Contents lists available at www.journal.unipdu.ac.id

Register

Journal Page is available to www.journal.unipdu.ac.id/index.php/register

Research article

Identifying Degree-of-Concern on COVID-19 topics with text classification of Twitters

Novrindah Alvi Hasanah [a], Nanik Suciati [b], Diana Purwitasari [c]

[a,b,c] Department of Informatics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

email: [a] alvinovrindah@gmail.com, [b] nanik@if.its.ac.id, [c] diana@if.its.ac.id

ARTICLE INFO

Please cite this article in IEEE style as:
N. A. Hasanah, N. Suciati and D. Purwitasari, "Identifying Degree-of-Concern on COVID-19 topics with text classification of Twitters," *Register: Jurnal Ilmiah Teknologi Sistem Informasi*, vol. 7, no. 1, pp. 50-62, 2021.

ABSTRACT

The COVID-19 pandemic has various impacts on changing people's behavior socially and individually. This study identifies the Degree-of-Concern topic of COVID-19 through citizen conversations on Twitter. It aims to help related parties make policies for developing appropriate emergency response strategies in dealing with changes in people's behavior due to the pandemic. The object of research is 12,000 data from verified Twitter accounts in Surabaya. The varied nature of Twitter needs to be classified to address specific COVID-19 topics. The first stage of classification is to separate Twitter data into COVID-19 and non-COVID-19. The second stage is to classify the COVID-19 data into seven classes: warnings and suggestions, notification of information, donations, emotional support, seeking help, criticism, and hoaxes. Classification is carried out using a combination of word embedding (Word2Vec and fastText) and deep learning methods (CNN, RNN, and LSTM). The trial was carried out with three scenarios with different numbers of train data for each scenario. The classification results show the highest accuracy is 97.3% and 99.4% for the first and second stage classification obtained from the combination of fastText and LSTM. The results show that the classification of the COVID-19 topic can be used to identify Degree-of-Concern properly. The results of the Degree-of-Concern identification based on the classification can be used as a basis for related parties in making policies to formulate appropriate emergency response strategies in dealing with changes in public behavior due to a pandemic.

1. Introduction

Since the World Health Organization (WHO) declared the COVID-19 outbreak a global pandemic, Indonesia, especially Surabaya, has become one of the affected areas. It has significant impacts on changing people's behavior socially and individually [1]. One example of the importance of knowing Degree-of-Concern is the panic buying incident when the COVID-19 outbreak first entered Indonesia, which resulted in the scarcity of certain items in some areas, for example, masks, hand sanitizers, and some basic food ingredients (CNN, 2020). Panic buying also occurred in China when hoaxes spread stating that iodized salt could help ward off Japan's nuclear radiation in March 2011 [2]. Because of these problems, it is necessary to identify Degree-of-Concern. Information on important topics that become Degree-of-Concern is needed by related parties to make appropriate emergency response strategies [3, 4].

Identification of Degree-of-Concern can be done through social media because people tend to take advantage of social media, for example, Twitter [3, 5, 6], to obtain information [7]. Public looking

for information on Twitter through a verified account that is trusted to provide the needed information. According to statistical data on social media users in Indonesia [8], it shows that approximately 56% of the total population of Indonesians use social media. The varied of Twitter text make it difficult to find out Degree-of-Concern on specific issues. So, a classification process related to the desired topic is necessary. A study to find public perceptions by analyzing Twitter content sentiment regarding H1N1 was done in 2009 [9]. Another research was carried out to determine the degree-of-control for several diseases such as listeria, measles, swine flu, and TB based on sentiment analysis results [10]. In this two studies, identification of Degree-of-Concern only refers to sentiment analysis, whereas not all tweets that show Degree-of-Concern have tweet sentiment, for example, shared information about the increase or decrease in COVID-19 cases which do not have an element of sentiment but remain a Degree-of-Concern. This study does not focus only on sentiment analysis, but this study also has identified all tweets related to COVID-19. The number of topics related to COVID-19 currently being discussed underlies such a research [11] classifying situational information related to COVID-19 from Weibo data. The study divides COVID-19 into seven types of classes, namely warnings and suggestions, notification of information, donations (materials and services), emotional support, seeking help, criticism, and hoax.

Text classification with deep learning is considered better than machine learning [12] on the same object. A previous study [13] showed that the results of deep learning accuracy were better than machine learning in the classification of health tweets. The superior performance of deep learning than machine learning underlies this research using deep learning as a classification method. The deep learning methods used are Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), and Long Short-Term Memory (LSTM) in accordance with several previous studies that used these methods to classify text, such as a research [14] using CNN as a deep learning method for sentence classification. In another study [12, 15] used CNN for text classification and showed good results. Most CNNs have a smaller number of parameters so that CNNs can be trained with a small amount of word data [11]. Another advantage of CNN for text classification is that it is a convolutional filter that automatically learns the suitable features of a given task. However, in study [16] the CNN performance results did not exceed the RNN because RNN was rated better for text classification than CNN. RNN can maintain word order sequence [16]. Based on this, this study will use the RNN method as a comparison of CNN performance. Previous studies used RNN [17] for static machine translation. RNN has a weakness, namely exploding gradient. Thus, the standard RNN was developed into LSTM. An earlier study [18] used the LSTM method to conduct sentiment analysis. The study stated that LSTM is more effective than CNN and RNN. In another study, LSTM and word embeddings showed higher accuracy than LSTM alone [19]. Based on several previous researches, this study uses the deep learning methods such as CNN, RNN, and LSTM for the text classification method.

In the previous studies that have been mentioned, some of these studies show that when the deep learning and word embedding methods are combined, it yield better accuracy results [15, 19]. A research [19] shows that Word2Vec's performance for medical text classification is better than other word embeddings used in the study when it's applied with deep learning. The study showed that this combination outperformed all combinations performed in experiments research. Another research [20] shows fastText is better than some word embeddings such as GloVe, ngram2vec, dict2vec, and CBOW in word equation performance.

Thus, this research combines deep learning and word embedding as feature extraction to achieve maximum accuracy. This study uses Word2Vec and fastText as a word embedding based on several previous studies. Word2Vec is a type of word embedding that has a good performance. Meanwhile, fastText is a toolkit developed by the Facebook Research Team [21] and an extension of the Word2Vec model, which presents each word as an n-gram bag of characters [22].

The COVID-19 pandemic significantly impacts people in many fields, both socially and individually. In order to solve this problem, a Degree-of-Concern identification is required. Identification can be made through Twitter social media. Identifying Degree-of-Concern on the topic of COVID-19 via Twitter text can be problematic. Twitter, which has a variety of topics, complicates the identification process. So it requires Twitter text classification for the COVID-19 topic. In fact, the process of identifying Degree-of-Concern on certain topics is critical to assist related parties in

developing appropriate emergency response strategies to deal with changes in community behavior socially and individually.

This paper proposes a method to identify the Degree-of-Concern on COVID-19 topic through the Twitter text object based on two stages of the classification process. Twitter text classification on the discussion of COVID-19 aims to get a specific topic of COVID-19 because Twitter data has a variety of topics. Classification is carried out with six pairs of combinations between two words embeddings methods, namely Word2Vec and fastText, with three deep learning methods: CNN, RNN, and LSTM. The proposed method eases the identification process more specifically by calculating the Degree-of-Concern because the classification process is carried out to omit non-COVID-19 data first. Our research can also show which combinations of deep learning and word embedding methods are good for Twitter's text classification process. The identification process proposed in this study, starting from the classification stage to measuring the Degree of Concern, can be used by related parties to develop appropriate emergency response strategies.

This paper is structured as follows: An introduction is presented in Section 1. Section 2 presents the experimental stages carried out in identifying Degree-of-Concern based on the results of Twitter text classification. Section 3 describes the research results, which consists of the results of the classification and the Degree-of-Concern identification. Section 4 describes the analysis of the research results. Section 5 provides a summary of the results of our experiment as well as a brief excerpt of our future work.

2. Research Methodology

In order to achieve the desired results, this research will go through several stages, as shown in Fig. 1. The following subsections present the explanation of the stages in detail.
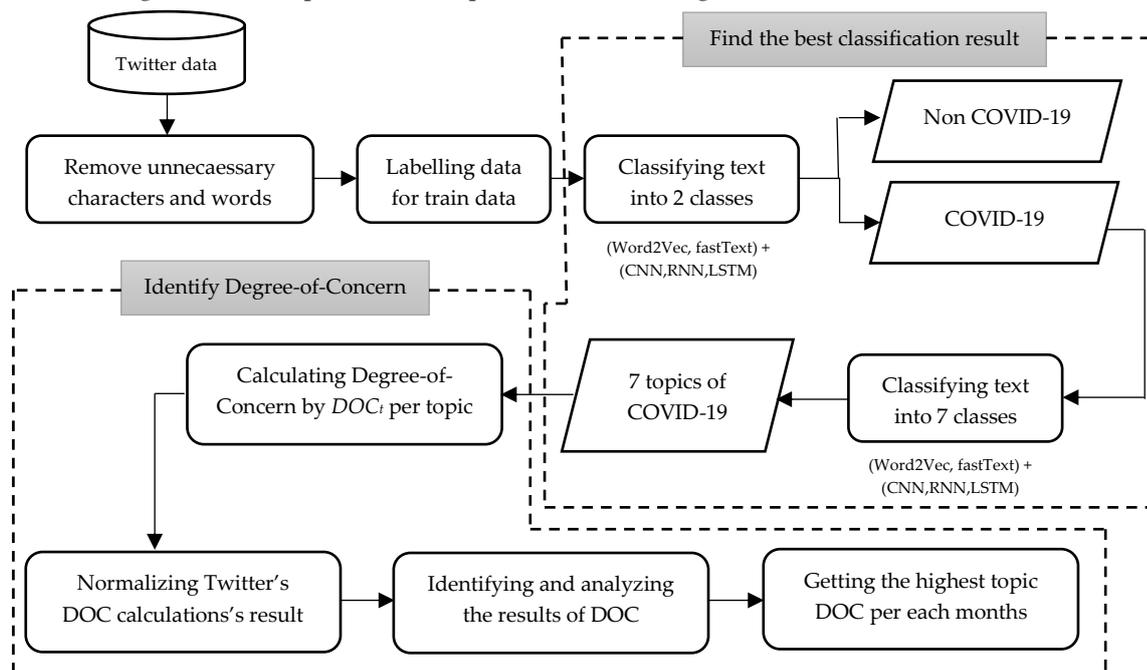

Fig. 1. Proposed method to identify Degree-of-Concern COVID-19's topic

2.1. Data preparation

The data used in this study is tweet data in Indonesian from the timeline of verified account in Surabaya, Suara Surabaya's Twitter account (@e100ss), in the vulnerable time between February and October 2020. The total tweet data that was collected in the study was more than 12,000 data. This data consists of COVID-19 and non-COVID-19 data. Tweets are retrieved by crawler technique using the Twitter API and saved in .csv format.

The Twitter dataset that was successfully obtained was processed to remove some unnecessary elements in the classification process. Noise in data such as punctuation marks, numbers, and uppercase letters are replaced with lowercase letters so that the data obtained is clean. The stemming process in the pre-processing data uses the Sastrawi library specifically developed for Indonesian. The clean data

from the pre-processing results are divided into test data and train data. The training data is manually labeled, which is used in the classification process. The data is divided into two labels for non COVID-19 and COVID-19. The second labeling is the labeling of COVID-19 data into seven classes according to reference [11].

Table 1. Definition of topic, sample tweet of COVID-19 in Suara Surabaya's Twitter account, and keyword per topic

No.	Topic*	Definition*	Twitter text	Keyword
1.	Warnings and suggestions	Preventive measures, such as frequent hand washing, wearing masks, physical distancing or recommendations to respond to an outbreak of crisis	Government Will Take Social Distancing Policy to Suppress Covid-19	masks, protocols, distancing, sanctions, discipline
2.	Notification of information	Announcement of increased or decreased outbreaks by the relevant department, such as how many cases occurred, virus characteristics, material reserves, etc.	75 Covid-19 Patients Healed, Positive Cases 1,414 and 122 Died	positive, cases, recovered, died, cluster
3.	Donations	Donations or wishes to provide assistance in efforts to prevent and control an outbreak	Chairman of IDI: Help Us Health Officers Assisted by PPE Donations	donations, groceries, social assistance, apd, blt
4.	Emotional support	Show support to others such as the medical team, public welfare organizations, Covid-19 sufferers and others.	Risma give support to Medical Workers at Dr. Soewandhie Hospital	enthusiasm, support, solidarity, optimism, pessimism
5.	Seeking assistance	a. medical institutions, individuals, etc. to seek support needs etc. b. seeking emotional support such as seeking comfort, or to express depression, etc.	Medical personnel need a lot of PPE and medical equipment to handle COVID-19 patients. Come together with IDI Surabaya Branch to care and ease the burden on those in the front line	aid, raise, care, donors, funds
6.	Criticism	To question the actions of the local government, central government, the Red Cross or other related parties, or parts of the community to mislead others.	Agus asked the government and related parties to be more serious in trying to stop the spread of the COVID-19 virus or at least reduce it don't be fail.	fail, evidence, crisis, bad, criticism
7.	Hoax	Responding to rumors	Hoax Corona Test Rates, Director of Unair Hospital: We Continue to Follow Government Policies	hoax, claims, clarifications, information, minimum

On the Degree-of-Concern identification process, to demonstrate the correct identification of the Twitter's Degree-of-Concern text, we built a ground truth. A ground truth was created by collecting articles obtained from Google News which had been set to get articles containing COVID-19 and the location specified in Surabaya. To determine the articles related to each topic, we found the most frequently used words in Suara Surabaya account tweets using the document frequency. We assign five keywords to each topic to find articles with related headlines. The collected articles are used to calculate the Degree-of-Concern using Formula (1). This Degree-of-Concern then become the ground truth used as the comparison with the Degree-of-Concern Twitter text. Table 1 shows the definition per topic, the

sample of the tweet data and keywords per topic. Samples of tweets that we gathered actually in Indonesian, but in Table 1 we have translated into English.

2.2. Classification text

The classification is carried out in two stages. The first classification is done by classifying the data into two classes, namely COVID-19 and non-COVID-19. The second classification is carried out to classify the COVID-19 data obtained from the first classification into seven classes, namely warnings and suggestions, notification of information, donations (materials and services), emotional support, seeking help, criticism, and rumors. Text classification experiments were carried out in three different scenarios to see the effect of the number of train data for text classification. The scenarios are shown in Table 2. The Train ID column in Table 2 is the code for each train data per scenario.

Table 2. Test scenario

Scenario	Train ID	Train Data	Test Data
1	S1.1	April-May	June
	S1.2	May-June	July
	S1.3	June-July	August
	S1.4	July-August	September
	S1.5	August-September	October
2	S2.1	March-April-May	June
	S2.2	April-May-June	July
	S2.3	May-June-July	August
	S2.4	June-July-August	September
	S2.5	July-August-Sept	October
3	S3.1	Feb-March-April-May	June
	S3.2	March-April-May-June	July
	S3.3	April-May-June-July	August
	S3.4	May-June-July-August	September
	S3.5	June-July-August-Sept	October

In each scenario, we perform two classifications using a combination pair of word embedding and deep learning. This method is a combination of two word embedding methods and three deep learning methods. Word embedding used in this experiment is Word2Vec and fastText. Word2Vec is a word embedding in a package gensim that is computationally efficient and works to process and learn words from raw text [23]. Word2Vec has two training algorithms, namely Continuous Bag-Of-Words and Skip-gram [24]. fastText is a toolkit developed by the Facebook Research Team [21]. fastText is an extension of the Word2Vec model representing each word as an n-gram bag of characters [22].

The deep learning methods used are CNN, RNN, and LSTM. CNN is a development of the Multi-Layer Perceptron (MLP). CNN has a workflow similar to MLP [25]. RNN is one architecture that is often used in dealing with NLP problems because of the repetitive structure that is suitable for processing text [26]. The basic idea of RNN is to create a network topology capable of representing sequential data or time series. LSTM addresses the vanishing gradient or burst problems experienced by vanilla RNNs by introducing a memory cell to remember random time interval values, and three gates (input gate, output gate, forget gate) to regulate the flow of incoming information and get out of the cell. Each method will be tested in each scenario. The combination of these methods is Word2Vec + CNN, Word2Vec + RNN, Word2Vec + LSTM, fastText + CNN, fastText + RNN, and fastText + LSTM. The results of the classification method with the best accuracy percentage will be used in find out Degree-of-Concern. In this experiment, a program was made using the Python 3 programming language and the gensim package.

2.3. Degree-of-Concern measurement

$$DOC_t = \frac{\sum tweet\ of\ topic_n}{\sum tweet} \tag{1}$$

The last step in the research stage is to get the Degree-of-Concern value after carrying out the calcification stage twice. For the identify Degree-of-Concern, we use Eq. 1, where DOC_t= Degree-of-Concern from a specific time (in this study we use measurement every month from June until October), $\sum tweet\ of\ topic_n$ = the number tweets of topic$_n$, and $\sum tweet$ = number of tweets on all topics.

All data obtained from the classification results of each scenario are used to calculate Degree-of-Concern using Eq. 1. In the experiment, each scenario has a different range of values on each topic. The difference in the range of values causes the malfunction of the scores on certain topics because they have smaller values than other topics. Thus, data transformation is needed to equalize the range of values on each topic with a certain scale in order to get better results. Normalization in this study can also determine which scenario is best used in calculating the Degree-of-Concern for each topic. Data transformation with normalization can be done in several ways, namely min-max normalization, z-score normalization, decimal scaling, sigmoid, and softmax [27]. This study uses min-max normalization. This study's experiments determine a new range of values taken from the range of values in the ground truth. The new minimum and maximum ground truth values, a subtraction of 10 is determined from each intensity of the original ground truth's Degree-of-Concern value. The Twitter's Degree-of-Concern value for each scenario was normalized using the normalization formula to determine which scenario was best used in Twitter's Degree-of-Concern calculation. The scenario formula used is shown in Eq. 2 below.

$$I_N = (I - Min) \frac{newMax - newMin}{Max - Min} + newMin \tag{2}$$

where I_N = new Twitter's Degree-of-Concern values per topic, I = old Twtter's Degree-of-Concern values per topic, Min = the original minimum Twitter's Degree-of-Concern value per scenario, Max = the original maximum Twitter's Degree-of-Concern value per scenario, $newMin$ = the new GroundTruth's Degree-of-Concern minimum value, and $newMax$ = the new GroundTruth' Degree-of-Concern maximum value.

3. Results

3.1. Classification text

The first classification is the classification of data into two classes, namely COVID-19 and non-COVID-19. The results of the first stage classification obtain good accuracy because the data is only classified into two classes. This stage's results are shown in Table 3, with the highest accuracy of the first stage classification is 97.3%. The two classes' classification results show that all experiments using fastText and LSTM outperform all the methods used in the experiments. Thus, all the classification results that using fastText and LSTM is used as material for the second stage of classification.

Table 3. The first stage classification into two classes of Suara Surabaya's Twitter Text

TEST ID	Word embedding with Word2Vec		Word embedding with fastText	
	Classifier	Accuracy (%)	Classifier	Accuracy (%)
S1.1	LSTM	92.1	LSTM	**95.3**
S1.2	RNN	91.5	RNN	**94.9**
S1.3	LSTM	91.0	LSTM	**94.8**
S1.4	LSTM	91.2	LSTM	**94.8**
S1.5	RNN	90.6	RNN	**94.7**
S2.1	RNN	92.8	RNN	**96.0**
S2.2	LSTM	92.9	LSTM	**95.9**
S2.3	LSTM	92.6	LSTM	**95.7**
S2.4	LSTM	92.4	LSTM	**95.6**
S2.5	LSTM	92.1	LSTM	**95.1**
S3.1	LSTM	96.5	LSTM	**97.3**
S3.2	LSTM	95.1	LSTM	**96.9**
S3.3	LSTM	94.5	LSTM	**96.1**
S3.4	LSTM	93.2	LSTM	**96.1**
S3.5	LSTM	92.7	LSTM	**95.9**

All COVID-19 tweets are going through to the second stage classification, which classifies the data into seven classes. The second stage classification accuracy is lower than the first stage because of the limited Twitter data. To solve this problem, we added an augmentation. The augmentation we use is synonym replacements, random deletions, and random exchanges. The second classification results

with the best accuracy are obtained from a combination of deep learning using fastText for all scenarios. The second stage classification result, in which the data is categorized into seven classes, is shown in Table 4. The best-accuracy classification results for the combination of deep learning and Word2Vec as well as deep learning and fastText show that LSTM's deep learning method outperformed CNN and RNN in all experiments, respectively. However, in scenario 1 and scenario 2 there are two experiments which show that RNN is better than LSTM and CNN. Nevertheless, the accuracy results of the RNN are still lower than all experiments in each scenario. The result data shows the difference in accuracy between before and after augmentation. The experiments show that the utilization of augmentation increases accuracy. The trial scenarios that were carried out show that the more training data used, the greater and better accuracy the classification results. It is proved by the accuracy in scenario 3, which shows better accuracy than other scenarios even without augmentation.

Table 4. The second stage classification result of Suara Surabaya's Twitter Text

Test ID	Word embedding with Word2Vec			Word embedding with fastText		
	Classifier	Non-Augment (%)	Augment (%)	Classifier	Non-Augment (%)	Augment (%)
S1.1	LSTM	69.6	98.5	LSTM	71.4	**98.8**
S1.2	LSTM	66.3	98.2	LSTM	71.2	**98.8**
S1.3	LSTM	66.9	97.3	LSTM	66.7	**98.3**
S1.4	LSTM	66.3	96.9	LSTM	69.4	**97.9**
S1.5	RNN	65.4	95.6	RNN	65.9	**97.7**
S2.1	RNN	65.9	97.2	RNN	70.7	**98.6**
S2.2	LSTM	70.2	98.7	LSTM	72.8	**98.8**
S2.3	LSTM	68.2	97.6	LSTM	70.9	**98.7**
S2.4	LSTM	67.4	97.5	LSTM	70.9	**98.6**
S2.5	LSTM	66.6	97.3	LSTM	69.6	**98.9**
S3.1	LSTM	75.6	98.8	LSTM	77.6	**99.1**
S3.2	LSTM	80.8	99.3	LSTM	80.8	**99.4**
S3.3	LSTM	78.2	98.9	LSTM	77.0	**99.3**
S3.4	LSTM	70.3	98.6	LSTM	71.6	**98.8**
S3.5	LSTM	68.2	97.6	LSTM	70.9	**98.7**

All experiments of the second stage classification results show that classification using fastText as word embedding combined with deep learning for classification has better accuracy than Word2Vec. The highest accuracy of the second stage classification is 99.4%. All of the second stage classification results from the combination of fastText and deep learning with augmentation are used to identifying the Degree-of-Concern.

3.2. Twitter's Degree-of-Concern measurement

Table 5. Degree-of-Concern measurement scenario 1 before and after normalization

Topic	June		July		August		September		October	
	Before	After	Before	After	Before	After	Before	After	Before	After
1	0.447	0.238	0.439	0.236	0.416	0.232	0.672	0.280	0.761	0.327
2	0.497	0.278	0.501	0.279	0.419	0.264	0.254	0.233	0.180	0.188
3	0.023	0.161	0.041	0.205	0.041	0.205	0.054	0.236	0.017	0.147
4	0.013	0.163	0.003	0.010	0.017	0.225	0.018	0.137	0.011	0.133
5	0.005	0.062	0.003	0.059	0.011	0.071	0.036	0.109	0.011	0.071
6	0.005	0.036	0.006	0.040	0.011	0.056	0.018	0.079	0.011	0.056
7	0.007	0.042	0.003	0.037	0.005	0.040	0.018	0.056	0.005	0.040

The results of the Degree-of-Concern measurement are shown in Table 5, Table 6, and Table 7. All scenarios are normalized according to Eq. 2 to generalize the ranges on each topic so that it can be determined which scenario is best used for calculation. Degree-of-Concern. The result of the Degree-of-Concern calculation with normalization is used to obtain the best scenario used in the calculation. In Table 5, the "before" column is the Degree-of-Concern value before normalization, while the "after" column is the result of normalizing the Degree-of-Concern value. The "after" column is compared with the ground truth's Degree-of-Concern results to show the correctness of Twitter's Degree-of-Concern approximation.

Table 6. Degree-of-Concern measurement scenario 2 before and after normalization

Topic	June		July		August		September		October	
	Before	After	Before	After	Before	After	Before	After	Before	After
1	0.390	0.205	0.390	0.232	0.414	0.214	0. 593	0.280	0.344	0.188
2	0.560	0.316	0.560	0.293	0.506	0.302	0. 237	0.233	0.601	0.327
3	0.021	0.171	0.021	0.187	0.036	0.205	0. 050	0.236	0.010	0.147
4	0.005	0.025	0.005	0.010	0.018	0.125	0.033	0.240	0.016	0.109
5	0.013	0.072	0.013	0.074	0.006	0.059	0.033	0.109	0.016	0.078
6	0.002	0.036	0.002	0.074	0.012	0.050	0.033	0.079	0.005	0.040
7	0.005	0.037	0.005	0.037	0.006	0.040	0.016	0.056	0.005	0.040

Table 7. Degree-of-Concern measurement scenario 3 before and after normalization

Topic	June		July		August		September		October	
	Before	After	Before	After	Before	After	Before	After	Before	After
1	0.379	0.196	0.348	0.188	0.678	0.277	0.689	0.280	0.436	0.212
2	0.565	0.316	0.629	0.327	0.275	0.268	0.068	0.233	0.490	0.304
3	0.031	0.175	0.059	0.236	0.018	0.147	0.051	0.219	0.024	0.160
4	0.012	0148	0.003	0.010	0.013	0.163	0.017	0.255	0.018	0.240
5	0.005	0.091	0.011	0.071	0.003	0.059	0.034	0.109	0.012	0.074
6	0.002	0.036	0.003	0.039	0.003	0.039	0.017	0.079	0.006	0.048
7	0.002	0.037	0.003	0.038	0.003	0.038	0.037	0.056	0.012	0.050

4. Discussion

In the Degree-of-Concern identification process, this research focuses not only on sentiment analysis because all tweets related to COVID-19 are considered a Degree-of-Concern. Basically, not all tweets that are Degree-of-Concern contain elements of sentiment, for example, on the topic of warnings and suggestions, notification of information, donations, emotional assistance, and help-seeking which do not necessarily have an element of sentiment but remain a Degree-of-Concern so this research processes all related tweets with COVID-19.

In the first stage, the classification process resulted in better accuracy than the later stage because the processed data is only classified into two classes. It is caused by the balanced distribution of data. The results of the first stage classification show that all the accuracy value has promising results. Whereas in the second stage classification, the first experiment shows poor accuracy results. It is because the limited amount of data must be classified into seven classes. Thus, the distribution of data becomes unbalanced, which causes the accuracy of the results to decrease due to the limited data train for learning seven different topics. To fix this, we added augmentation. We use synonym replacements, random deletions, and random exchanges. Because the data we use is in Indonesian, we get augmentation of synonym replacement with a list of synonyms from the "Kamus Besar Bahasa Indonesia (KBBI)". After adding augmentation, the accuracy results achieved the best results when compared to the results before adding augmentation. This study applies three different scenarios. The accuracy results show that scenario 3 is the scenario with the best accuracy. Scenario 3 is the scenario with the highest number of training data. This result indicates that the more training data used to test the data, the higher the accuracy of the classification results. Even without augmentation, scenario 3 was also superior to the others.

The accuracy results show that the combination of the deep learning method with fastText shows better accuracy than the combination using Word2Vec. It is because fastText is the latest model from Word2Vec and has the advantage of being able to improve word representation for morphologically rich languages containing verbs and nouns in various forms **[28]**. The fastText word representation approach is different from the word representation in Word2Vec. The representation of the word fastText is that assuming a word is composed of n-gram characters, the length n can be changed. The benefit of this approach is to find vector representations for words that cannot be directly found in a word dictionary **[21]**. fastText can improve performance on syntactic tasks significantly but not much on semantics **[20]**.

According to all of the classification experiments' results in the first stage, 12 out of 15 experiments show LSTM is better than all the available method combinations. The other three were

58
N. A. Hasanah et al.
regist. j. ilm. teknol. sist. inf.

ISSN 2502-3357 (online) | ISSN 2503-0477 (print)
7 (1) January 2021 50-62

RNN. For the second stage classification, 13 out of 15 experiments and the second stage classification show that LSTM has the best accuracy among all available combination pairs. The other two experiments were RNN. It is because LSTM can solve the problem of explosion or gradient removal in a standard RNN. LSTM can better study the long-run dependencies of the order of higher-level representations than other models [29]. According to [19, 30, 31], the LSTM method is superior when used for time-series models. So LSTM is suitable for solving Twitter text problems with the time-series model.

According to the theory of RNN, RNN is one of the architectures that is often used in dealing with NLP problems because of the repetitive structure suitable for processing text [26]. This is indicated by good sequence data modeling so that the RNN can predict the next word in a sentence. It is made possible because of its ability to capture good contextual information in sequences, or if it is concluded that the RNN has a memory containing the results of recorded information produced previously [32]. Some researchers argue that the RNN is better for text classification than the CNN, because it maintains the order of the word order [16]. In this experiment, CNN has less accuracy than other methods because CNN has better performance when applied to image objects [25], unlike RNN and LSTM, which are suitable for text.

The results of the classification second stage of each scenario with the best accuracy are calculated as the value of Degree-of-Concern. To demonstrate the correctness of our experiment, we made a ground truth measurement with the same formula but using different data. The data we use to create ground truth is the number of articles from Google News we previously set using a special COVID-19 news filter and the location is Surabaya. Each topic was determined by five keywords, which were obtained by calculating document frequency. Each topic contains the amount of news obtained through manual collection and analysis calculated using Eq. 1. The results of the calculation of Twitter's Degree-of-Concern are shown in Table 8.

Table 8. Ground truth of Degree-of-Concern measurement

Topic	June	July	August	September	October
1	0.271	0.267	0.198	0.328	0.250
2	0.302	0.337	0.309	0.244	0.270
3	0.157	0.157	0.192	0.188	0.246
4	0.076	0.055	0.057	0.020	0.250
5	0.082	0.069	0.119	0.076	0.096
6	0.046	0.049	0.077	0.089	0.067
7	0.065	0.066	0.047	0.054	0.058

In the experiment, each scenario has a different range of values on each topic. The difference in the range of values causes the malfunction of the scores on certain topics because they have smaller values than other topics. Thus, data transformation is needed to equalize the range of values on each topic with a certain scale to get better results. Normalization in this study can also determine which scenario is best used in calculating the Degree-of-Concern for each topic. This study uses min-max normalization by Eq. 2. In the experiment, the Degree-of-Concern GroundTruth results were compared with the normalized Degree-of-Concern Twitter Text results. After being analyzed, the best scenario used to calculate the Degree-of-Concern text for Twitter is scenario 3 because in that scenario the difference between the Degree-of-Concern values between Twitter text and the Degree-of-Concern ground truth is minimum. The comparison of the results of the Degree-of-Concern Text Twitter normalized results from each scenario with the results of the Degree-of-Concern ground truth is illustrated in Fig. 2.

In June, July, August 2020, Twitter's Degree-of-Concern score showed the same result compared to the Degree-of-Concern result of the ground truth on topic 2, namely information notification. For example, information about the increasing number of COVID-19 cases that have occurred has a high Degree-of-Concern. In September the Twitter's Degree-of-Concern score showed the same result compared to the ground truth's Degree-of-Concern result, namely topic 1, the information notification, showing the highest result compared to other topics. According to an article obtained from Google News, the analysis states that in September, the daily addition of COVID-19 cases in Surabaya continued to decline. There are many articles regarding warnings and suggestions from the city government to the

public to adhere to health protocols so that cases tend not to increase. Thus, public social media activities focus on warning topics and suggestions as an effort to support the public for government decisions. In October, the Degree-of-Concern measurement value showed that the topic with the highest score is topic 1, namely warnings and suggestions, for scenario 1. Meanwhile, for scenario 2, scenario 3, and ground truth show the same results, namely topic 2, notification of information.
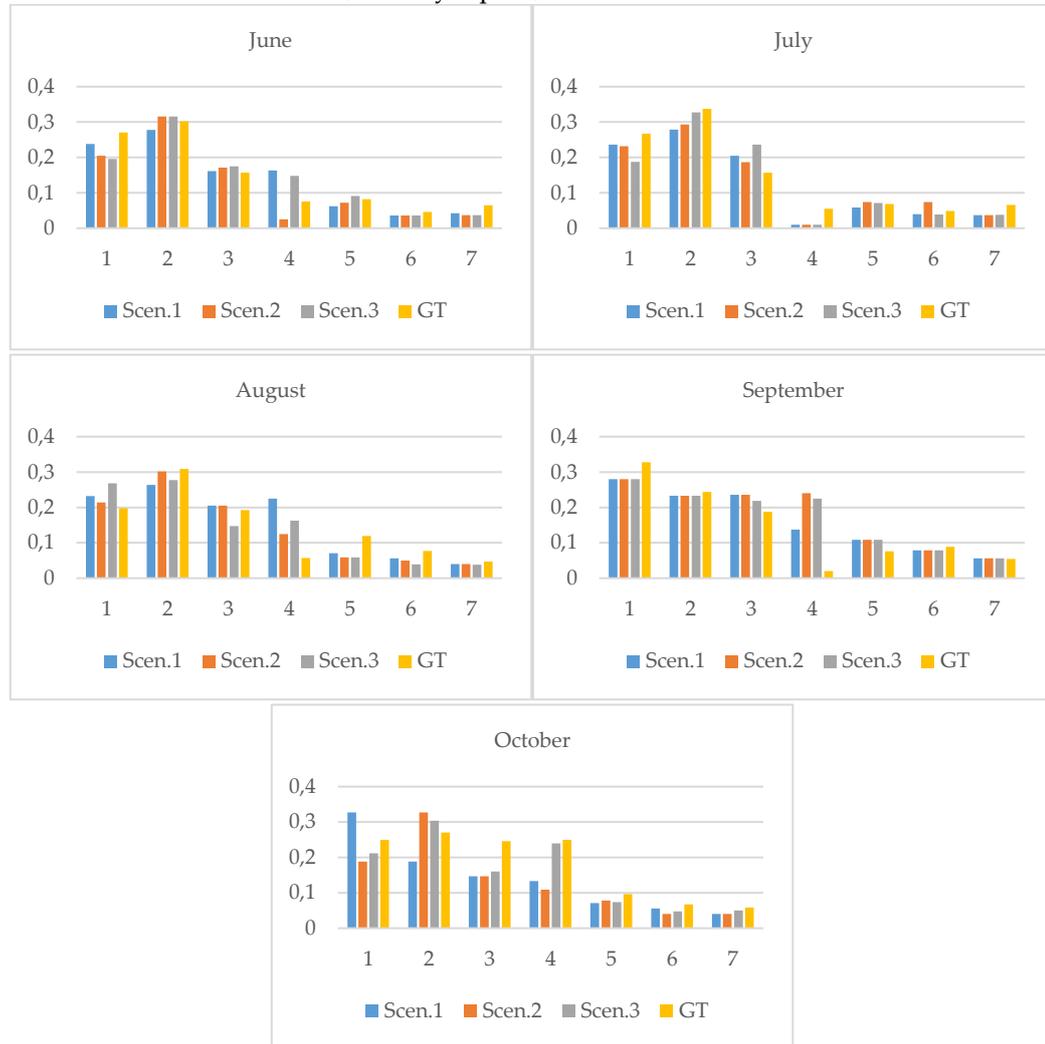

Fig. 2. Comparison of Twitter's Degree-of-Concern calculations per scenario with ground truth

All Degree-of-Concern measurements based on Twitter text classification results show that the topic with the highest score in each month has the same result as the ground truth result. The difference in resulted value because each class's amount of data in each scenario is not the same. Likewise, the amount of data in the ground truth's Degree-of-Concern with the classification results, because the number of collected Twitter texts and Google News articles is different. However, the Twitter's Degree-of-Concern value is not significantly different from the ground truth's Degree-of-Concern because normalization is applied so that the transformed data becomes more balanced.

Fig. 2 shows that the sequence of Twitter's Degree-of-Concern values for the first, second, and third topics is the same as the result from ground truth. Meanwhile, the fourth to last order has a slight difference. The cause is the data in ground truth is more balanced than the data in Twitter text since most people express their heart through tweets that only focus on a specific topic. So that there is an imbalance of data on each classification result topic.

From all the results of the Degree-of-Concern calculation, topic 2, which is information notification, obtains a relatively high score. This is indicated by the high value of the Degree-of-Concern on that topic in June, July and August for all scenarios, as well as in September. In October, the scenario 1 show that topic 1, which is warning and suggestions is highest In the conducted experiments, it appears that the topic of information is growing and declining in COVID-19 cases. It is indicated in the results of the Degree-of-Concern identification in June, July, August, and October.

5. Conclusion

The identification of the Degree-of-Concern COVID-19's topic was carried out based on the results of Twitter text classification through two classification stages using a combination of word embedding method (Word2Vec and fastText) and deep learning methods (CNN, RNN, and LSTM). The first stage separates COVID-19 and non-COVID-19 with the highest accuracy of 97.3%. The second stage classifies the COVID-19 data into seven class topics with the highest accuracy of 99.4%. Each classification stage shows that the highest accuracy is obtained in the scenario with the highest number of training data. This proves that the more number of training data, the better the classification results. The method pairs with the highest accuracy are fastText and LSTM, because fastText can split words that do not appear during training models into n-grams to become vectors. LSTM also works well for Twitter text objects with time series models.

Meanwhile, the results of the Degree-of-Concern identification in June, July, August and October indicated that the topic of information notification was the topic with the highest score. But, the topic with the highest Degree-of-Concern score in September was the topic of warnings and suggestions. The results show that the classification of the COVID-19 topic can be used to identify Degree-of-Concern correctly. For future work, we will use data that has a broader reach, not only in Surabaya. This proposed method can also be further developed for Degree-of-Concern on various issues to assist the related parties in making an appropriate emergency response strategy on these issues.

Declaration of Competing Interest
We declare that we have no conflict of interests.

References

[1] X. Zhu, S. Wu, D. Miao and Y. Li, "Changes in Emotion of The Chinese Public in Regard to the Sars Period," *Social Behavior and Personality: an international journal,* vol. 36, no. 4, pp. 447-454, 2008.

[2] X. Ji, S. A. Chun, Z. Wei and J. Geller, "Twitter sentiment classification for measuring public health concerns," *Soc. Netw. Anal. Min.,* vol. 5, no. 13, 2015.

[3] K. Lee, D. Palsetia, R. Narayanan, M. M. A. Patwary, A. Agrawal and A. Choudhary, "Twitter Trending Topic Classification," in *2011 IEEE 11th International Conference on Data Mining Workshops*, Vancouver, BC, Canada, 2011.

[4] E. M. Glowacki, J. B. Glowacki and G. B. Wilcox, "A text-mining analysis of the public's reactions to the opioid crisis," *Substance Abuse,* vol. 39, no. 2, pp. 129-133, 2018.

[5] T. Rathod and M. Barot, "Trend Analysis on Twitter for Predicting Public Opinion on Ongoing Events," *International Journal of Computer Applications,* vol. 180, no. 26, pp. 13-17, 2018.

[6] L. Yan and A. J. Pedraza-Martinez, "Social Media for Disaster Management: Operational Value of the Social Conversation," *Production and Operations Management ,* vol. 28, no. 10, pp. 2514-2532, 2019.

[7] S. Vieweg, A. L. Hughes, K. Starbird and L. Palen, "Microblogging During Two Natural Hazards Events: What Twitter May Contribute to Situational Awareness," in *CHI '10: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Atlanta, GA, USA, 2010.

[8] S. Kemp, "Digital 2019: Indonesia," DataReportal, 2019.

[9] C. Chew and G. Eysenbach, "Pandemics in the Age of Twitter: Content Analysis of Tweets during the 2009 H1N1 Outbreak," *PLoS ONE,* vol. 5, no. 11, p. e14118, 2010.

[10] X. Ji, S. A. Chun and J. Geller, "Monitoring Public Health Concerns Using Twitter Sentiment Classifications," in *2013 IEEE International Conference on Healthcare Informatics*, Philadelphia, PA, USA, 2013.

[11] L. Li, Q. Zhang, X. Wang, J. Zhang, T. Wang, T.-L. Gao, W. Duan, K. K.-f. Tsoi and F.-Y. Wan, "Characterizing the Propagation of Situational Information in Social Media During COVID-19 Epidemic: A Case Study on Weibo," *IEEE Transactions on Computational Social Systems,* vol. 7, no. 2, pp. 556-562, 2020.

[12] S. Boukil, M. Biniz, F. E. Adnani, L. Cherrat and A. E. E. Moutaouakkil, "Arabic Text Classification Using Deep Learning Technics," *International Journal of Grid and Distributed Computing,* vol. 11, no. 9, pp. 103-114, 2018.

[13] R. A. Calix, R. Gupta, M. Gupta and K. Jiang, "Deep gramulator: Improving precision in the classification of personal health-experience tweets with deep learning," in *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Kansas City, MO, 2017.

[14] Y. Kim, "Convolutional Neural Networks for Sentence Classification," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar, 2014.

[15] M. Hughes, I. Li, S. Kotoulas and T. Suzumura, "Medical Text Classification Using Convolutional Neural Networks," *Stud Health Technol Inform,* vol. 235, pp. 246-250, 2017.

[16] A. Severyn and A. Moschitti, "UNITN: Training Deep Convolutional Neural Network for Twitter Sentiment Classification," in *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, Denver, Colorado, 2015.

[17] K. Cho, B. v. Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk and Y. Bengio, "Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar, 2014.

[18] A. Tholusuri, M. Anumala, B. Malapolu and G. J. Lakshmi, "Sentiment Analysis using LSTM," *International Journal of Engineering and Advanced Technology (IJEAT),* vol. 8, no. 6S3, pp. 1338-1340, 2019.

[19] A. Rao and N. Spasojevic, "Actionable and Political Text Classification using Word Embeddings and LSTM," arXiv, 2016.

[20] B. Wang, A. Wang, F. Chen, Y. Wang and C.-C. J. Kuo, "Evaluating word embedding models: Methods and experimental results," *APSIPA Transactions on Signal and Information Processing,* vol. 8, no. E19, 2019.

[21] B. Kuyumcu, C. Aksakalli and S. Delil, "An automated new approach in fast text classification (fastText): A case study for Turkish text classification without pre-processing," in *NLPIR 2019: Proceedings of the 2019 3rd International Conference on Natural Language Processing and Information Retrieval*, Tokushima, Japan, 2019.

[22] F. K. Khattak, S. Jeblee, C. Pou-Prom, M. Abdalla, C. Meaney and F. Rudzicz, "A survey of word embeddings for clinical text," *Journal of Biomedical Informatics: X,* vol. 4, 2019.

[23] A. Mandelbaum and A. Shalev, "Word Embeddings and Their Use In Sentence Classification Tasks," arXiv, 2016.

[24] T. Mikolov, K. Chen, G. Corrado and J. Dean, "Efficient Estimation of Word Representations in Vector Space," arXiv, 2013.

[25] I. W. S. E. Putra, "Klasifikasi Citra Menggunakan Convolutional Neural Network (CNN) Pada Caltech 101," Institut Teknologi Sepuluh Nopember, Surabaya, 2016.

[26] P. Liu, X. Qiu and X. Huang, "Recurrent neural network for text classification with multi-task learning," in *IJCAI'16: Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, 2016.

[27] D. A. Nasution, H. H. Khotimah and N. Chamidah, "Perbandingan Normalisasi Data untuk Klasifikasi Wine Menggunakan Algoritma k-NN," *CESS (Journal of Computer Engineering, System and Science),* vol. 4, no. 1, pp. 78-82, 2019.

[28] S. Dabiri and K. Heaslip, "Developing a Twitter-based traffic event detection model using deep learning architectures," *Expert Systems with Applications,* vol. 118, pp. 425-439, 2019.

[29] C. Zhou, C. Sun, Z. Liu and F. C. Lau, "A C-LSTM Neural Network for Text Classification," arXiv, 2015.

[30] Z. Zhang, Q. He, J. Gao and M. Ni, "A deep learning approach for detecting traffic accidents from social media data," *Transportation Research Part C: Emerging Technologies,* vol. 86, pp. 580-596, 2018.

[31] G. Liu, X. Xu, B. Deng, S. Chen and L. Li, "A hybrid method for bilingual text sentiment classification based on deep learning," in *2016 17th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, Shanghai, China, 2016.

[32] C. Du and L. Huang, "Text Classification Research with Attention-based Recurrent Neural Networks," *International Journal of Computers Communications & Control,* vol. 13, no. 1, pp. 50-61, 2018.