Sentiment Analysis on Social Distancing and Physical Distancing on Twitter Social Media using Recurrent Neural Network (RNN) Algorithm

Fikri Aldi Nugraha[1], Nisa Hanum Harani[2], Roni Habibi[3], Rd. Nuraini Siti Fatonah[4]
[1,2,3,4] Department of Informatics, Politeknik Pos Indonesia, Bandung, Indonesia

Article Info	ABSTRACT

The government is seeking preventive steps to reduce the risk of the spread of Covid-19, one of which is social restrictions that have become popular with social distancing and physical distancing. One way to assess whether the steps taken by the government regarding social and physical distancing are accepted or not by the community is by conducting sentiment analysis. The process of sentiment analysis is carried out using a variant of the Recurrent Neural Network (RNN), namely Long Short-Term Memory (LSTM). In this study, the results obtained from the sentiment analysis, where the public response to social distancing and physical distancing has more positive sentiments than negative sentiments. To measure the accuracy level of sentiment analysis using the Recurrent Neural Network (RNN) algorithm and evaluation of the modeling is done using confusion matrix where the results obtained for the training dataset are 89% accuracy, 89% recall, 89% precision, and 89% F1 Score. Meanwhile, for the test dataset, an accuracy of 80% was obtained, a recall of 79%, a precision of 81%, and an F1 score of 80%.

Corresponding Author:

Fikri Aldi Nugraha,
Department of Informatics,
Politeknik Pos Indonesia,
Jl. Sari Asih No.54, Kel. Sarijadi, Kec. Sukasari, Kota Bandung, Jawa Barat, 40151, Indonesia
Email: fikrialdinugraha@gmail.com

1. INTRODUCTION

Currently, there are widespread infectious diseases. Infectious disease can be transmitted or transmitted to humans caused by viruses, bacteria, fungi, and parasites, whether those that only infect a person can even cause extraordinary events or outbreaks. A disease outbreak can be defined as an outbreak of an infectious disease that occurs in a community where the number of infected patients exceeds the usual conditions in a certain time and area and can cause catastrophe [1]. One of the current outbreaks of disease, namely the Novel Coronavirus 2019 (COVID-19), Wuhan, the capital of Hubei province in Central China, is the first capital city to diagnose sufferers of unusual pneumonia caused by the COVID-19 virus. Coronavirus is a virus that belongs to the superdomain of the biota, virus kingdom, and is the largest group of viruses in the order Nidovirales [2].

Indonesia is one of the countries affected by COVID-19 exposure. Quoted from the worldometer.info website as of March 25, 2020, the number of people who were positively infected was 790 people, 31 people were declared cured, and 58 people died [3]. From this data, the mortality rate is greater than the number of recovered patients. Therefore, it is necessary to make efforts to prevent COVID-19 from spreading to all Indonesian citizens.

The government strives for preventive measures to reduce the risk of spread, one of which is social distancing, which has become popular with the terms social distancing and physical distancing. The efforts made by the government are in line with Law No. 6 of 2018 article 59 paragraph 2, namely "Large-scale social restrictions aim to prevent the widespread of public health emergencies that are occurring between people in certain areas" [4]. However, the government's move has received a different response from the public, one of which can be seen on social media.

Based on a survey conducted by Hootsuite We Are Social in January 2020, the total active users of social media via the internet in Indonesia reached 175.4 million people, meaning that it reached 66% of Indonesia's total population. Users of social media applications in Indonesia, including youtube, WhatsApp, Facebook, Instagram, Line, Twitter, FB Messenger, etc., reach 160 million [5]. Therefore, the response of the Indonesian people on social media regarding social distancing and physical distancing must be different. One way to assess whether the steps taken by the government regarding social and physical distancing are accepted or not by the community is by conducting sentiment analysis. Sentiment analysis is a process carried out to analyze and determine the sentiment or opinion of a person or group which is visualized in the form of text and can be categorized as positive or negative sentiment [6].

The research conducted is inseparable from the results of previous studies that have been carried out for consideration and study. The research results that are used as a comparison are inseparable from the research topic, namely regarding the sentiment analysis of the public response to social media.

Based on the results of research conducted by AH, Alamoodi (2020), which conducted research aimed at reviewing and analyzing articles about the incidence of various types of infectious diseases, such as epidemics, pandemics, viruses or outbreaks, during the last ten years, understand the application of sentiment analysis and obtain the most important literature findings. . Articles on related topics are searched systematically in five main databases, namely, ScienceDirect, PubMed, Web of Science, IEEE Xplore, and Scopus. This Index is deemed broad and reliable enough to cover our coverage of the literature. Articles were selected based on our inclusion and exclusion criteria for systematic review, with a total of n = 28 articles selected. All of these articles are formed into a coherent taxonomy to describe the current viewpoints that fit in the literature according to four main categories: lexicon-based, machine learning-based, hybrid-based, and individual models. Despite the relatively low number of studies in this area, up-to-date data is critical for fighting future similar outbreaks and for dealing with this crisis, not only as physicians and medical researchers but also as scientists from all domains, communities, and decision-making bodies. From a computer science perspective, integrating other technologies such as AI, ML, and different analytical procedures can contribute to making a difference in making solutions to problems. [7].

Gonzalo A.Ruz (2020) presents research on sentiment analysis through machine learning using Twitter data. This study discusses the issue of sentiment analysis during critical events such as natural disasters or social movements. The study considered classifying the Bayesian network to conduct sentiment analysis on two Spanish data sets: the 2010 Chile earthquake and the 2017 Catalan independence referendum. To automatically control the number of edges supported by training examples in Bayesian network classifiers, the study adopted the Bayes factor approach to this purpose, resulting in a more realistic network. The results showed the effectiveness of Bayes factor size use and its competitive prediction results compared to vector support machines and random forests, with an adequate number of training examples. In addition, the resulting network makes it possible to identify relationships between words, offering interesting qualitative information to historically and socially understand the main features of event dynamics [8].

This research will discuss the sentiment analysis of Indonesian people's response on Twitter social media to the topic that is being discussed, namely about social distancing and physical distancing on Twitter using Recurrent Neural Network (RNN) algorithm because RNN algorithm is a multi-layer neural network that can be used to analyze sequential inputs, such as text, speech or video, for classification and prediction purposes. In addition, RNN algorithms are not limited by input length and can use temporal context to better predict meaning [9]. RNN method also has higher accuracy in processing results compared to other methods [10]. The purpose of this study is to group people's responses to positive or negative sentiment categories.

2. METHOD

Recurrent Neural Network (RNN) is one of the deep learning methods and part of a neural network that can be used to analyze sentiment with output results based on previous calculation results using sequential information. RNN is an efficient model for sentiment analysis because it uses memory cells capable of capturing sequential information without being limited by input length [11].

Long Short-Term Memory (LSTM) is one of the RNN variants proposed by Hochreiter in 1977 that can be used to study remote dependencies for text sequences. LSTM aims to solve RNN's problem of gradient vanishing and exploding, LSTM replaces RNN hidden vectors with memory blocks equipped with gates to control text flow [12]. There are three main components in the memory block in the LSTM layer, namely the input gate, forget gate, and output gate, which are useful to control the flow of information [13]. The main components of the LSTM layer are processed in the model development as shown in Figure 1.
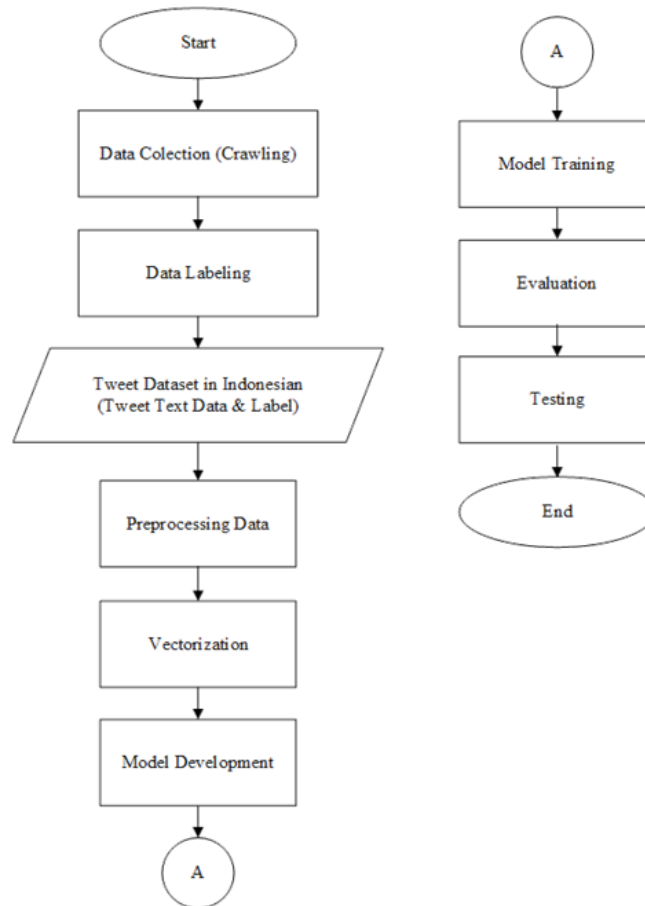
Figure 1. Flowchart of the Design Process for Determining Sentiment Classification

The following are the steps taken in conducting a sentiment analysis using the RNN algorithm in Figure 1 [14].

1. Data Collection: This process is a step taken to obtain Indonesian-language tweets from social media Twitter based on the keywords physical distancing and social distancing using crawling techniques. Then the data were divided into training data and test data.
2. Data Labeling: This process is a step taken to label the training data so that it can be used in the processing process.
3. Tweet Dataset in Indonesian: Dataset is tweet data that has been cleaned and labeled to be used as training data in the data processing.
4. Pre-process Data: It is from preparing data before processing or learning by computers. Several steps will be carried out in pre-processing the data, such as converting text to lowercase (case folding), cleaning data (data cleaning), and tokenizing.
5. Vectorization: After the dataset for training data goes through the pre-processing stage, next is vectorization, which is the process of converting the pre-processed data into a vector so that words can be counted and used as input for the processing using the RNN algorithm.
6. Model Development: In this stage, the modeling process is carried out. The model is built using the RNN algorithm, where the type of RNN used is LSTM because it has a more accurate ability to process information. LSTM has the advantage that its architecture can remember and forget the output that will be processed back to the input.
7. Model Training: At the model training stage, the data used is the result of the word embedding process. Then the input will be processed at the encoder layer. In the encoder layer, the input will be processed by the LSTM cell into an output, which is then forwarded to the hidden layer. In the hidden layer, data will be recorded that are trained in the form of logistics which is used to update the weights and values needed so that output can be achieved. The decoder layer will then receive the process results from the hidden layer, which will then be focused on using attention to produce a weight value that can approach the target output value.
8. Evaluation: At this stage, an evaluation of the model will be carried out as well as a performance test of the model, such as measurement of accuracy, precision, recall, f-measure (f1-score).

9. Testing: In the testing phase, the model will be tested using test data to measure the extent to which the classifier can classify correctly. Therefore, it should be noted that the data in the test data must be different from the data in the training data so that it can be seen that the classifier has done its job correctly.

2.1 Experiment

Modeling is a process to create a model of the system. The purpose of making modeling is to analyze and provide predictive results that can be closer to reality before the system is applied in the field [15]. The data used in modeling is public opinion in the form of tweets from social media Twitter. Data retrieval is done by crawling data using the Twitter API, to get tweets with the keywords social distancing and physical distancing within a certain period. The data collection period was carried out from April to June 2020 with the amount of data obtained as many as 9594 for tweets with the keyword social distancing and 9414 for tweets with the keyword physical distancing. Then the crawled data is filtered, namely by removing duplicate data and tweet data that tends not to be Indonesian. The data that has gone through the filtering process is further divided into training data, and test data with a percentage of 80% and 20%, where the training data is 12804 and the test data is 3298 data consisting of 1815 data with physical distancing keywords and 1483 data with words the key to social distancing. Then, the accuracy obtained was still below 70% after modeling, so the data was added for training data from the tweet in July as many as 7136 data so that the overall training data used was 19410 data. Jupyter Notebook is an application that is used to design, develop, implement, and communicate with special code [16]. The following is a flow chart for modeling the tweet data sentiment analysis using a jupyter notebook as shown in Figure 2.
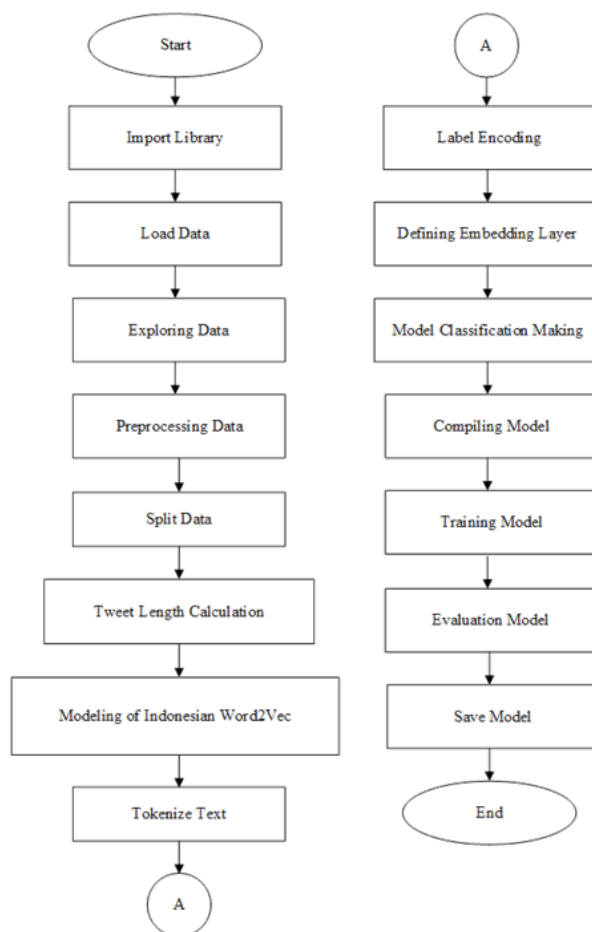

Figure 2. Modeling Flowchart

2.1.1 Import Library

The process carried out in the import library stage is to call the functions used during the modeling process. Some of the libraries used in modeling include Keras, Gensim, Pandas, Matplotlib, Sklearn, and NumPy as shown in Figure 3.

```
# DataFrame
import pandas as pd

# Matplot
import matplotlib.pyplot as plt
%matplotlib inline

# Scikit-Learn
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder
from sklearn.metrics import confusion_matrix, classification_report, accuracy_score, recall_score, precision_score, f1_score

# Keras
from keras.preprocessing.text import Tokenizer
from keras.preprocessing.sequence import pad_sequences
from keras.models import Sequential, load_model, Model
from keras.layers import Activation, Dense, Dropout, Embedding, LSTM
from keras import utils
from keras.optimizers import RMSprop

# Word2vec
import gensim
from gensim.models import Word2Vec

# Utility
import re
import numpy as np
import os
from collections import Counter
import logging
import time
import pickle
import itertools
import html
import string as str
from elang.plot.utils import plot2d, plotNeighbours

# Widget
from ipywidgets import widgets
```

Figure 3. Import Library

2.1.2 Load Data and Explore Data

The command used in this stage is pd.read_csv, where this function is a function owned by the Pandas library and is used to read data that has csv and txt formats and read data can be formed into a data frame, which later can be initialized into a variable that serves to facilitate data call. Then in this stage, the process of exploring the data is also carried out, which aims to see the distribution of each data. In this study, explored data was carried out to see the comparison of the amount of data, where tweets that have positive sentiment are 9970 data, and negative sentiment are 9970 data as shown in Figure 4.

```
DATA_KBBI    = [kamus.strip('\n').strip('\r') for kamus in open('kamus/slangword.txt')]

wiki_corpus = pd.read_csv('idwiki_new_lower.txt', header = None, names = ['teks_wikipedia'])

df = pd.read_csv('data/data_train_V3.csv')
df.head(5)
```

	Tweet	Label
0	@AXISgsm Semoga Kota Jakarta makin bebas dari ...	1
1	Lagi physical distancing mba. Masa mendem E ad...	0
2	@Aprilaola @shitlicious Kita ketemu dionline a...	1
3	Mas, mangaaaaaat mas.\r\nTetep physical distan...	1
4	@yolandwio Apakah saya stay at home dan physic...	0

```
Label_cnt = Counter(df.Label)

plt.figure(figsize=(16,8))
plt.bar(Label_cnt.keys(), Label_cnt.values())
plt.title("Dataset Labels Distribution", fontsize="x-large")
```

Text(0.5, 1.0, 'Dataset Labels Distribution')

Figure 4. Load and Explore Data

2.1.3 Preprocessing and Data Split

The pre-processing process or data preparation is a step taken to clean data and convert text. Some of the steps taken at the pre-processing stage are deleting URLs, deleting emojis, and special symbols on Twitter, deleting punctuation marks, and the process of replacing abbreviated words and slang words. The data split process is carried out using one of the functions of the Sklearn library, namely train_test_split. This process

aims to divide the cleaned dataset into training datasets and test datasets with a percentage of 80% or 1592 records and 20% or 3988 records.

Table 1. Tweets Before and After Preprocessing

Tweet Before Preprocessing	Tweet After Preprocessing
Di tengah pandemi covid-19 (virus corona), alangkah baiknya jika kita tetap di rumah saja dan jangan lupa terus menerapkan physical distancing. Bersama kita bisa memutus rantai penyebaran virus corona. https://t.co/ADaiOcQIJe	di tengah pandemi covid virus corona alangkah baiknya jika kita tetap di rumah saja dan jangan lupa terus menerapkan physical distancing bersama kita bisa memutus rantai penyebaran virus corona
@vintaratu Ga ada yang jual diri soalnya lagi physical distancing kak	tidak ada yang jual diri soalnya lagi physical distancing kak
@roseyannepark Gamauu ah, kan physical distancing :(	gamauu ah kan physical distancing

2.1.4 Tweet Length Calculation

This process aims to find out the length of the tweets in the dataset, where later tweets that are less than the maximum length will be padded to match the size. For example, in this study, the maximum length of identified tweets was 55, so that a tweet that was less than 55 would be padded at the tokenize stage to make it the same length.

2.1.5 Modeling of Indonesian Word2vec

Several processes are carried out to make the word2vec model, including separating the words from each sentence and collecting them in a variable for the next training process. After the word separation process is carried out, the process of making a model for word2vec is carried out. There are several parameters used in the modeling process, namely size, which is the size or dimension of the vector to be generated, then min_count, which is the minimum frequency of a word, and window, which is the range between the words context and the position of the current word. A training process then carries out the model that has been created to train the word2vec model so that it can convert or index words into vectors. After the training process is complete, a test can be done to see the vector of a word that is close to other words according to word2vec.

2.1.6 Tokenize Text

The tokenize text process is needed to divide the words contained in a sentence into a token, making it easier to carry out the classification process. The process of dividing words from a sentence into a token is carried out using one of the classes in the Keras library, namely Tokenizer. The use of the utility tokenizer class aims to vector the existing corpus of text into a list of binary numbers. This makes each number/integer map a value in the dictionary, which then encodes the entire corpus, with the keywords in the dictionary being the terms in the vocabulary itself. Then after tokenizing the text, the padding process is also carried out which aims to equalize the size of each sentence that will be classified, where the missing tweet-length will be filled using the number 0.

2.1.7 Label Encoding

Label encoding is a process carried out to change data from a category in which in this study the label becomes numeric so that the category or label can be processed and understood by machines. The label encoding process is carried out by using one of the functions contained in the Sklearn library, namely the label encoder.

2.1.8 Defining Embedding Layer

The embedding layer can be defined as the first hidden layer which converts the word token to a specific embedding size. There are several parameters contained in the embedding layer, namely input_dim, namely the size of the vocabulary in the text (vocab_size), output_dim, which is the size of the vector space where words will be embedded (embed_dim), and input_length, which is the length of the input sequence (max_len). Then also included the weight (embedding_matrix) to initialize the existing weights on this layer.

2.1.9 Classification Model Making

Figure 5 is the process of making a sentiment analysis classification model using the LSTM method. In this classification model, several layers are used, including the embedding layer, then the LSTM layer with 128 neurons, then the dense layer, which functions as the fully connected layer, namely the layer that maps the LSTM output layer to the desired output size, besides that the dense layer is also given sigmoid activation which functions to change all output values is in a value between 0 and 1 then the output of sigmoid activation which is the final output of the network.
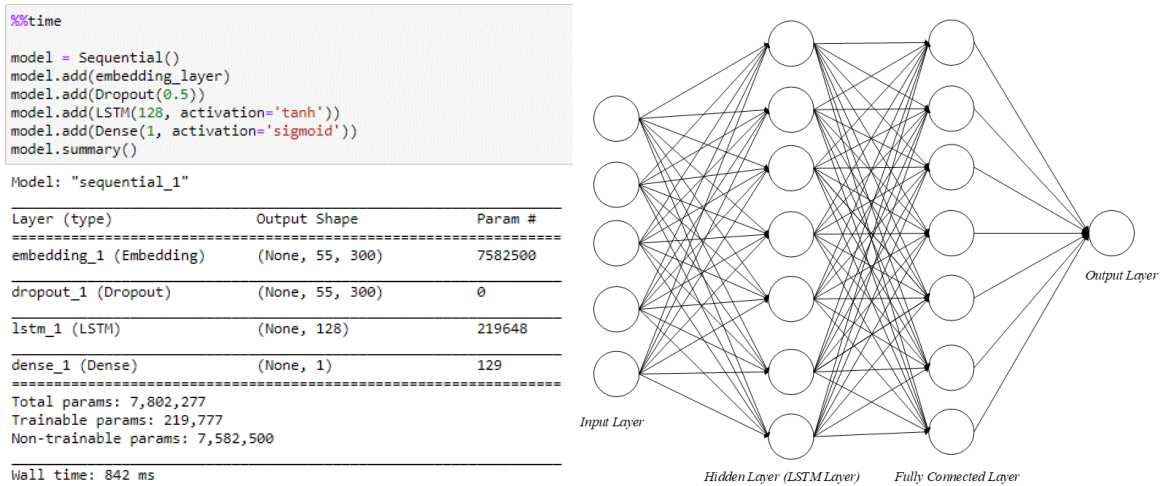
Sentiment Analysis on Social Distancing and Physical Distancing Using the RNN Algorithm
(Fikri Aldi Nugraha[1], Nisa Hanum Haranir[2], Roni Habibi[3], Rd. Nuraini Siti Fatonah[4])

200

```
%%time

model = Sequential()
model.add(embedding_layer)
model.add(Dropout(0.5))
model.add(LSTM(128, activation='tanh'))
model.add(Dense(1, activation='sigmoid'))
model.summary()

Model: "sequential_1"

Layer (type)                 Output Shape              Param #
=================================================================
embedding_1 (Embedding)      (None, 55, 300)           7582500

dropout_1 (Dropout)          (None, 55, 300)           0

lstm_1 (LSTM)                (None, 128)               219648

dense_1 (Dense)              (None, 1)                 129
=================================================================
Total params: 7,802,277
Trainable params: 219,777
Non-trainable params: 7,582,500
_____
Wall time: 842 ms
```

Figure 5. RNN / LSTM Classification Model and Architecture

2.1.10 Compiling Model

Proses *compiling model* bertujuan untuk mengkonfigurasi model yang telah dibuat, yaitu untuk menentukan *loss function, optimizer* serta matriks untuk prediksi. Adapun loss function yang digunakan yaitu binary cross entropy dan optimizer RMSprop dengan learning rate 0.0001.

2.1.11 Training Model

The compiling model process aims to configure the model that has been created, namely, to determine the loss function, optimizer, and matrix for prediction. The loss function used is the binary cross-entropy and the RMSprop optimizer with a learning rate of 0.0001.

2.1.12 Model Evaluation

The Model Evaluation Process aims to measure the performance of the model that has gone through the training process. Model evaluation is done by using a confusion matrix, where confusion matrix is a method that is usually used to calculate accuracy in data mining and supervised machine learning concepts. From a classification point of view, there are 4 (four) categories that can be used to compare labels from each class on the classification results, namely True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN) such as shown in Figure 7 and Figure 8 [17].
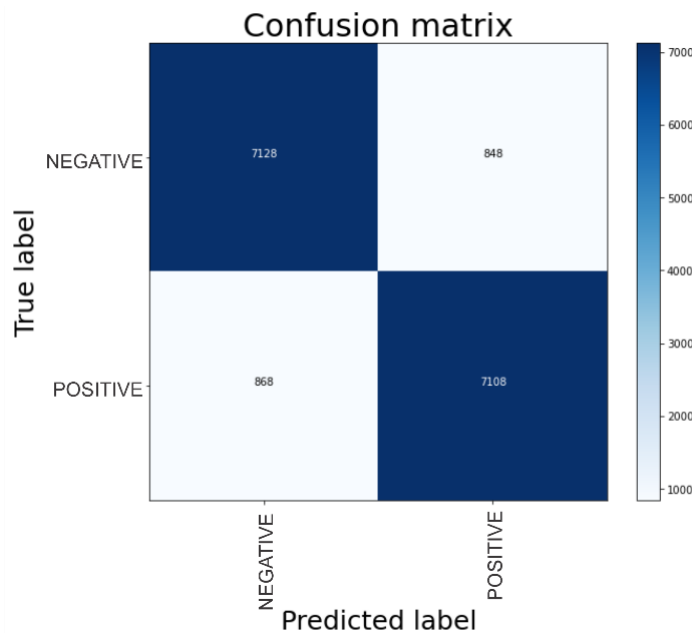
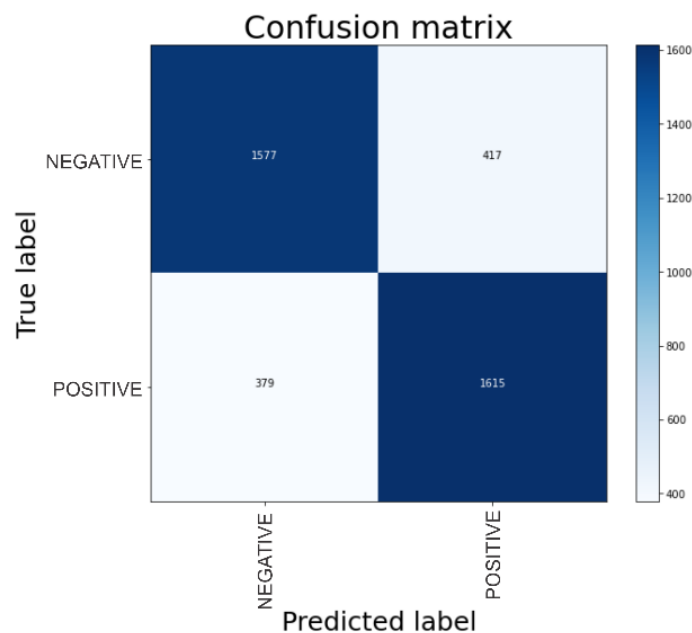

Figure 6. Confusion Matrix Training Data

Figure 7. Confusion Matrix Test Data

Based on the value obtained from the confusion matrix, four parameters are measured as a performance evaluation material, which can be as follows [18].

1. Accuracy is the percentage of data records classified correctly by the algorithm. To measure accuracy, it can be done using Equation 1.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

2. Precision is a measure of the accuracy of the classification results obtained from dividing the number of records classified correctly as positive by the number of records classified as positive. To calculate the value of precision, it can be done using Equation 2.

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

3. Recall is an evaluation carried out to measure the completeness of the classification results. The recall value is obtained by dividing the number of records labeled positive by the number of records correctly classified as positive. To measure the value of recall, it can be done using Equation 3.

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

4. F-Measure (F1 Score) is an evaluation carried out to measure the average alignment between precision and recall. The value of the f-measure is needed to optimize the system and to see which is the greater effect of recall and precision on the final result. To measure the value of the f-measure can be done using Equation 4.

$$F - Measure = \frac{2 * Precision * Recall}{Precision + Recall} \tag{4}$$

2.1.13 Save Model

This stage is a process for storing a classification model where the purpose of storing is to don't need to run back training when you return to predicting new data but just classify with an existing model as shown in Figure 6.

```
model.save("model/model_3_(89-80)/model.h5")
w2v_model.save("model/model_1/model.w2v")
pickle.dump(tokenizer, open("model/model_3_(89-80)/tokenizer.pkl", "wb"), protocol=0)
pickle.dump(encoder, open("model/model_3_(89-80)/encoder.pkl", "wb"), protocol=0)
```

Figure 8. Save Model

Sentiment Analysis on Social Distancing and Physical Distancing Using the RNN Algorithm
(Fikri Aldi Nugraha[1], Nisa Hanum Haranir[2], Roni Habibi[3], Rd. Nuraini Siti Fatonah[4])

202

3.   RESULTS AND DISCUSSION

Based on the research that has been done, the results obtained from a sentiment analysis where the public response to social distancing and physical distancing has more positive sentiments than negative sentiments. This can be seen in Figure 9, where the results of the predictions that have been made on the tweet data with the keyword physical distancing, amounting to 1815 data, the results are 1069 tweets or 58.9% with positive sentiment labels and 746 tweets or 41.1% with negative sentiment labels. Meanwhile, based on the predictions that have been made on the tweet data with the keyword social distancing, amounting to 1483, the results obtained are 1065 tweets or 71.8% with a positive sentiment label and 418 tweets or 28.2% with a negative sentiment label as in Figure 10.
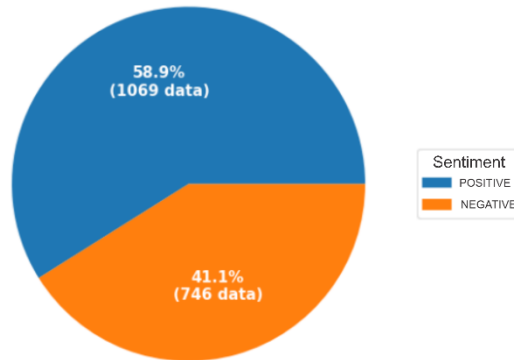
Figure 9. Visualization of the Results of Classification Dataset Test Keywords Physical Distancing
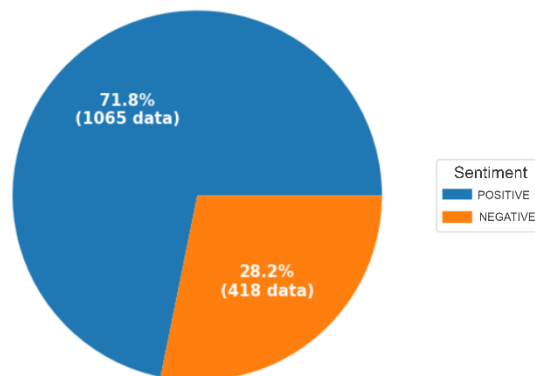

Figure 10. Visualization of the Results of Classification Dataset Test Keywords Social Distancing

To see how accurate the model can describe the classification correctly, accuracy is needed. Accuracy is the ratio of true predictions (positive and negative) to the overall data. In other words, accuracy is the level of closeness of the predicted value to the actual or actual value. Measuring the level of accuracy of sentiment analysis using the RNN algorithm and evaluating the modeling is done using confusion matrix where the results obtained for the training dataset are 89% accuracy, 89% recall, 89% precision, and 89% F1 Score. Meanwhile, for the test dataset, an accuracy of 80% was obtained, a recall of 79%, a precision of 81%, and an F1 score of 80%.

4.   CONCLUSION

The conclusion that can be drawn from the research is Sentiment Analysis on Social Distancing (Social Distancing and Physical Distancing) on Twitter Social Media Using the Recurrent Neural Network (RNN) Algorithm, namely the RNN algorithm can be used to classify tweet data from social media Twitter against social distancing and physical distancing. Meanwhile, the feature extraction process from the text is carried out using the word embedding concept called word2vec by using 300 dimensions of the word2vec embedding to become an input feature in LSTM. Then, the results of the evaluation of the modelling using confusion matrix are obtained, that is, for the training dataset, an accuracy of 89% is obtained, a recall of 89%, a precision of

89%, and an F1 score of 89%, while for the test dataset an accuracy of 80% is obtained, a recall of 79%, precision is 81%, and F1 Score is 80%.

The results of the classification that have been carried out on the tweet data with the keyword physical distancing, which amounted to 1815, the results obtained were 1069 tweets or 58.9% with a positive sentiment label and 746 tweets or 41.1% with a negative sentiment label. While the results of the classification of the tweet data with the keyword social distancing, amounting to 1483, the results obtained were 1065 tweets or 71.8% with a positive sentiment label and 418 tweets or 28.2% with a negative sentiment label.

There are several suggestions that can be made for the development of further research, namely that it is necessary to increase the number of training data to get the results from the evaluation of the model to be more satisfying and to carry out the process of turning back the parameters of the model to increase the accuracy of the sentiment classification.

5. REFERENCES

[1] K. K. R. Indoenesia, "Peraturan Menteri Kesehatan Republik Indonesia No. 82 Tahun 2014 Tentang Penanggulangan Penyakit Menular," *Ber. Negara Republik Indones.*, vol. Nomor 65, no. 879, pp. 2004–2006, 2014.

[2] M. Parwanto, "Virus Corona (2019-nCoV) penyebab COVID-19," *J. Biomedika dan Kesehat.*, vol. 13, no. 8, pp. 751–752, 2006.

[3] "Indonesia Corona Virus Report," 2020. [Online]. Available: https://www.worldometers.info/coronavirus/country/indonesia/.

[4] "Undang-Undang Republik Indonesia Nomor 6 Tahun 2018 Tentang Kekarantinaan Kesehatan."

[5] "Indonesia Digital Report 2020," 2020. [Online]. Available: https://datareportal.com/reports/digital-2020-indonesia.

[6] R. Arief and K. Imanuel, "Analisis Sentimen Topik Viral Desa Penari Pada Media Sosial Twitter Dengan Metode Lexicon Based," *J. Ilm. Matrik*, vol. 21, no. 3, pp. 242–250, 2019.

[7] A. H. Alamoodi *et al.*, "Sentiment analysis and its applications in fighting COVID-19 and infectious diseases: A systematic review," *Expert Syst. Appl.*, no. October, p. 114155, 2020.

[8] 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.

[9] K. Kowsari, D. E. Brown, M. Heidarysafa, K. Jafari Meimandi, M. S. Gerber, and L. E. Barnes, "HDLTex: Hierarchical Deep Learning for Text Classification," *Proc. - 16th IEEE Int. Conf. Mach. Learn. Appl. ICMLA 2017*, vol. 2017-Decem, pp. 364–371, 2017.

[10] M. A. Nurrohmat and A. SN, "Sentiment Analysis of Novel Review Using Long Short-Term Memory Method," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 13, no. 3, p. 209, 2019.

[11] A. Patel and A. K. Tiwari, "Sentiment Analysis by using Recurrent Neural Network," in *2nd International Conference On Advanced Computing And Software Engineering (ICACSE-2019)*, 2019, pp. 1–4.

[12] C. Chiu *et al.*, "State-Of-The-Art Speech Recognition With Sequence-To-Sequence Models," pp. 4774–4778, 2018.

[13] A. Kumar and R. Rastogi nee Khemchandani, "Attentional recurrent neural networks for sentence classification," *Advances in Intelligent Systems and Computing*, vol. 757. pp. 549–559, 2019.

[14] W. C. F. Mariel, S. Mariyah, and S. Pramana, "Sentiment analysis: A comparison of deep learning neural network algorithm with SVM and naïve Bayes for Indonesian text," *J. Phys. Conf. Ser.*, vol. 971, no. 1, 2018.

[15] N. H. Harani, C. Prianto, and F. A. Nugraha, "Segmentasi Pelanggan Produk Digital Service Indihome Menggunakan Algoritma K-Means Berbasis Python," *J. Manaj. Inform.*, vol. 10, no. 2, pp. 133–146, 2020.

[16] D. Sarkar, R. Bali, and T. Sharma, *Practical Machine Learning with Python: A Problem-Solver's Guide to Building Real-World Intelligent Systems*. Apress, 2017.

[17] N. H. Harani, C. Prianto, and M. Hasanah, "Deteksi Objek Dan Pengenalan Karakter Plat Nomor Kendaraan Indonesia Menggunakan Metode Convolutional Neural Network (CNN) Berbasis Python," *J. Tek. Inform.*, vol. 11, no. 3, pp. 47–53, 2019.

[18] A. Tripathy, "Classification of Sentiment of Reviews using Supervised Machine Learning Techniques," *Int. J. Rough Sets Data Anal. ·*, vol. 4, no. 1, pp. 56–74, 2016.

Sentiment Analysis on Social Distancing and Physical Distancing Using the RNN Algorithm
(Fikri Aldi Nugraha[1], Nisa Hanum Haranir[2], Roni Habibi[3], Rd. Nuraini Siti Fatonah[4])

204