Sentiment Analysis of State Capital Relocation of Indonesia using Convolutional Neural Network

Aditya Welly Andi
Department of Informatics
UIN Sunan Gunung Djati Bandung
Bandung, Indonesia
wellyadit@gmail.com

Cepy Slamet
Department of Informatics
UIN Sunan Gunung Djati Bandung
Bandung, Indonesia
cepy_lucky@uinsgd.ac.id
https://orcid.org/0000-0001-8351-8523

Dian Sa'adillah Maylawati
Department of Informatics
UIN Sunan Gunung Djati Bandung
Bandung, Indonesia
diansm@uinsgd.ac.id
https://orcid.org/0000-0002-1193-3370

Jumadi Jumadi
Department of Informatics
UIN Sunan Gunung Djati Bandung
Bandung, Indonesia
jumadi@uinsgd.ac.id
https://orcid.org/0000-0002-4055-5065

Aldy Rialdy Atmadja
Department of Informatics
UIN Sunan Gunung Djati Bandung
Bandung, Indonesia
aldy@if.uinsgd.ac.id
https://orcid.org/0000-0001-7798-8078

Muhammad Ali Ramdhani
Department of Informatics
UIN Sunan Gunung Djati Bandung
Bandung, Indonesia
m_ali_ramdhani@uinsgd.ac.id
https://orcid.org/0000-0002-6492-067X

***Abstract*— The current government policy led by President Joko Widodo regarding relocating the capital city from *Daerah Khusus Ibukota* (Capital Special Region) of Jakarta to East Kalimantan has drawn a variety of comments, ranging from praise, criticism, suggestions, innuendo to hate speech. This is supported by many Indonesian political figures who have Twitter accounts to provide support or opinions on this policy. This study aims to analyze the sentiments about this issue. There is very varied sentiment from this issue, either positive or negative responses. This research used Convolutional Neural Network (CNN) algorithm as a part of the Deep Learning method to classify sentiments towards government policy on moving capital city of Indonesia with data obtained from Twitter. The process begins with text pre-processing containing case folding, tokenizing, stop-words removing, stemming, and changing the emoticon to word. Then, the word embedding process used Word2Vec. The result of experiment of CNN algorithm with 1,515 tweets in the Indonesian language and 15 times of experiment shows that the average accuracy is 66.68% with the highest accuracy is 70.3%. The experiment used five training and testing data splitting variations, with three epochs, among others: 10 epochs, 30 epochs, and 100 epochs.**

Keywords*— *convolutional neural network, deep learning, Indonesia issue, sentiment analysis, state capital relocation

I. INTRODUCTION

Since August 17, 1945 the Unitary State of the Republic of Indonesia (NKRI: *Negara Kesatuan Republik Indonesia*) proclaimed its independence, and there has never been a serious discussion about establishing a city as the capital of the country. The city that became the nation's capital (Jakarta) is now a relic of Dutch colonialism. The relocation of the national capital has been carried out by several countries, for various reasons [1]. The government's policy currently led by President Joko Widodo regarding relocating the capital city from the Special Capital Region Province (DKI: *Daerah Khusus Ibukota*) Jakarta to East Kalimantan has received various comments, ranging from praise, criticism, suggestions, innuendo to hate speech.

A popular communication tool among internet users at this time is a microblogging site. Where there are millions of messages present every day on microblogging websites such as Twitter. This makes many users upload issues that are developing or express their opinions on something [2]. Of the 212.35 million Internet users in Indonesia or reaching 76.8 percent of the 276.3 million Indonesian population, around 61.8 percent spend their activities on the internet to access social media [3]. Twitter is the most data source that used in text analytics research. This is because of the easy access and permission to pull data from Twitter compared to other social media. However, apart from more uncomplicated data collection techniques, Twitter is a place that is more open, open-minded, egalitarian, and a place for various trending news [4]. Compared to Facebook, which has many fake accounts and is spreading hoax news, and Instagram, which is mostly used for self-promotion and self-exposure, Twitter has far more educated users. The lack of mental judgment shows this, and if there is news that is a hoax, has a toxic smell, and is fighting with each other, Twitter users are wiser in dealing with it. In essence, social media analysis technology is currently one of the research strengths in the digital era [5]. Therefore, Twitter is considered to be more in demand by the public because it is easy to express their opinions. With so many Twitter users expressing this opinion, it can be used to find information. However, its use requires good analysis so that the information obtained can help many parties to support a decision.

There are several sentiment analysis research to find public opinion about government or politic issues, among others: (1) sentiment analysis of presidential election in Indonesia using polarity method and texblob with only 250 tweet data for training and 100 data for testing [6]; (2) comparison between Support Vector Machine and K-Nearest Neighbor to analyze the sentiment of Indonesia presidential election [7]; sentiment analysis for public policy and bureaucracy service issues in Indonesia [8]; (3) sentiment analysis to find the public opinion COVID-19 vaccination policy in Indonesia using Support Vector Machine [9]; (4) sentiment analysis to get public response about how to Indonesian government in handling COVID-19 pandemic [10]; (5) polarization of Indonesia politics before presidential election using with social network and sentiment analysis [11]; (6) sentiment analysis to monitor the performance of politics issue in Thailand [12]; (7) sentiment analysis to know the influence of politics in COVID-19 pandemic [13]; (8) sentiment analysis technique also used to find political communication through social media [14]–[16]; and many more.

Today, many sentiment analysis researchers use the Deep Learning method. Deep Learning method is considered to have higher accuracy and is consistent when faced with big data. From these advantages, Deep Learning can produce better accuracy [7]. Several of Indonesia's largest NLP communities, such as IndoLEM, are developing Indonesian NLP benchmarks (Indonesian Language Evaluation Montage) [17], IndoNLU (Indonesian Natural Language Understanding) [18], and IndoNLG (Indonesian Language for Natural Language Generation) [19]. These studies also perform NLP research, which includes sentiment analysis utilizing Deep Learning. Deep learning, which is the creation of artificial neural networks, is widely used in sentiment analysis: (1) sentiment analysis for Indonesian News using Deep Learning [20]; (2) sentiment analysis based on emoticon using Convolutional Neural Network and Support Vector Machine [21]; (3) sentiment analysis using Deep Belief Network for the Indonesian language [22]; (4) a review and comparative study of sentiment analysis using deep learning [23]–[25]; (5) sentiment analysis about Covid-19 using deep learning [26]; and so on.

For sentiment classification, of the many deep learning algorithms, the most widely used in text classification are Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), because traditional machine learning algorithms such as Support Vector Machine (SVM) and Naïve Bayes Classifier reach the limit. Where adding more training data will not increase accuracy. On the other hand, classification using deep learning improves with more data being trained [7]. The Deep Learning method currently has the most significant results in the classification is the Convolutional Neural Network (CNN). This is because CNN is trying to mimic the system in the human visual cortex. CNN has a smaller number of parameters, so it can be trained with small data [7].

The sentiments of the Indonesian people regarding the relocation of the new capital city are certainly interesting and important to study. Because moving the capital city is a big and historic decision for a country. Public opinion is certainly an important consideration, especially in a democratic country like Indonesia. Therefore, by utilizing sentiment analysis technology, this study aims to analyze the public sentiment about state capital relocation as hot issue in Indonesia using Convolutional Neural Networks. The next sections will be explained the research method, result and discussion, and the conclusion of this research.

II. Research Methods

A. *Sentiment Analysis*

Sentiment analysis is a technique to find public sentiment on specific issues [27], [28]. Sentiment analysis groups data with positive, neutral, and negative class labels related to public opinion or views from many sources, such as social media [29]–[34], movie review [35]–[37], news portal [20], review of a product [38], [39], and others. The sentiment analysis technique uses a classification approach, where the training data must be labeled (positive, negative, or neutral) so that the computer can learn to recognize patterns and predict labels on other text data.

B. *Research Activities*

This section contains research initiatives that demonstrate CNN's performance in sentiment analysis research. These tasks include data collection and labeling, text pre-processing, word embedding, sentiment analysis using CNN, and evaluation using confusion matrix, as shown in Fig 1.

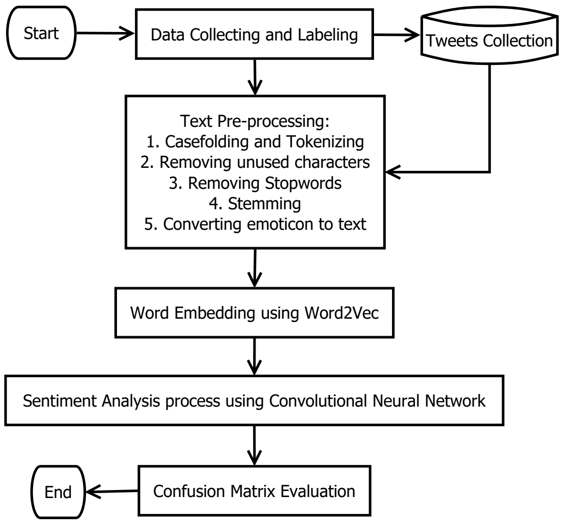

Fig. 1. Research Activities

1) *Data Collecting and Labeling*

Dataset was collected from Twitter using Twitter API. The data taken are tweets related to the relocation of the capital city of Indonesia with several keywords such as the new capital city and moving the capital city. A sampling of tweet data was carried out from 1 February 2020 to 30 May 2020. There are 1,515 tweets in Indonesian language for this research. Then, data was labeled by positive or negative sentiment.

2) *Text Pre-processing*

Text pre-processing is a key step in the mining process that prepares the data. Text pre-processing ensures the quality of input data, ensuring that the results are of the anticipated quality [40], [41]. Several text pre-processing processes are utilized in this study, including case-folding and tokenizing, deleting unneeded letters, removing stop-words, and stemming. This study uses the Sastrawi library, employing the Nazief-Adriani stemming algorithm for Indonesian language stemming [42]. Then, the emoticons that many find from social media will be converted to text.

3) *Convolutional Neural Network*

Convolutional Neural Network is a neural network method that is extensively used in image and text processing. Convolution, or simply convolution, is a matrix that filters and classifies images or text [43]. The layers of a Convolutional Neural Network are utilized to conduct filtering in each phase. The procedure is known as the training procedure. There are three layers in the training process: Convolutional, Pooling, and Fully-connected layers [44][45]. Fig 2 shows the CNN layer architecture. Where the key layer for producing new features from input data is the Convolutional layer. The Pooling layer, like the Convolutional Layer, is in charge of reducing the spatial size of the Convolved Feature. The computer power required to process the data is lowered because of dimensionality reduction. It also aids the training process by extracting rotational and positional invariant dominating features. There are two types of pooling: maximal pooling and average pooling. Max Pooling returns the maximum value from the image's Kernel-covered area. The average of all values from the Kernel's segment of the data is returned by Average Pooling. The Fully-Connected layer then begins to learn a non-linear function. Using a Fully-Connected layer to learn non-linear combinations of the high-level qualities represented by the convolutional layer's output is a cheap technique to learn non-linear combinations of the high-level characteristics represented by the convolutional layer's output.

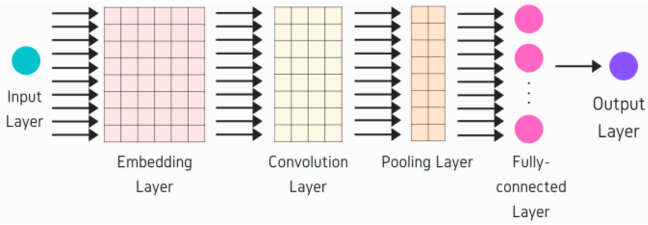

Fig. 2. CNN Architecture Illustration

4) Confusion Matrix

The classification metric used in sentiment analysis techniques is the confusion matrix to evaluate the sentiment classification result model [46]. There are values of accuracy, precision, and recall, which are calculated by the formula (1)-(3) [47]–[49].

$$precision = \frac{|TP|}{|TP|+|TN|} \quad (1)$$

$$recall = \frac{|TP|}{|TP|+|FP|} \quad (2)$$

$$accuracy = \frac{|TP|+|FN|}{|D|} \quad (3)$$

Precision is calculated by dividing the total sentiment classifications that are correctly predicted to be positive (True Positive - TP) by the number of sentiment classifications that are correctly predicted to be positive and the correct sentiment classifications are predicted to be negative in the document (True Negative - TN), so that the prediction results correctly determine whether the sentiment classification is correct, both predicted by humans and predicted by the system is correct. In a document, Recall is equal to the total number of predicted true sentiment classifications divided by the number of false-positive predicted and false-positive sentiment classifications (FP). Recall evaluates the proportion of sentiment classifications predicted by humans and those generated by the system. At the same time, accuracy is defined as the proportion of the total sentiment classification that is predicted correctly and incorrectly divided by the total sentiment classification in a document.

III. Result and Discussion

A. Text Pre-processing Result

Test pre-processing is an essential process of preparing unstructured text data into structured data so that the data can be appropriately processed. In pre-processing, several stages are divided into five parts, namely case-folding, tokenizing, filtering (including removing regular expressions and stop-words), stemming, and emoticon conversion. Table I shows the example process of text pre-processing (begin from case-folding, filtering, stop-words removing, stemming, until converting the emoticons) from the example of three tweets in the Indonesian language below:

1. *Ibu Kota pindah ke Kalimantan? Mau ngerusak paru-paru dunia? Wkwkw* (Capital city moved to Kalimantan? Want to destroy the lungs of the world? Wkwkw)
2. *Pemindahan Ibu Kota dari DKI Jakarta ke Kalimantan Timur penting untuk pemerataan wilayah pembangunan :) #Ibu KotaBaru* (Moving the capital city from DKI Jakarta to East Kalimantan is important for equitable distribution of development areas :) #New Capital)
3. *INDONESIA GAK BUTUH IBU KOTA BARU, CUMA BUTUH PRESIDEN BARUUU* (INDONESIA DON'T NEED A NEW CAPITAL, JUST NEED A NEW PRESIDENT)

TABLE I. The Example of Text Pre-processing Result

Process	**Result**
Case-folding	1. ibu kota pindah ke kalimantan? mau ngerusak paru-paru dunia? Wkwkw 2. pemindahan ibu kota dari dki jakarta ke kalimantan timur penting untuk pemerataan wilayah pembangunan :) #ibu kotabaru 3. indonesia gak butuh ibu kota baru, cuma butuh presiden baruuu
Tokenizing	1. ‘ibu’, ‘kota’, ‘pindah’, ‘ke’, ‘kalimantan?’, ‘mau’, ‘ngerusak’, ‘paru-paru’, ‘dunia?’, ‘wkwkwk’ 2. ‘pemindahan’, ‘ibu’, ‘kota’, ‘dari’, ‘dki’, ‘jakarta’, ‘ke’, ‘kalimantan’, ‘timur’, ‘penting’, ‘untuk’, ‘pemerataan’, ‘wilayah’, ‘pembangunan’, ‘:)’, ‘#ibukotabaru’ 3. ‘indonesia’, ‘gak’, ‘butuh’, ‘ibu’, ‘kota’, ‘baru,’, ‘cuma’, ‘butuh’, ‘presiden’, ‘baruuu’
Filtering and Stop-words Removing	1. ‘ibu’, ‘kota’, ‘pindah’, ‘kalimantan’, ‘ngerusak’, ‘paru paru’, ‘dunia’ 2. ‘pemindahan’, ‘ibu’, ‘kota’, ‘dki’, ‘jakarta’, ‘kalimantan’, ‘timur’, ‘penting’, ‘pemerataan’, ‘wilayah’, ‘pembangunan’, ‘:)’, ‘ibukotabaru’ 3. ‘indonesia’, ‘gak’, ‘butuh’, ‘ibu’, ‘kota’, ‘baru,’, ‘butuh’, ‘presiden’, ‘baruuu’
Stemming	1. ‘ibu’, ‘kota’, ‘pindah’, ‘kalimantan’, ‘rusak’, ‘paru-paru’, ‘dunia’ 2. ‘pindah’, ‘ibu’, ‘kota’, ‘dki’, ‘jakarta’, ‘kalimantan’, ‘timur’, ‘penting’, ‘rata’, ‘wilayah’, ‘bangun’, ‘:)’, ‘ibukotabaru’ 3. ‘indonesia’, ‘gak’, ‘butuh’, ‘ibu’, ‘kota’, ‘baru,’, ‘butuh’, ‘presiden’, ‘baruuu’
Converting emoticons to word	1. ‘ibu’, ‘kota’, ‘pindah’, ‘kalimantan’, ‘rusak’, ‘paru-paru’, ‘dunia’ 2. ‘pindah’, ‘ibu’, ‘kota’, ‘dki’, ‘jakarta’, ‘kalimantan’, ‘timur’, ‘penting’, ‘rata’, ‘wilayah’, ‘bangun’, ‘senyum’, ‘ibukotabaru’ 3. ‘indonesia’, ‘gak’, ‘butuh’, ‘ibu’, ‘kota’, ‘baru,’, ‘butuh’, ‘presiden’, ‘baruuu’

B. Word Embedding using Word2Vec

Word embedding is a type of word representation that allows for the representation of words that have similar meanings [50]. There are many word embedding representations, such as Word2Vec [51], GloVe [52], and FastText [53]. The embedding layer in this study is Word2Vec. Word2vec is a method for expressing words in N-dimensional vector form [54]. Word2Vec uses a neural network to determine contextual and semantic similarity (contextual and semantic similarity) in the form of one-hot encoded vectors when presenting a word. Contextual and semantic similarity can be used to illustrate a word's relationship to other words. Word embedding representation varies by language.

C. *Implementation of Convolutional Neural Network*

By using the linear regression layer as the output layer, we can create a simple CNN model for our purpose. The task's CNN architecture is given here. The embedding layer, which is the initial layer of the model, is found within the CNN algorithm. Words are represented as real-valued vectors in a high-dimensional space using layer embedding. This layer allows for the pre-trained word vector matrix to be used to initialize vocabulary word vectors. The input tweets are converted into numeric word token sequences, such as t_1, t_2, ..., t_N, where t_N represents the original word, and N is the token vector length. To keep the result size identical for tweets of varying length, the authors limit the maximum value of N to the maximum tweet length of all tweets. Any tweet was shorter than N will be populated to N using zeros. In the convolutional layer, the m filter is used to extract the local n-gram features from the matrix of the previous layer. In a sliding window of width w indicating that w-grams can be extracted, the F_l filter ($1 \le l \le m$) studies the feature map as formula below:

$$y_i^l = f(T_{i;i+w-1} \circ W^l + b^l)$$

Where $\circ$ is a convolutional operation, $W \in Rw \times d$ is the weight matrix of the token vector record output ($t_i\ t_{i+1}$, ..., t_{i+w-1}, where ifk> previous layer, b is bias, and $T_{i:\,i+w-1}$ dari mana $y_l = 0$, $t_k = 0$). The result of F_l is $y_l \in R_d$, i is an element of y_l. This research uses Re-LU as an activation function for a more simple and accurate calculation [55].

There are several stages in the classification process, namely Convolutional Layer, Pooling Layer, and Fully Connected Layer. After the pre-processing stage, the module's word (tweet) input will be converted into an array of tokens with specific dimensions using word embedding and then mapped in the feature matrix or sentence matrix by the embedding layer. Furthermore, in the data convolution layer, the weight is calculated layer by layer with a filter layer to perform calculations on each embedding layer. This layer is used to form the weights that the input layer has with the hidden layer. Then, the largest value weight is taken from each filter layer using a pooling layer. The fully connected layer finally composes these features to output the final regression results with a linear decoder. After doing the pre-training then it is done to train the backpropagation algorithm to train the error rate. A loss function is an error in the training data set. Loss validation is an error after running a data validation set.

D. *Experiment Scenario and Result*

The experiment was carried out on 1,515 data that had been obtained from Twitter, the data was first processed until the data was ready for the classification process. The processing consists of several processes, namely case-folding, filtering, and tokenizing. The testing process is carried out by changing three types: the amount of training data, test data, and the number of epochs. The data for testing is divided into five times testing. The first test was carried out with 90% training data, 10% test data, and ten epochs. The second test was carried out with 80% training data, 20% test data, and ten epochs. The third test used 70% training data and 30% test data with epoch 10. In the fourth test, the amount of training data is 60%, 40% is test data, and epoch 10. The last test or fifth test is with the amount of training data is 50%, 50% test data, and epoch 10. After that, the next test was carried out by dividing the amount of training data and testing data the same as before but increasing the number of epochs to 30 and 100. Totally, there are 15 experiment variations in this research. Then, the result of the experiment is evaluated using a confusion matrix. Table II shows the accuracy result from the confusion matrix evaluation for each experiment scenario.

TABLE II. ACCURACY RESULT OF EXPERIMENT

Epoch	**Percentage of Data**		**Accuracy**
	Training	**Testing**	
10	90%	10%	69.70%
	80%	20%	67.80%
	70%	30%	67.60%
	60%	40%	67.30%
	50%	50%	65.20%
30	90%	10%	68.40%
	80%	20%	66.90%
	70%	30%	66.50%
	60%	40%	67.90%
	50%	50%	67.50%
100	90%	10%	70.30%
	80%	20%	69.30%
	70%	30%	66.30%
	60%	40%	66.30%
	50%	50%	64.90%
Average of Accuracy			**67.46%**

E. *Discussion*

The graph in Fig. 3 shows tests 1-5 with epochs of 10 producing an accuracy value of 65.2% for 50:50 data sharing, then in the second test it rose to 67.3% for 60:40 data division. Then the third test rose again to 67.6% with a data share of 70:30. Furthermore, the accuracy becomes 67.8% in the fourth test for 80:20 data sharing, and the last one increases the accuracy on 90:10 data sharing with an accuracy value of 69.7%. Furthermore, the value of epochs or iterations was added to 30 and retested in the order as before and resulted in an accuracy of 67.5%, 67.9%, 66.5%, 66.9%, and 68.4%. For the epochs value of 100, the accuracy values obtained are 64.9%, 66.3%, 66.3%, 69.3%, and 70.3%. Thus, it is clear that the predictions generated from the Convolutional Neural Network algorithm depend on the epochs and the amount of data being trained. This is because the accuracy of each test increases when more training data and epochs are used.

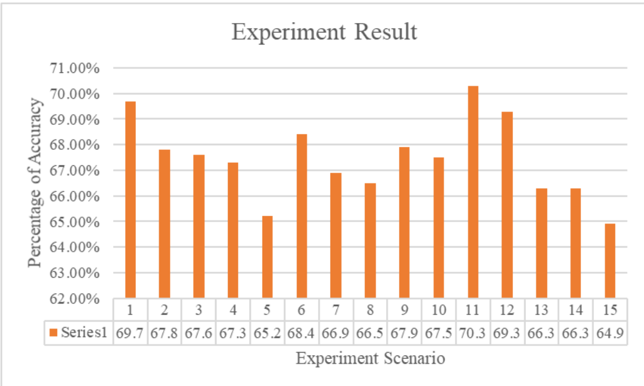

Fig. 3. Graphic of Experiment Result

After testing 15 times with different epochs, the results obtained are as Tabel II, where the average test performed with ten epochs is 67.42%. Meanwhile, if more epochs are added, the accuracy will increase to 67.44% and 67.52%. Epochs are when the entire dataset has gone through the training process on the Neural Network until it is returned to the beginning for one round. Another thing that affects is the batch size. However, in this study, the batch size was not changed to examine the best epochs and networks in classifying this convolutional neural network.

The evaluation process shows that the system is not running well. This is due to various factors, one of which is that less data is used for deep learning. The data used is only 1,515 tweets. This is because the data retrieval time using the API with the library has limited data crawling, which can only be retrieved a maximum of 7 days from the time of retrieval, the sample data was taken on February 24. This indicates that the maximum tweet data will be retrieved only February 17 and so on. In addition to data with less amount, labeling is also less than optimal because linguists do not carry out the labeling, so the labeling between one tweet and another tweet is irrelevant.

Another influencing factor is unbalancing data, meaning that the data used is not balanced between positive and negative. In several experiments carried out, several results showed the system classified the test tweets to be negative. More than 50% of the data are tweets with negative labels so that when the machine does learning, it cannot represent a positive label. Other factors that affect the success of this system are libraries in Indonesian in python programming that are less qualified and less than optimal. In addition, the limitation in modifying data in the form of text is more complicated than data in the form of images.

CONCLUSION

Sentiment analysis of data on the transfer of the Indonesian capital using the Convolutional Neural Network algorithm has been successfully carried out by retrieving data from Twitter, then processing it by pre-processing and weighting it with Word2Vec the data is ready to be classified by CNN. Then the CNN algorithm predicts the testing data. After previously labeling all data so that the algorithm accuracy value can be calculated. The results of testing the CNN algorithm show that the highest accuracy is 70.3%, with an average value of 66.68%. After labeling the data generated from Twitter data related to the policy of moving the Indonesian capital city, taking positive and negative sentiments into account. From the results of classification and labeling, it is found that changing emoticons into words affects the accuracy value, and negative sentiments have more numbers than positive sentiments. However, the prediction of positive sentiment is more accurate than negative sentiment. For further works, the library of emoticons should be more complete so that the converting emoticons process can be optimized to provide maximum accuracy values. Then, another deep learning algorithm can be used to enhance the accuracy and sentiment prediction.

REFERENCES

[1] M. Yahya, “Pemindahan Ibu Kota Negara Maju dan Sejahtera,” *J. Stud. Agama dan Masy.*, vol. 14, no. 1, p. 21, 2018, doi: 10.23971/jsam.v14i1.779.

[2] I. Taufik dan S.A.Pamungkas, “Analisis Sentimen Terhadap Tokoh Publik Menggunakan Algoritma Support Vector Machine (SVM),” *J. LOGIK@*, vol. 8, no. 1, pp. 71–79, 2018.

[3] R. K. Nistanto, “Berapa Lama Orang Indonesia Akses Internet dan Medsos Setiap Hari?,” *Kompas.com*, 2021. https://tekno.kompas.com/read/2021/02/23/11320087/berapa-lama-orang-indonesia-akses-internet-dan-medsos-setiap-hari-?page=all#:~:text=Dari total populasi Indonesia sebanyak,3 persen dibandingkan tahun lalu.

[4] “Ciri Khas Yang Membedakan Pengguna Twitter, Instagram dan Facebook,” *dipa14.web.id*, 2020. https://dipa14.web.id/2020/12/16/ciri-khas-yang-membedakan-pengguna-twitter-instagram-dan-facebook/ (accessed Aug. 11, 2021).

[5] A. I. Kabir, R. Karim, S. Newaz, and M. I. Hossain, “The Power of Social Media Analytics: Text Analytics Based on Sentiment Analysis and Word Clouds on R.,” *Inform. Econ.*, vol. 22, no. 1, 2018.

[6] W. Budiharto and M. Meiliana, “Prediction and analysis of Indonesia Presidential election from Twitter using sentiment analysis,” *J. Big data*, vol. 5, no. 1, pp. 1–10, 2018.

[7] F. Firmansyah *et al.*, “Comparing Sentiment Analysis of Indonesian Presidential Election 2019 with Support Vector Machine and K-Nearest Neighbor Algorithm,” in *2020 6th International Conference on Computing Engineering and Design (ICCED)*, 2020, pp. 1–6.

[8] L. Muliawaty, K. Alamsyah, U. Salamah, and D. S. Maylawati, “The concept of big data in bureaucratic service using sentiment analysis,” *Int. J. Sociotechnology Knowl. Dev.*, 2019, doi: 10.4018/IJSKD.2019070101.

[9] M. Ritonga, M. A. Al Ihsan, A. Anjar, and F. H. Rambe, “Sentiment analysis of COVID-19 vaccine in Indonesia using Naive Bayes Algorithm,” in *IOP Conference Series: Materials Science and Engineering*, 2021, vol. 1088, no. 1, p. 12045.

[10] P. H. Prastyo, A. S. Sumi, A. W. Dian, and A. E. Permanasari, “Tweets responding to the Indonesian Government’s handling of COVID-19: Sentiment analysis using SVM with normalized poly kernel,” *J. Inf. Syst. Eng. Bus. Intell.*, vol. 6, no. 2, pp. 112–122, 2020.

[11] M. N. Habibi, “Analysis of Indonesia Politics Polarization before 2019 President Election Using Sentiment Analysis and Social Network Analysis.,” *Int. J. Mod. Educ. Comput. Sci.*, vol. 11, no. 11, 2019.

[12] W. Chansanam and K. Tuamsuk, “Thai Twitter Sentiment Analysis: Performance Monitoring of Politics in Thailand using Text Mining Techniques,” *Int. J. Innov. Creat. Chang.*, vol. 11, no. 12, pp. 436–452, 2020.

[13] A. Mukherjee, S. Bhattacharyya, K. Ray, B. Gupta, and A. K. Das, “A Study of Public Sentiment and Influence of Politics in COVID-19 Related Tweets,” in *Computational Intelligence in Pattern Recognition*, Springer, 2022, pp. 655–665.

[14] M. Haselmayer and M. Jenny, “Sentiment analysis of political communication: combining a dictionary approach with crowdcoding,” *Qual. Quant.*, vol. 51, no. 6, pp. 2623–2646, 2017.

[15] Y. Matalon, O. Magdaci, A. Almozlino, and D. Yamin, “Using sentiment analysis to predict opinion inversion in Tweets of political communication,” *Sci. Rep.*, vol. 11, no. 1, pp. 1–9, 2021.

[16] J. Pal, “Studying political communication on Twitter: the case for small data,” *Curr. Opin. Behav. Sci.*, vol. 18, pp. 97–102, 2017.

[17] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, “IndoLEM and IndoBERT: A benchmark dataset and pre-trained language model for Indonesian NLP,” *arXiv Prepr. arXiv2011.00677*, 2020.

[18] B. Wilie *et al.*, “IndoNLU: Benchmark and resources for evaluating Indonesian natural language understanding,” *arXiv Prepr. arXiv2009.05387*, 2020.

[19] S. Cahyawijaya *et al.*, “IndoNLG: Benchmark and Resources for Evaluating Indonesian Natural Language Generation,” *arXiv Prepr. arXiv2104.08200*, 2021.

[20] M. Z. Negara, G. F. Shidik, A. Z. Fanani, and E. Noersasongko, “Sentiment Analysis of Indonesian News Using Deep Learning (Case Study: TVKU Broadcast),” in *2018 International Seminar on Application for Technology of Information and Communication*, 2018, pp. 261–265.

[21] C. Daniel and S. Loganathan, “A Comparison of Machine Learning and Deep Learning Methods with Rule Based Features for Mixed

Emotion Analysis," *Int. J. Intell. Eng. Syst.*, vol. 14, no. 1, pp. 42–53, 2021, doi: 10.22266/IJIES2021.0228.05.

[22] I. Zulfa and E. Winarko, "Sentimen Analisis Tweet Berbahasa Indonesia Dengan Deep Belief Network," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 11, no. 2, p. 187, 2017, doi: 10.22146/ijccs.24716.

[23] A. Yadav and D. K. Vishwakarma, "Sentiment analysis using deep learning architectures: a review," *Artif. Intell. Rev.*, vol. 53, no. 6, pp. 4335–4385, 2020.

[24] N. C. Dang, M. N. Moreno-Garcia, and F. De la Prieta, "Sentiment analysis based on deep learning: A comparative study," *Electronics*, vol. 9, no. 3, p. 483, 2020.

[25] O. Habimana, Y. Li, R. Li, X. Gu, and G. Yu, "Sentiment analysis using deep learning approaches: an overview," *Sci. China Inf. Sci.*, vol. 63, no. 1, pp. 1–36, 2020.

[26] K. Chakraborty, S. Bhatia, S. Bhattacharyya, J. Platos, R. Bag, and A. E. Hassanien, "Sentiment Analysis of COVID-19 tweets by Deep Learning Classifiers - A study to show how popularity is affecting accuracy in social media," *Appl. Soft Comput.*, vol. 97, p. 106754, 2020.

[27] B. Lin, F. Zampetti, M. Di Penta, R. Oliveto, G. Bavota, and M. Lanza, "Sentiment Analysis for Software Engineering: How Far Can We Go?," *2018 ACM/IEEE 40th Int. Conf. Softw. Eng. 117*, 2018, doi: 10.1145/3180155.3180195.

[28] B. Liu, "Sentiment analysis and opinion mining," *Synth. Lect. Hum. Lang. Technol.*, vol. 5, no. 1, pp. 1–184, 2012, doi: 10.2200/S00416ED1V01Y201204HLT016.

[29] A. P. Rodrigues, A. Rao, and N. N. Chiplunkar, "Sentiment Analysis of Real Time Twitter Data Using Big Data Approach," *2nd Int. Conf. Comput. Syst. Inf. Technol. Sustain. Solut. CSITSS 2017*, pp. 1–6, 2018, doi: 10.1109/CSITSS.2017.8447656.

[30] A. Ali, "Sentiment Analysis on Twitter Data using KNN and SVM," vol. 8, no. 6, pp. 19–25, 2017.

[31] S. Rosenthal, N. Farra, and P. Nakov, "SemEval-2017 Task 4: Sentiment Analysis in Twitter," in *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, 2017, doi: 10.18653/v1/S17-2088.

[32] L. Branz and P. Brockmann, "Sentiment Analysis of Twitter Data," in *Proceedings of the 12th ACM International Conference on Distributed and Event-based Systems - DEBS '18*, 2018, doi: 10.1145/3210284.3219769.

[33] A. Alrumaih, A. Al-Sabbagh, R. Alsabah, H. Kharrufa, and J. Baldwin, "Sentiment analysis of comments in social media," *Int. J. Electr. Comput. Eng.*, vol. 10, no. 6, pp. 5917–5922, 2020, doi: 10.11591/ijece.v10i6.pp5917-5922.

[34] G. A. Ruz, P. A. Henríquez, and A. Mascareño, "Sentiment analysis of Twitter data during critical events through Bayesian networks classifiers," *Futur. Gener. Comput. Syst.*, vol. 106, pp. 92–104, 2020, doi: 10.1016/j.future.2020.01.005.

[35] D. S. Maylawati, M. N. Mudyawati, M. H. Wahisyam, and R. A. Maulana, "Comparison of Classification Algorithms for Sentiment Analysis on Movie Comments," in *Gunung Djati Conference Series*, 2021, vol. 3, pp. 68–80.

[36] C. Nanda, M. Dua, and G. Nanda, "Sentiment Analysis of Movie Reviews in Hindi Language Using Machine Learning," *Proc. 2018 IEEE Int. Conf. Commun. Signal Process. ICCSP 2018*, pp. 1069–1072, 2018, doi: 10.1109/ICCSP.2018.8524223.

[37] A. Z. Amrullah, A. Sofyan Anas, and M. A. J. Hidayat, "Analisis Sentimen Movie Review Menggunakan Naive Bayes Classifier Dengan Seleksi Fitur Chi Square," *J. Bumigora Inf. Technol.*, vol. 2, no. 1, pp. 40–44, 2020, doi: 10.30812/bite.v2i1.804.

[38] X. Fang and J. Zhan, "Sentiment analysis using product review data," *J. Big Data*, 2015, doi: 10.1186/s40537-015-0015-2.

[39] K. Marimuthu, A. Shankar, R. Ranganathan, and R. Niranchana, "Product opinion analysis using text mining and analysis," *Int. J. Smart Grid Green Commun.*, vol. 1, no. 3, p. 227, 2018, doi: 10.1504/IJSGGC.2018.091351.

[40] D. S. Maylawati, H. Sugilar, and R. A. Yudhiantara, "KUALITAS PERANGKAT LUNAK: MODULARITAS PUSTAKA TEXT PRE-PROCESSING," *J. Perspekt.*, 2018, doi: 10.15575/jp.v1i2.11.

[41] S. Kannan *et al.*, "Preprocessing Techniques for Text Mining," *Int. J. Comput. Sci. Commun. Networks*, 2015.

[42] M. Adriani, J. Asian, S. M. M. T. . Nazief, and H. . Williams, "Stemming Indonesian: A Confix-stripping approach," *ACM Trans. Asian Lang. Inf. Process.*, vol. 6, no. 1, pp. 1 – 33, 2007.

[43] A. Santoso and G. Ariyanto, "Implementasi Deep Learning Berbasis Keras Untuk Pengenalan Wajah," *Emit. J. Tek. Elektro*, vol. 18, no. 01, pp. 15 – 21, 2018.

[44] S. Saha, "A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way," *towardsdatascience.com*, 2018. https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53.

[45] S. Sena, "Pengenalan Deep Learning Part 7 : Convolutional Neural Network (CNN)," *Medium.com*, 2017. https://medium.com/@samuelsena/pengenalan-deep-learning-part-7-convolutional-neural-network-cnn-b003b477dc94 (accessed Aug. 02, 2019).

[46] K. M. Ting, "Confusion Matrix," in *Encyclopedia of Machine Learning and Data Mining*, 2017.

[47] M. Gambhir and V. Gupta, "Recent automatic text summarization techniques: a survey," *Artif. Intell. Rev.*, 2017, doi: 10.1007/s10462-016-9475-9.

[48] R. Rautray and R. C. Balabantaray, "An evolutionary framework for multi document summarization using Cuckoo search approach: MDSCSA," *Appl. Comput. informatics*, vol. 14, no. 2, pp. 134–144, 2018.

[49] P. Verma, S. Pal, and H. Om, "A Comparative Analysis on Hindi and English Extractive Text Summarization," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 18, no. 3, pp. 1–39, 2019.

[50] J. Brownlee, "What Are Word Embeddings for Text?," *Machine Learning Mastery*, 2019. https://machinelearningmastery.com/what-are-word-embeddings/ (accessed Apr. 13, 2020).

[51] S. Banerjee, "Word2Vec — a baby step in Deep Learning but a giant leap towards Natural Language Processing," 2018. https://medium.com/explore-artificial-intelligence/word2vec-a-baby-step-in-deep-learning-but-a-giant-leap-towards-natural-language-processing-40fe4e8602ba.

[52] H. Zamani and W. B. Croft, "Relevance-based Word Embedding," in *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval - SIGIR '17*, 2017, doi: 10.1145/3077136.3080831.

[53] M. Tezgider, B. Yildiz, and G. Aydin, "Improving Word Representation by Tuning Word2Vec Parameters with Deep Learning Model," in *2018 International Conference on Artificial Intelligence and Data Processing, IDAP 2018*, Jan. 2019, doi: 10.1109/IDAP.2018.8620919.

[54] K. K. D. Assistant, "WORD2VEC AND TF-IDF WORD EMBEDDING," vol. 5, no. 3, pp. 2348–2350, 2018.

[55] D. Liu, "A Practical Guide to ReLU," *Medium.com*, 2017. https://medium.com/@danqing/a-practical-guide-to-relu-b83ca804f1f7 (accessed May 09, 2020).

[56] N. Ahmed, T. Natarajan, and K. R. Rao, "Discrete Cosine Transform," *IEEE Trans. Comput.*, vol. C–23, no. 1, pp. 90–93, 1974, doi: 10.1109/T-C.1974.223784.

[57] Z. M. Hafed and M. D. Levine, "Face recognition using the discrete cosine transform," *Int. J. Comput. Vis*, vol. 43, no. 3, pp. 167–188, 2001.

[58] R. Krasmala, A. Budimansyah, and U. T. Lenggana, "Kompresi Citra Dengan Menggabungkan Metode Discrete Cosine Transform (DCT) dan Algoritma Huffman," *J. Online Inform.*, vol. 2, no. 1, pp. 1–9, 2017, doi: 10.15575/join.v2i1.79.

[59] Q. Yang and Y. Cai, "A digital image watermarking algorithm based on discrete wavelet transform and discrete cosine transform," in *Proceedings of 2012 International Symposium on Information Technologies in Medicine and Education, ITME 2012*, 2012, vol. 2, pp. 1102–1105, doi: 10.1109/ITiME.2012.6291492.

[60] K. Singh and P. A. Mishra, "A Survey on Digital Image Authentication by DCT and RPM Based Watermarking," *IOSR J. Comput. Eng.*, vol. 18, no. 04, pp. 37–42, 2016, doi: 10.9790/0661-1804013742.