the input gate value is in equation 2, while the odor candidate value is in equation 3.

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

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

After that update the last cell state, C_{t-1} , to the new cell state C_t . The trick is to multiply the old cell state by forget gate f_t then add $i_t * \tilde{c}_t$. For more details can be seen in equation 4.

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

The last process is the output gate. The sigmoid layer that is run will determine what part of the cell will be the output, then put the cell state through the tanh and multiply the output from the sigmoid gate, so that the specified part can be used as output. The calculation of the output gate is in equation 5 and equation 6.

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

$$h_t = O_t * \tanh(C_t) \quad (6)$$

In RNN there is a bidirectional layer extension, this is because the structure processes data sequences back and forth. The existing standard in the RNN can only move forward to the last time step. This is referred to as "Bidirectional". Bidirectional LSTM brings together two hidden layers in the same output layer. Therefore, according to a fully bidirectional network is better than a unidirectional network, especially for the case of classification and speech recognition, and comparisons of LSTM and BiLSTM in Figure 2.

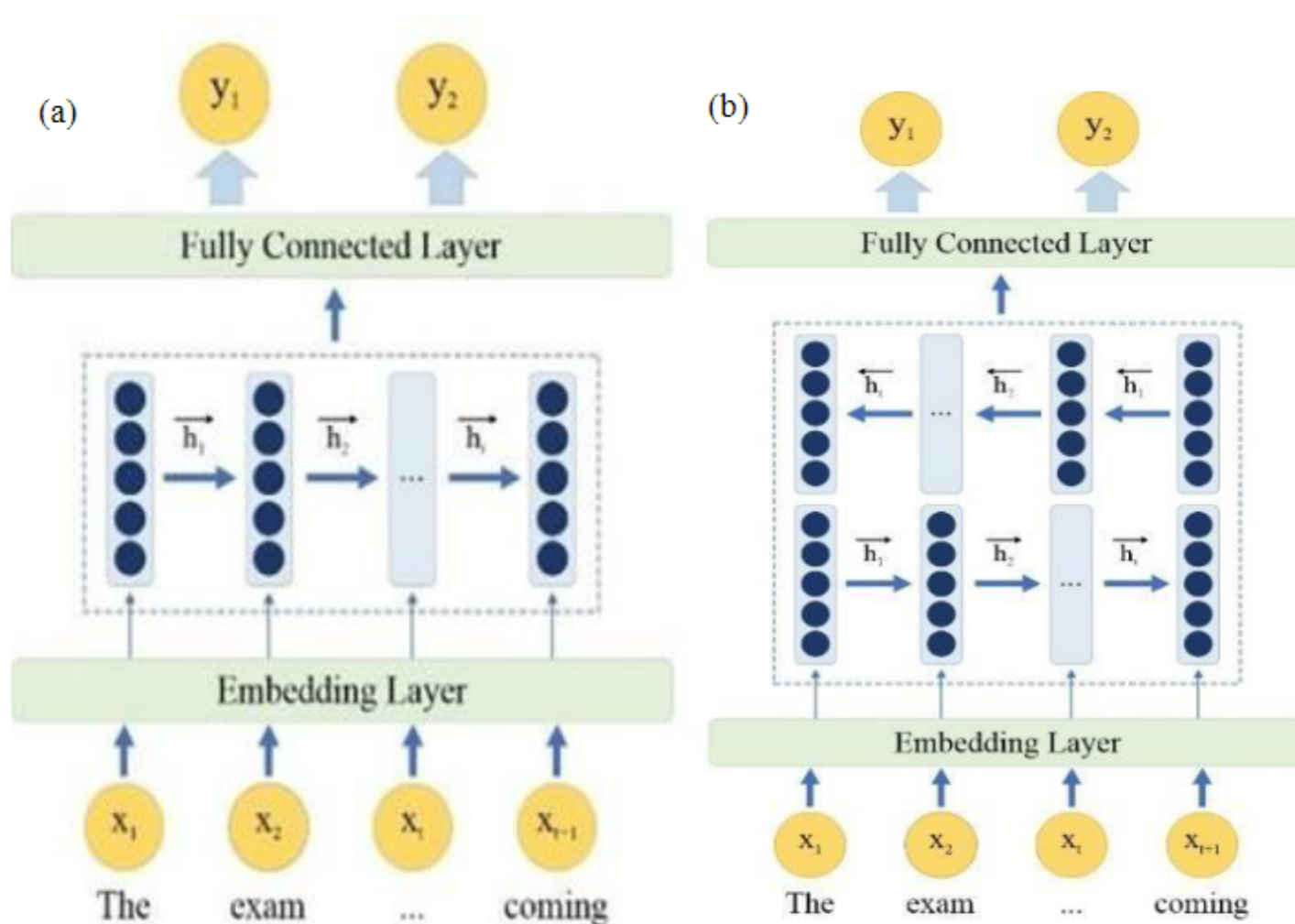


Figure 2. Structure of: (a) LSTM (b) BiLSTM (Winata et al., 2018)

2.4. Result Evaluation

The BiLSTM method will be evaluated with a confusion matrix, to see the performance of the model that has been created. The confusion matrix table is a tool that is used as a performance measurement tool for document classification in one or more classes. In Table 2 is a description of the confusion matrix of the three classes.

		Prediction Class		
		A	B	C
Actual Class	A	TP _A	E _{AB}	E _{AC}
	B	E _{BA}	TP _B	E _{BC}
	C	E _{CA}	E _{CB}	TP _C

Accuracy is a measure of how well a method can classify text correctly. A high accuracy value is obtained when a large amount of data has been classified correctly according to the sentiment class [9]. Calculation of accuracy can be obtained by using the following equation 7.

$$\text{Accuracy} = \frac{\text{Amount of sentiments is correct}}{\text{Amount of the test data}} \times 100\% \quad (7)$$

3. Results and Discussion

At this stage, testing is carried out to determine the performance of the BiLSTM model that has been made. The test used in this study uses accuracy, based on the accuracy value that has been obtained, it can be seen what parameters produce the optimal accuracy value of the BiLSTM model. Due to limited resources, in this study, not all combinations of parameter values were tested. This refers to previous research [4].

When the first parameter has obtained the optimal accuracy value, then that value will be used in the next parameter test. Meanwhile, the parameters tested include word2vec, number of neurons, number

of epochs, learning rate, and activation function. The default parameter values to be tested are presented in Table 3.

Table 3. Parameters and Values tested

Parameter	Value
Word2Vec	CBOW dan Skip-gram
Total Neuron	50, 100, 150, 200, 250
Total Epoch	5,10,20,25
Learning Rate	0.01, 0.001, 0.0001, 0.00001
Activation Function	Sigmoid dan Softmax

3.1. Tuning Word2vec

In this first test, word2vec, there is a CBOW and skip-gram architecture. This stage is carried out so that the CBOW or skip-gram architecture can be seen that has optimal performance. Then the optimal accuracy of the word2vec model will be used in the next parameter tuning. This word2vec test uses random parameter values, considering this is the first test. The test results can be seen in Table 4.

Table 4. Tuning Word2vec on BiLSTM

Word2vec	Time (minute)	accuracy (%)
CBOW	2:55	58.25
Skip-gram	3:21	55.45

According to the results from Table 4, it is found that the results of the CBOW architecture are more optimal than the skip-gram. This happens because the CBOW architecture can obtain a more optimal word embedding and can observe the semantic meaning of each word. And from the test results above, CBOW is also suitable for large amounts of data. Therefore, the accuracy of CBOW is more optimal than skip-gram.

3.2. Tuning Number of Neurons

The next stage is testing the number of neurons. In this study, the number of neurons to be tested was 50, 100, 150, 200, and 250. This test was carried out using the BiLSTM method. The test results can be seen in Table 5.

Table 5. Tuning the Number of Neurons on BiLSTM

Total Neuron	Time (minute)	Accuracy (%)
50	2:11	57.5
100	2:40	57.25
150	3:10	58
200	3:30	58.55
250	3:55	57.6

Based on the test of the neurons above, it can be seen that the optimal result is that the number of neurons is 200. The more the number of neurons tested, the longer it takes to process. A large number of neurons can not guarantee increased accuracy. In this test, there is no special method used to obtain optimal accuracy results. Therefore, to determine the number of neurons with optimal results, it is necessary to experiment or tune the number of neurons

3.3. Tuning Learning Rate

Based on previous tests, optimal results have been obtained from several parameters, namely word2vec and the number of neurons. Then the next step is to test the learning rate using the optimal parameter values from the previous test. The details can be seen in Table 6.

Table 6. Tuning Learning Rate on BiLSTM

Learning Rate	Time (Minute)	Accuracy (%)
0.01	2:32	58.15

0.001	2:45	58.22
0.0001	3:15	58.45
0.00001	3:56	58.6

It can be seen in Table 6 that optimal accuracy is obtained from the learning rate value of 0.00001 with a total of 58.60% so that in this study it can be concluded that the learning rate value is directly proportional to the length of time required.

3.4. Tuning the Number of Epoch

In this study, the number of epochs that will be tested is 5, 10, 20, 25 with the word2vec CBOW architecture, the number of neurons is 200 and the learning rate is 0.00001. More details can be seen in Table 7.

Table 7. Tuning the number of epochs on BiLSTM

Epoch	Time (Minute)	Accuracy (%)
5	2:35	57.2
10	2:50	57.65
20	3:45	58.11
25	4:30	58.45

According to Table 7 that the most optimal number of epoch tests obtained an accuracy value of 58.45%. Based on the results of these tests, it can be concluded that the large number of epochs used can affect the accuracy results, although the accuracy obtained is not too significant for the time required to test the number of epochs. The more epoch values, the longer the time needed for training data.

3.5. Tuning Activation Function

In this test, there are two activation values tested, namely sigmoid and softmax. Other parameters used are derived from the optimal accuracy value in the previous test. Details of the activation function tests are presented in Table 8 below.

Table 8. Tuning the Activation Function on BiLSTM

Fungsi Aktivasi	Waktu (menit)	Akurasi (%)
Sigmoid	3:45	56.2
Softmax	3:55	58.65

After the test is carried out, the softmax activation function gets the optimal accuracy value, namely the accuracy result is 58.65%.

3.6. Overall Result of BiLSTM Metode Method

After testing the parameters, the optimal value of each parameter is obtained. With the word2vec CBOW value, the number of neurons is 200, the learning rate is 0.00001, the number of epochs is 25 and the activation function is softmax, the accuracy result is 58.65%.

Conclusion

The method proposed in this study is BiLSTM. After doing the research, it can be concluded that the parameters that produce the optimal accuracy value are word2vec CBOW, number of neurons 200, learning rate 0.00001, number of epochs 25, and softmax activation function. Then, with this parameter value, the accuracy is 58.56%.

References

- [1] M. M. Munir, M. A. Fauzi, and R. S. Perdana, "Implementasi Metode Backpropagation Neural Network berbasis Lexicon Based Features dan Bag of Words Untuk Identifikasi Ujaran Kebencian Pada Twitter," p. 10.

- [2] M. A. Nurrohmat and A. Sn, "Sentiment Analysis of Novel Review Using Long Short-Term Memory Method," *Indonesian J. Comput. Cybern. Syst.*, vol. 13, no. 3, p. 209, Jul. 2019, doi: 10.22146/ijccs.41236.
- [3] A. M. Ramadhani and H. S. Goo, "Twitter sentiment analysis using deep learning methods," presented at the th International Annual Engineering Seminar (InAES), Lecture Notes in Computer Science, IEEE, Cham, 2017.
- [4] R. W. Pratiwi, Y. Sari, and Y. Suyanto, "Attention-Based BiLSTM for Negation Handling in Sentimen Analysis," *Indonesian J. Comput. Cybern. Syst.*, vol. 14, no. 4, p. 397, Oct. 2020, doi: 10.22146/ijccs.60733.
- [5] F. Nausheen and S. H. Begum, "Sentiment analysis to predict election results using Python," in *2018 2nd International Conference on Inventive Systems and Control (ICISC)*, Coimbatore, Jan. 2018, pp. 1259–1262. doi: 10.1109/ICISC.2018.8399007.
- [6] A. Santoso and G. Ariyanto, "IMPLEMENTASI DEEP LEARNING BERBASIS KERAS UNTUK PENGENALAN WAJAH," *emitor*, vol. 18, no. 01, pp. 15–21, Jun. 2018, doi: 10.23917/emitor.v18i01.6235.
- [7] R. Ferdiana, F. Jatmiko, D. D. Purwanti, A. S. T. Ayu, and W. F. Dicka, "Dataset Indonesia untuk Analisis Sentimen," *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, vol. 8, no. 4, p. 334, Nov. 2019, doi: 10.22146/jnteti.v8i4.533.
- [8] A. R. Isnain, A. Sihabuddin, and Y. Suyanto, "Bidirectional Long Short Term Memory Method and Word2vec Extraction Approach for Hate Speech Detection," *Indonesian J. Comput. Cybern. Syst.*, vol. 14, no. 2, p. 169, Apr. 2020, doi: 10.22146/ijccs.51743.
- [9] A. D. Rahmawan, "Analisis Emosi Pada Twitter Review Film Dalam Bahasa Indonesia," Gadjah Mada University, Yogyakarta, 2018.