





# Aspect-based Sentiment and Correlation-based Emotion Detection on Tweets for Understanding Public Opinion of Covid-19

Salsabila <sup>1)</sup> , Salsabila Mazya Permataning Tyas <sup>2)</sup> , Yasinta Romadhona <sup>3)</sup> ,  
Diana Purwitasari <sup>4)\*</sup> 

<sup>1,2,3,4)</sup> *Department of Informatics, Institut Teknologi Sepuluh Nopember, Indonesia  
Jl. Teknik Kimia, Keputih, Sukolilo, Surabaya*

<sup>1)</sup>salsabilasabil151199@gmail.com, <sup>2)</sup>salsa25mazya@gmail.com, <sup>3)</sup>yasinta.rmd@gmail.com, <sup>4)</sup>diana@if.its.ac.id

---

## Abstract

**Background:** During the Covid-19 period, the government made policies dealing with it. Policies issued by the government invited public opinion as a form of public reaction to these policies. The easiest way to find out the public's response is through Twitter's social media. However, Twitter data have limitations. There is a mix between facts and personal opinions. It is necessary to distinguish between these. Opinions expressed by the public can be both positive and negative, so correlation is needed to link opinions and their emotions.

**Objective:** This study discusses sentiment and emotion detection to understand public opinion accurately. Sentiment and emotion are analyzed using Pearson correlation to determine the correlation.

**Methods:** The datasets were about public opinion of Covid-19 retrieved from Twitter. The data were annotated into sentiment and emotion using Pearson correlation. After the annotation process, the data were preprocessed. Afterward, single model classification was carried out using machine learning methods (Support Vector Machine, Random Forest, Naïve Bayes) and deep learning method (Bidirectional Encoder Representation from Transformers). The classification process was focused on accuracy and F1-score evaluation.

**Results:** There were three scenarios for determining sentiment and emotion, namely the factor of aspect-based and correlation-based, without those factors, and aspect-based sentiment only. The scenario using the two aforementioned factors obtained an accuracy value of 97%, while an accuracy of 96% was acquired without them.

**Conclusion:** The use of aspect and correlation with Pearson correlation has helped better understand public opinion regarding sentiment and emotion more accurately.

**Keywords:** Aspect-based sentiment, Deep learning, Emotion detection, Machine learning, Pearson correlation, Public opinion.

**Article history:** Received 1 January 2023, First decision 26 January 2023, Accepted 7 March 2023, Available online 28 April 2023

---

## I. INTRODUCTION

As the adherents of a decentralized system, some countries have various institutions and also implement multiple regional policies. As a result, policies from one region to another may differ in response to Covid-19[1]. The number of guidelines issued by the government invites public opinion as a form of public reaction. We can take public opinion from social media, one of which is Twitter. Twitter is a good source of information because it publishes news and information from various sources, be it facts or personal opinions [2]. Opinions expressed by the public on Twitter can be in the form of positive, negative, or neutral responses or sentiments. In addition, sentiment usually contains certain aspects that are related to something that is happening. One technique used to identify public sentiment is sentiment analysis. Furthermore, sentiment analysis of public opinion on Twitter have been conducted by various studies [3]–[5]. The studies [3] and [4] analyze using aspects, while [6] analyzes by detecting emotions.

Aspect-based is a new challenge of sentiment analysis, revealing various aspects of an entity [7]. Aspect-Based Sentiment Analysis (ABSA) involves identifying specific aspects mentioned in a given text and determining the sentiment associated with each of those aspects. ABSA provides a more precise way to approach sentiment analysis by focusing directly on sentiment rather than language structure [8] while an aspect is connected to an entity, its basic

---

\* Corresponding author

concept goes beyond evaluation and includes thoughts, viewpoints, ways of thinking, perspectives, underlying themes, and social influences on an event.

Meanwhile, emotion detection is a type of sentiment analysis related to extracting and analyzing emotions [9]. Emotions describe intense feelings directed at something or someone in response to internal or external events that have special meaning for the individual. In psychology, complex states of feeling lead to a change in thoughts, actions, behavior, and personality, referred to as emotions [10]. We can use this emotion-related information to study how people react to different situations and circumstances. Several previous studies use one of these two things, though analyzing sentiment from one aspect alone is only enough to consider the community's emotions. Because emotions can be a supporting factor in understanding public opinion, understanding related to public opinion can be more accurate. This study aims to analyze public sentiment and emotions during the Covid-19 pandemic when the government implemented policies to address the situation. Such policies elicited public opinion, which can be captured through Twitter's social media platform. However, Twitter data have their limitations, as they include a blend of factual information and personal views, requiring a clear differentiation between the two. The public's opinions expressed on Twitter may be either positive or negative, necessitating the establishment of a correlation between their opinions and associated emotions. Therefore, our main contribution is taking advantage of these two things using the Pearson correlation to measure the linear correlation of the two datasets. Pearson correlation was used due to its ability to examine the correlation between different aspects and emotions derived from public opinion. By utilizing Pearson correlation, we can determine the correlation between aspects and public opinion by computing the correlation value between polarity and subjectivity. Additionally, to verify the emotions expressed, we can calculate the correlation value between the emotions.

After knowing the aspects and emotions of public opinion, a classification process uses several machine learning and deep learning methods. The machine learning methods used are Naive Bayes (NB), Random Forest (RF), and Support Vector Machine (SVM). At the same time, the method based on deep learning used is Bidirectional Encoder Representations from Transformer (BERT). The results of the implementation of these methods are compared to determine the most effective method for classification.

The subsequent explanation of this paper in Section 2 describes previous research related to our study. Section 3 contains the design stages of carrying out this experiment. Section 4 describes the results obtained from the experiments that must carry out. Section 5 describes the interpretation of the results to generate the main conclusions. Furthermore, section 6 ends with the experimental results and contains things that need to be corrected for further research.

## II. LITERATURE REVIEW

Studying Twitter posts during the early stages of the Covid-19 pandemic can provide insights into the emotions, beliefs, and thoughts of the general public. This information can be valuable for policymakers seeking to raise awareness about the virus [11]. Previous studies have employed sentiment analysis techniques that focused on either aspects or emotions. Aspect-based sentiment analysis involves identifying topics related to the object under review before classifying the sentiments expressed on each topic [5]. This technique leads to a more detailed and user-friendly clustering system since it allows for the identification of specific features that receive positive or negative sentiments. In a previous study [3], the researchers utilized aspect extraction for aspect-based sentiment analysis on a Twitter dataset. They employed TF-IDF and Word2Vec to identify the aspects that were being discussed and conducted emotional analysis to determine the crucial topics that elicited positive, negative, or neutral sentiments. TF-IDF was used to identify the most used words and corpus expressions, while the Word2Vec model identified the most relevant words for every 100 words. Based on the study's findings, the researchers selected 'politics,' 'health,' and 'media' as the top three frequently used words and analyzed them as aspects. If tweets that were not related to these three aspects were analyzed, a fourth category called 'other' was added. In another study [3], aspect-based sentiment analysis was employed to evaluate online customer reviews. The researchers examined their customers' attitudes toward particular aspects, such as 'food,' 'service,' 'ambience,' 'drinks,' and 'location,' They directly identified these aspects because they already knew which aspects they wanted to analyze. The two studies showcase distinct methods for identifying the aspects to be analyzed, one using frequently used words to identify aspects and the other directly selecting the aspects. In this study [4] the approach of identifying frequently discussed aspects was utilized because it suggests that these aspects are likely to be linked with significant issues and widely discussed. The study conducted aspect-based sentiment analysis to assess Covid-19-related data from two distinct datasets and identified three aspects for analysis.

The field of affective science has extensively researched the classification of emotions into distinct groups and categories. There are two main approaches to emotion classification: discrete and dimensional [12]. On the one hand, dimensional emotions usually categorize emotions based on one or more dimensions, such as valence, intensity, and

arousal. In contrast, discrete emotion theory divides emotions into eight distinct categories, including surprise, interest, joy, rage, fear, disgust, shame, and anguish [12]. It is believed that these emotions can be recognized across different cultures. These fundamental emotions are instinctive emotional reactions that are biologically predetermined, and their expression and recognition are the same for all people, regardless of their cultural or ethnic backgrounds. In terms of detecting emotions, [13] the system conducts an emotional analysis of each tweet shared by a Twitter user. The system's goal is to sort tweets into positive and negative categories and then categorize them according to the six fundamental emotions identified by Ekman: joy, sadness, anger, fear, disgust, and surprise. These emotions possess distinct characteristics that enable them to be expressed to varying degrees. As for [14], the analysis was conducted to comprehend the public's emotions and determine the extent and consequences of the distribution's implementation. In the case of Nigeria's Covid-19 palliative and aid distribution, they categorized emotions into five categories: anger, sadness, joy, fear, and disgust. Considering the two studies mentioned earlier, this study selected anger, sadness, happy, surprise, and fear as the emotions to focus on, while disgust was excluded because it was deemed unsuitable for aspects related to government policy.

Once the aspects and emotions to be used are identified, the next step is to extract them from the tweet data. In this case, a polarity score is chosen, and its correlation is analyzed using the Pearson correlation (PC) method. The PC method has been employed in various fields and is used to measure the degree of strength and direction of the linear relationship between random variables [14]. In previous research, Cheng Qian et al. [13] used the PC method in their financial data analysis. All emotions will be evaluated by calculating their correlation with the emotion score using the Pearson correlation, which ranges from -1 to +1. The calculation will also be applied to the aspects, with a value close to zero indicating no correlation, a value close to +1 indicating a direct relationship, and a value close to -1 indicating an inverse relationship.

After implementing PC, classification is then carried out using several methods based on Machine Learning (ML) and Deep Learning (DL). A study [15] used six machine learning algorithms, namely Multinomial Naïve Bayes (MNB), Support Vector Machine (SVM), Random Forest (RF), Logistics Regression (LR), K-Nearest Neighbor (KNN), and Decision Tree (DT). Another study [5] employed Support Vector Machine (SVM) which is a supervised learning algorithm. In this study, the ML used was SVM, NB and RF. SVM is a supervised learning approach that involves a range of learning algorithms used for classification and regression data analysis [16]. By using a set of training data with corresponding class labels, SVM creates a model that can predict the class of a new testing sample [17]. The algorithm maximizes the margin that separates classes while minimizing misclassification. This method involves dividing the sample into classes by identifying the optimal hyperplane (n-dimensional region) within the hyperspace (n-dimensional space). Meanwhile, Naïve Bayes is an algorithm based on polarity for classifying tweets. This algorithm uses a theoretical approach regarding data consistency and calculation classification [18]. NB was chosen because the implementation of this model is easy to implement, and this algorithm is known for its ability to work with small training datasets [18]. Specifically, the Naive Bayes algorithm operates under the assumption that individual words within a given text are independent of each other [19]. Next up is Random Forest, a type of ensemble model that is based on decision trees and can be utilized for both classification and regression tasks [20]. RF prediction is done by combining several decision trees. RF was chosen because this method has gained popularity in recent years. This is because the performance of this type of algorithm is extraordinary for classification tasks in several domains. Random Forest was formally introduced in 2001 by Leo Breiman and Adèle Cutler and is considered one of the machine learning techniques that fall under the category of supervised learning algorithms techniques [21]. The Random Forest algorithm combines 'bagging' and random subspaces to train multiple decision trees on slightly different subsets of data, leading to higher accuracy and reduced risk of overfitting. Another study [3] uses transformers-based DL, namely Bidirectional Encoder Representation from Transformers (BERT). BERT (Bidirectional Encoder Representations from Transformers) is regarded as one of the most effective language models in use today. It is built on top of a transformer encoder architecture, which is a type of sequence-to-sequence model that relies exclusively on attention mechanisms for both the encoder and decoder components [22]. BERT evaluates sentence components one by one to deal with sentences holistically and two-way (before and after). Thus, the previous word insertion technique obtained a different structure. It also creates predictions to fill in semantic deficiencies by masking between words. Thanks to these predictions, the model can successfully understand semantic integrity. Each token representation is based on the representation of all tokens, which is the main strength of the transformer model [16]. During the process of fine-tuning BERT on text classification tasks, the output representation of the generated tokens is utilized to feed the classification layer, which is typically a softmax layer [23]. BERT has proved to be a potent language model. This model has a transformer layer that allows parallelization, leading to more speed [24].

### III. METHODS

The stages of the proposed method are shown in Fig. 1. Starting from data preparation, annotation for sentiment and emotion, and aspect extraction, until the result of the classification with the classification report.

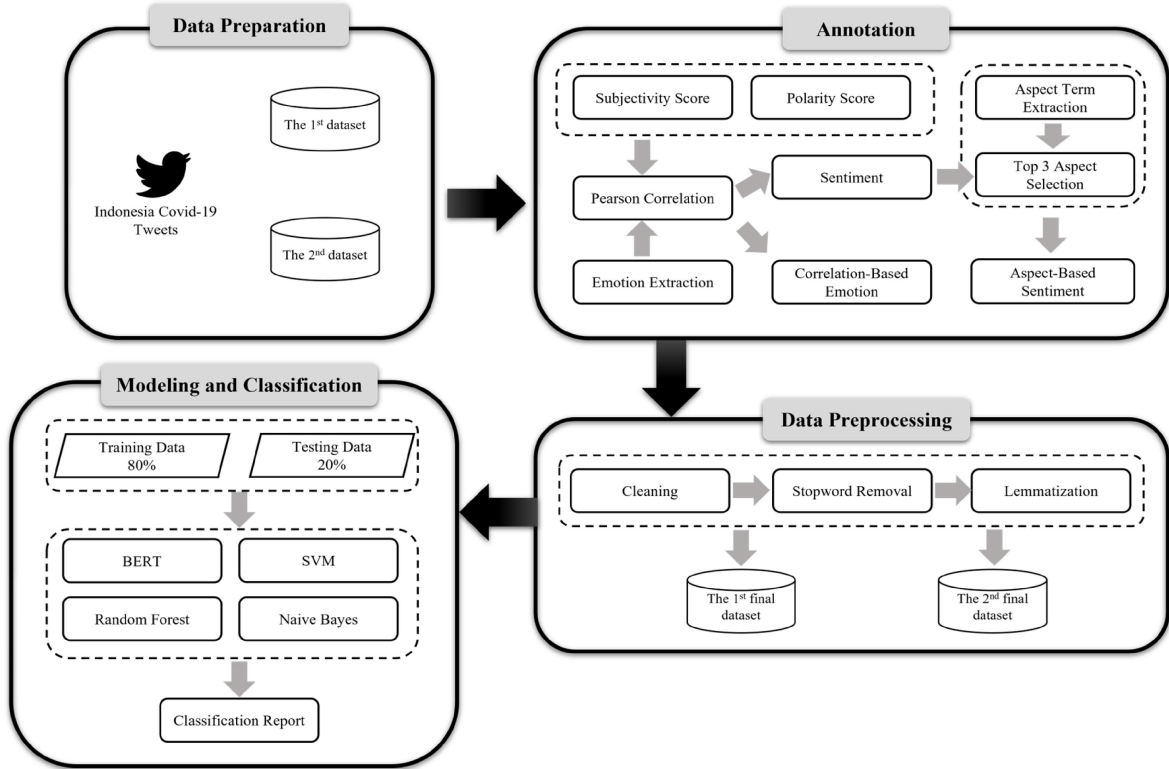


Fig. 1 The proposed method

#### A. Data Preparation

This research first used the Twitter dataset from the Indonesian language using Twitter API combined with *Twint library* to get big tweets fast. In this phase, tweets data were retrieved from the time of the start of Covid-19 pandemic, which was between March 14th 2020 and June 13th 2020 [25]. The data retrieval was focused on the area with the highest number of Covid-19 cases in Indonesia during the period of data collection, which has the potential as the source of investigation to find out people's reaction during crisis situations. Besides that, we also used another dataset taken from Kaggle with the same topic, created by Dionius Darryl Hermansyah [26]. The data distribution of each dataset is shown in Table 1.

TABLE 1  
 DATA DISTRIBUTION

Dataset	Emotion	Data Distribution	
		Sentiment	Aspect
1 <sup>st</sup> Dataset (Scrapped dataset beginning Covid-19)	Happy : 6332	Neutral: 6165 Negative: 188 Positive: 409	House Fasting Price
	Surprise : 100		
	Fear : 130		
	Angry : 21		
	Sad : 179		
2 <sup>nd</sup> Dataset (Dataset public opinion about Covid-19 from Kaggle)	Happy : 7691	Neutral: 14991 Negative: 6908 Positive: 15012	Government Homecoming Instruction
	Surprise : 5653		
	Fear : 15545		
	Angry : 1871		
	Sad : 6151		

B. Annotation

Data annotation process mainly describes about sentiment and emotion. The annotation process is implemented based on the correlation of several related variables to build sentiments and emotions respectively using the Pearson coefficient. First thing to do is the annotation of aspect-based sentiment. In determining the sentiment correlation, the extraction of polarity and subjectivity score was performed by using textblob Python package for text processing. Afterwards, the correlation between the two variables was checked using Pearson function. Based on this, the sentiments were generated into ‘negative,’ ‘positive,’ and ‘neutral.’ Meanwhile, the top three aspects were selected after the aspect term extraction process. Then, each sentiment was assigned into an aspect, which are used for the case of aspect-based sentiment analysis. Table 2 shows the example of feature representation on each tweet.

In addition to those three aspects, annotation is also made on emotion of the tweets. The emotion extraction was performed by using text2emotion Python package. Afterwards, the emotion correlation was built by calculating the two most contradictory emotions, namely happy and sad, using the method. If the calculation process obtained a good value, then the whole emotions could be implemented in the datasets. Otherwise, if the calculation didn’t bring in a good score, then it should be re-extracted. There are five categories of emotions namely ‘happy,’ ‘surprise,’ ‘fear,’ ‘sad’ and ‘angry.’ Happy emotion is detected by the words like ‘happy,’ ‘cool,’ ‘perfect,’ ‘fun,’ ‘decent,’ etc. The emotion of surprise is detected if there are words like ‘surprise,’ ‘announce,’ ‘result,’ ‘attraction,’ etc. The fear emotion is indicated by the words such as ‘afraid,’ ‘worried,’ etc. The sad emotion is identified if there are words like ‘sad,’ ‘disappointed,’ ‘sad,’ etc. Meanwhile, the angry emotion is noticed if words such as ‘angry,’ ‘annoying,’ ‘bad,’ ‘stupid,’ ‘damn,’ etc. are found.

TABLE 2  
 DATA EXAMPLE

Text	Aspect	Aspect Term	Sentiment	Emotion
sobat tanjungemas wabah covid 19 musim mudik himbauan pemerintah tidak mudik tidak piknik salah upaya memutus rantai penyebaran covid 19 sayangi keluarga ( <i>buddy tanjungemas plague covid 19 homecoming                      season appeals unmudic government no picnic                      wrong attempt to break the chain of the spread of                      covid 19 love the family</i> )	Homecoming	['penyebaran', 'mudik', 'rantai', 'keluarga', 'upaya', 'pemerintah', 'musim', 'himbauan', 'wabah', 'memutus']	Neutral	Surprise
		['deployment', 'homecoming', 'chain', 'family', 'effort', 'government', 'season', 'appeal', 'plague', 'disconnect']		
keblinger pemerintah tandanya pemerintah gagal total mengatasi pandemi covid indonesia kesimpulannya warga disuruh cari selamat menghindari covid ( <i>the government's confusion is a                      sign that the government has failed miserably in                      overcoming the Indonesian covid pandemic. In                      conclusion, residents are told to look for safety to                      avoid covid</i> )	Government	['disuruh', 'total', 'mengatasi', 'pemerintah', 'gagal', 'selamat', 'kesimpulan', 'pandemi', 'warga', 'tanda']	Negative	Sad
		['told', 'total', 'resolve', 'government', 'fail', 'safe', 'conclusion', 'pandemic', inhabitant', 'sign']		

C. Data Preprocessing

After annotation process, the next step is data preprocessing, as shown in Table 3. The purpose of data preprocessing is to reduce noise and irrelevant text in the classification process. There is a mechanism of the preprocessing text based on machine learning and it consists of a few steps. The first step is data cleaning, which is done by removing some unnecessary components such as emoji, hashtag, RT, Twitter username, site address, and non-alphabetical characters such as “.”, “!”, “@”, etc. Those characters have no significant effect. Then, the next step is case folding. This stage is to change the upper-case text into lowercase. Afterwards, stopword removal was done. This process is to remove the stopword that occurred commonly across the documents in the corpus. The last step is lemmatization, to convert the sentence into the basic form without affixes. This process only approved texts that are more than three words, and the maximum text length is 144.

TABLE 3  
 DATA PREPROCESSING

Preprocessing	Raw Data	Data Cleaning
Cleaning (Remove emoji, hashtag, number, RT, twitter username, site address, non-alphabetical characters)	Pemerintah Lampung bantuan mahasiswa dasarnya donasi yang masuk ke pemerintah provinsi Lampung tim gugus tugas Covid-19. #Covid19 (The Government of Lampung assisted students with the torrential donations that went to the Lampung provincial government for the Covid-19 task force team. #Covid19)	Pemerintah Lampung bantuan mahasiswa dasarnya donasi yang masuk ke pemerintah provinsi Lampung tim gugus tugas Covid Covid (The Government of Lampung assisted students with the torrential donations that went to the Lampung provincial government for the Covid Covid task force team)
Case Folding	Pemerintah Lampung bantuan mahasiswa dasarnya donasi yang masuk ke pemerintah provinsi Lampung tim gugus tugas Covid Covid (The Government of Lampung assisted students with the torrential of donations that went to the Lampung provincial government for the Covid Covid task force team)	pemerintah lampung bantuan mahasiswa dasarnya donasi yang masuk ke pemerintah provinsi lampung tim gugus tugas covid covid (the government of lampung assisted student with the torrential of donations that went to the lampung provincial government for the covid covid task force team)
Stopword Removal	pemerintah lampung bantuan mahasiswa dasarnya donasi yang masuk ke pemerintah provinsi lampung tim gugus tugas covid covid (the government of lampung assisted student with the torrential of donations that went to the lampung provincial government for the covid covid task force team)	pemerintah lampung bantuan mahasiswa dasarnya donasi masuk pemerintah provinsi lampung tim gugus tugas covid covid (government lampung assisted student torrential donations went lampung provincial government covid covid task force team)
Lemmatization	pemerintah lampung bantuan mahasiswa dasarnya donasi masuk pemerintah provinsi lampung tim gugus tugas covid covid (government lampung assisted student torrential donations went lampung provincial government covid covid task force team)	pemerintah lampung bantuan mahasiswa deras donasi masuk pemerintah provinsi lampung tim gugus tugas covid covid (government lampung assistance student torrent donation went lampung provincial government covid covid task force team)

#### D. Modeling and Classification

The classification was done using machine learning (ML) and deep learning (DL) methods. Before the classification process, feature extraction was implemented. As for the ML classification method, namely NB, SVM, and RF, the feature extraction is TF-IDF. Each ML method uses default parameters. On the other hand, both the DL classification method and the feature extraction is BERT. Afterward, the BERT tokenizer was used to generate the tokens. In determining the maximum tokenized sentence length, the API tweet was tagged with the ‘encode’ method of the BERT tokenizer and looked at the shortest sentence. Before the classification was carried out using the BERT method, the batch size was set to 32. In this study, the authors used the BERT-base-uncased pre-trained model where the BERT hidden size was 768, the classifier hidden size was 50, and 6 for the number of labels. This process runs on a GPU with two epochs. Furthermore, this study used default parameters for each classifier which is the best practice of previous research [27], [28].

Moreover, the accuracy and F1-score are conducted to measure the performance evaluation of each classifier on the two datasets. Accuracy is defined as the level of closeness between the predicted value and the actual value. Meanwhile, F1-score is comprehensive evaluation of the weighted average precision and recall, where precision is the exactness value between the ground truth information and the response decided by the system, and recall is the success rate to regain information by the system [29].

## IV. RESULTS

#### A. Data Preparation, Annotation, Data Preprocessing

Two different wordclouds for the first and second datasets are shown in Fig. 2a and Fig. 2b. There are several topics discussed in the first dataset, such as house, virus, people, fasting, price, family, sick, etc. Meanwhile, the words that often appear on the second dataset are country, government, economy, instruction, citizen, policy, etc. The aspects used for this study were obtained by analyzing the wordclouds. At this stage, there are top three aspects for each dataset. As for the first dataset, there are three aspects, i.e., ‘rumah’ (‘house’), ‘puasa’ (‘fasting’), and ‘harga’ (‘price’). Meanwhile, the second dataset got the top three aspects, i.e., ‘pemerintah’ (‘government’), ‘mudik’

(‘homecoming’), ‘*himbauan*’ (‘instruction’). Furthermore, the first dataset got a score of sentiment correlation by 0.024 and the second dataset obtained a score of 0.122 of which both are positive, so each of them has a strong correlation. As for emotion correlation, we obtained the scores of -0.011 for the first dataset and -0.117 for the second dataset, proving that the two variables of emotions, i.e., happy and sad, were not connected at all.



Fig. 2a Wordcloud of the (a) 1<sup>st</sup> Dataset and (b) 2<sup>nd</sup> Dataset

**B. Modeling and Classification**

There are several scenarios to calculate the performance metrics in determining non aspect-based sentiment and correlation-based emotion, aspect-based sentiment, and aspect-based sentiment and correlation-based emotion. BERT, Naïve Bayes, Random Forest, and SVM are the primary classification method of these experiments. Moreover, the experiments were conducted among all scenarios on two datasets, splitting the data into training and testing data with a ratio of 80%:20%. In the first case, as shown in Table 4, we conducted a single model classification on non-aspect-based and correlation-based scenarios. The classifiers were implemented for each aspect of the two datasets. Parameter specifications in Table 4 are also used for the experiments in Table 5 and Table 6. In the second case, as presented in Table 5, we performed a single model classification for aspect-based sentiment analysis. The first dataset’s aspects consist of house, fasting, and price. Meanwhile, the second dataset holds the aspects of government, homecoming, and instruction. In the last case, as in Table 6, we also experimented with aspect-based sentiment and correlation-based emotion. The significant difference of all is the comparison results between them.

TABLE 4  
 CLASSIFICATION RESULT OF NON ASPECT-BASED & CORRELATION-BASED

Dataset	Method	Parameter Specification	Sentiment		Emotion	
			F1-Score	Accuracy	F1-Score	Accuracy
1 <sup>st</sup> Dataset	BERT	Hidden size=768, batch size=32, epochs=2	59%	95%	45%	96%
	SVM	C=1.0, kernel='rbf', degree=3, gamma='scale', random_state=None	54%	94%	30%	94%
	Naïve Bayes	force_alpha='warn', fit_prior=True, class_prior=None	32%	91%	19%	94%
	Random Forest	n_estimators=100, criterion='gini', random_state=None	73%	96%	67%	96%
2 <sup>nd</sup> Dataset	BERT	Hidden size=768, batch size=32, epochs=2	77%	79%	59%	65%
	SVM	C=1.0, kernel='rbf', degree=3, gamma='scale', random_state=None	74%	78%	58%	67%
	Naïve Bayes	force_alpha='warn', fit_prior=True, class_prior=None	53%	64%	22%	47%
	Random Forest	n_estimators=100, criterion='gini', random_state=None	71%	76%	52%	61%

Table 4 shows different classification results from sentiment and emotion analysis on the first and second datasets, following non-aspect-based and correlation-based. The first dataset obtained the best results using the Random Forest method and consistently outperformed the second dataset in most performance metrics. Meanwhile, the second dataset acquired the best score using BERT and SVM methods, showing mixed and unstable results. Both sentiment and emotion analysis have a lower range of F1-scores compared to accuracy in all datasets and appear to have significantly different scores between the two metrics.

TABLE 5  
 CLASSIFICATION RESULT OF ASPECT-BASED SENTIMENT

Dataset	Aspect	Method	Sentiment	
			F1-Score	Accuracy
1 <sup>st</sup> Dataset	House	BERT	59%	95%
		SVM	56%	94%
		Naïve Bayes	32%	91%
		Random Forest	76%	96%
	Fasting	BERT	32%	91%
		SVM	32%	91%
		Naïve Bayes	32%	91%
		Random Forest	37%	92%
	Price	BERT	50%	93%
		SVM	45%	92%
		Naïve Bayes	32%	91%
		Random Forest	70%	95%
2 <sup>nd</sup> Dataset	Government	BERT	78%	80%
		SVM	75%	78%
		Naïve Bayes	54%	64%
		Random Forest	72%	77%
	Homecoming	BERT	44%	61%
		SVM	74%	78%
		Naïve Bayes	52%	64%
		Random Forest	71%	75%
	Instruction	BERT	76%	79%
		SVM	69%	74%
		Naïve Bayes	49%	62%
		Random Forest	66%	72%

Table 5 shows different classification results from sentiment analysis on the first and second datasets, following their various aspects. The top three aspects of the first dataset consistently obtained the best results using the Random Forest method. Meanwhile, the second dataset acquired the best result using the BERT method in two aspects, except 'homecoming,' which got the best score using the SVM method. Considering the results of the two datasets above, more widely used aspects tend to have better scores.

TABLE 6  
 CLASSIFICATION RESULT OF ASPECT-BASED & CORRELATION-BASED

Dataset	Aspect	Method	Sentiment		Emotion	
			F1-Score	Accuracy	F1-Score	Accuracy
1 <sup>st</sup> Dataset	House	BERT	94%	95%	96%	96%
		SVM	92%	94%	93%	95%
		Naïve Bayes	87%	91%	91%	94%
		Random Forest	95%	96%	96%	97%
	Fasting	BERT	87%	91%	93%	95%
		SVM	87%	91%	94%	92%
		Naïve Bayes	87%	91%	88%	94%
		Random Forest	88%	92%	96%	96%
	Price	BERT	91%	93%	91%	93%
		SVM	90%	92%	93%	91%
		Naïve Bayes	87%	91%	88%	93%
		Random Forest	95%	95%	95%	96%
2 <sup>nd</sup> Dataset	Government	BERT	80%	80%	69%	69%
		SVM	78%	78%	68%	69%
		Naïve Bayes	61%	64%	36%	47%
		Random Forest	76%	77%	60%	62%
	Homecoming	BERT	54%	61%	65%	65%
		SVM	77%	78%	67%	68%
		Naïve Bayes	60%	64%	36%	47%
		Random Forest	74%	75%	61%	63%
	Instruction	BERT	78%	79%	58%	63%
		SVM	72%	74%	67%	68%
		Naïve Bayes	57%	62%	35%	47%
		Random Forest	71%	72%	59%	62%

Based on Table 6, different classification results are obtained from the first and second datasets. The best F1-score and accuracy are consistently obtained using Random Forest on the first dataset in all scenarios, with the highest scores



of 95% and 96% for sentiment, while 96% and 97% for emotion, obtained by the 'house' aspect, respectively. Meanwhile, using BERT and SVM, the second dataset acquired composite scores for both performance metrics. The highest scores of the second dataset were consistently obtained using the BERT method, with a score of 80% and 80% for sentiment, 69% and 69% for emotion. Overall, BERT, SVM, and Random Forest methods dominate each performance metric's best result, with the first dataset acquiring a better range of scores.

## V. DISCUSSION

This study has two major limitations. First, this study used two datasets in practice because the first dataset was taken at the beginning of the Covid-19 outbreak, and it mostly didn't discuss about some issues during the pandemic. So, it made it a bit difficult to determine the aspects related to the public opinion of pandemic which is the main topic in this study. Therefore, we needed another dataset with information coming from the midst of the pandemic period as a comparison. Afterwards, we found the second dataset from Kaggle with many more samples that were quite relevant to the pandemic period, and finding aspects related to Covid-19 became easier. Second, both datasets have imbalanced distribution regarding the sentiment and emotion. This could be the main reason why our research results have so much variation in the end [25]. Regardless, an imbalanced distribution is more realistic than a balanced one when it comes to a real-world case. In the future, the imbalanced datasets of this study can be handled using oversampling and undersampling methods [26].

Pearson coefficient is usually used in statistics process to check the correlation within two variables, and indicates the more positive score they get, the more connected they are, and vice versa [14]. However, the second dataset obtained a better correlation for both sentiment and emotion. It consistently outperformed the first dataset in those calculations due to the big difference of total samples in each dataset. As for sentiment, the second dataset's score of polarity and subjectivity correlation is better by 83% compared to the first dataset. However, they both obtained positive scores, which shows the polarity and subjectivity on each dataset have a strong correlation, but the second dataset is stronger. Meanwhile, the second dataset is better by 91% compared to the first dataset in terms of emotions correlation between two variables that we used as the main parameter, i.e., happy and sad. Those two variables should have a minus score due to their inequalities [14]. Thus, the second dataset has a weaker connection between the two variables of emotion, which is good.

As explained above, the second dataset has a better correlation in both sentiment and emotion. Fig. 2b shows public opinion, which has much to do with the government, economy, policy, instruction, pandemic, and many others. Based on several studies, such as [30]–[36], research on Covid-19 that was spread to the public during those periods mainly discussed these topics as well. There is compatibility between the wordcloud analysis of the second dataset retrieved from Twitter and the public opinion naturally spread in society about Covid-19. Hence, these topics become fascinating things to be reviewed further.

Furthermore, the result comparison among all conducted scenarios can be generally analyzed. Unlike the first dataset, which consistently obtained the best results using only Random Forest method, the second dataset acquired the mixed highest results using two different methods, i.e., BERT and SVM, and is considered inconsistent. By comparing the overall results, we can also see that the average sentiment result, aspect-based or not, has a better range than the result of emotion scenarios. The experimental results continue to improve when aspect-based and correlation-based are applied to more complex sentiment and emotion scenarios. This advantage yields the best F1-score and accuracy for emotion scenario with the 'house' aspect as much as 96% and 97%, respectively.

## VI. CONCLUSIONS

This study aims to understand public opinion on Covid-19 based on two tweet datasets using aspect-based sentiment analysis and emotion detection. The following aspects of the first dataset are house, fasting, and price, while the second one has aspects of government, homecoming, and instruction. Since aspects could not sufficiently represent public opinion, our work has shown that the Pearson coefficient was able to check the correlation between two variables, i.e., subjectivity and polarity to build the sentiment, as well as happy and sad variables to build the emotion. As a result, the government's efforts to overcome Covid-19 effects can be discovered through tweets, considering the second dataset showed a better correlation than the first dataset in those both ways. With classifiers of BERT, Naïve Bayes, Random Forest, and SVM, as well as the correlation usage, the outcome of classification performance metrics obtained various best results for each scenario. The average score of all sentiment scenarios is better than the emotion scenarios. Even so, sentiment and emotion cases have inconsistent results and are gradually better after using aspect and correlation. These are evidenced by several different ranges of scores on the two performance metrics, even though using the same method on the same dataset. Moreover, most people use emojis when expressing their opinions, so when preprocessing is done, much data are wasted because the words obtained have to meet the requirements to be

analyzed. On the other hand, an emoji is a form of emotional expression that can also describe public opinion. So future work can convert emojis into text before preprocessing to get more accurate results during the experiments.

**Author Contributions:** *Salsabila*: Conceptualization, Methodology, Writing - Original Draft, Writing - Review & Editing. *Tyas*: Conceptualization, Methodology, Writing - Original Draft, Writing - Review & Editing. *Romadhona*: Software, Conceptualization, Methodology, Writing - Original Draft, Writing - Review & Editing. *Purwitasari*: Validation, Investigation, Supervision.

**Funding:** Research received no specific grant from any funding agency.

**Conflicts of Interest:** *Purwitasari* is member of Editorial Teams, but had no role in the decision to publish this article. No other potential conflict of interest relevant to this article was reported.

#### REFERENCES

- [1] A. Roziqin, S. Y. F. Mas'udi, and I. T. Sihidi, "An analysis of Indonesian government policies against COVID-19," *Public Administration and Policy*, vol. 24, no. 1, pp. 92–107, 2021, doi: 10.1108/PAP-08-2020-0039.
- [2] S. R. Anggraeni, N. A. Ranggianto, I. Ghazali, C. Faticah, and D. Purwitasari, "Deep learning approaches for multi-label incidents classification from Twitter textual information," *Journal of Information Systems Engineering and Business Intelligence*, vol. 8, no. 1, pp. 31–41, 2022, doi: 10.20473/jisebi.8.1.31-41.
- [3] I. Aygun, B. Kaya, and M. Kaya, "Aspect based Twitter sentiment analysis on vaccination and vaccine types in COVID-19 pandemic with deep learning," *IEEE J Biomed Health Inform*, vol. 26, no. 5, pp. 2360–2369, 2022, doi: 10.1109/JBHI.2021.3133103.
- [4] A. García-Pablos, M. Cuadros, and G. Rigau, "W2VLDA: Almost unsupervised system for aspect based sentiment analysis," *Expert Syst Appl*, vol. 91, pp. 127–137, 2018, doi: 10.1016/j.eswa.2017.08.049.
- [5] I. Surjandari, R. A. Wayasti, E. Laoh, Zulkarnain, A. M. M. Rus, and I. Prawiradinata, "Mining public opinion on ride-hailing service providers using aspect-based sentiment analysis," *International Journal of Technology*, vol. 10, no. 4, pp. 818–828, 2019, doi: 10.14716/ijtech.v10i4.2860.
- [6] A. Mathur, P. Kubde, and S. Vaidya, "Emotional analysis using Twitter data during pandemic situation: COVID-19."
- [7] M. E. Mowlaei, M. Saniee Abadeh, and H. Keshavarz, "Aspect-based sentiment analysis using adaptive aspect-based lexicons," *Expert Syst Appl*, vol. 148, 2020, doi: 10.1016/j.eswa.2020.113234.
- [8] A. Nazir, Y. Rao, L. Wu, and L. Sun, "Issues and challenges of aspect-based sentiment analysis: a comprehensive survey," *IEEE Transactions on Affective Computing*, vol. 13, no. 2. Institute of Electrical and Electronics Engineers Inc., pp. 845–863, 2022. doi: 10.1109/TAFFC.2020.2970399.
- [9] F. A. Acheampong, C. Wenyu, and H. Nunoo-Mensah, "Text-based emotion detection: advances, challenges, and opportunities," *Engineering Reports*, vol. 2, no. 7. John Wiley and Sons Inc, 2020. doi: 10.1002/eng2.12189.
- [10] P. Nandwani and R. Verma, "A review on sentiment analysis and emotion detection from text," *Social Network Analysis and Mining*, vol. 11, no. 1. Springer, 2021. doi: 10.1007/s13278-021-00776-6.
- [11] S. Boon-Itt and Y. Skunkan, "Public perception of the COVID-19 pandemic on twitter: sentiment analysis and topic modeling study," *JMIR Public Health Surveill*, vol. 6, no. 4, 2020, doi: 10.2196/21978.
- [12] A. S. Imran, S. M. Daudpota, Z. Kastrati, and R. Batra, "Cross-cultural polarity and emotion detection using sentiment analysis and deep learning on covid-19 related tweets," *IEEE Access*, vol. 8, pp. 181074–181090, 2020, doi: 10.1109/ACCESS.2020.3027350.
- [13] C. Qian, N. Mathur, N. H. Zakaria, R. Arora, V. Gupta, and M. Ali, "Understanding public opinions on social media for financial sentiment analysis using AI-based techniques," *InfProcess Manag*, vol. 59, no. 6, 2022, doi: 10.1016/j.ipm.2022.103098.
- [14] W. J. Wu and Y. Xu, "Correlation analysis of visual verbs' subcategorization based on Pearson's correlation coefficient," in *2010 International Conference on Machine Learning and Cybernetics, ICMLC 2010*, 2010, vol. 4, pp. 2042–2046. doi: 10.1109/ICMLC.2010.5580507.
- [15] H. Adamu, S. L. Lutfi, N. H. A. H. Malim, R. Hassan, A. di Vaio, and A. S. A. Mohamed, "Framing twitter public sentiment on Nigerian government COVID-19 palliatives distribution using machine learning," *Sustainability (Switzerland)*, vol. 13, no. 6, 2021, doi: 10.3390/su13063497.
- [16] F. A. Acheampong, H. Nunoo-Mensah, and W. Chen, "Transformer models for text-based emotion detection: a review of BERT-based approaches," *Artif Intell Rev*, vol. 54, no. 8, pp. 5789–5829, 2021, doi: 10.1007/s10462-021-09958-2.
- [17] S. Thara and S. Sidharth, "Aspect based sentiment classification: Svd features," in *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 2017, pp. 2370–2374. doi: 10.1109/ICACCI.2017.8126201.
- [18] S. S. Aljameel *et al.*, "A sentiment analysis approach to predict an individual's awareness of the precautionary procedures to prevent covid-19 outbreaks in Saudi Arabia," *Int J Environ Res Public Health*, vol. 18, no. 1, pp. 1–12, 2021, doi: 10.3390/ijerph18010218.
- [19] I. Perikos and I. Hatzilygeroudis, "Aspect based sentiment analysis in social media with classifier ensembles," in *Proceedings - 16th IEEE/ACIS International Conference on Computer and Information Science, ICIS 2017*, 2017, pp. 273–278. doi: 10.1109/ICIS.2017.7960005.
- [20] N. Leelawat *et al.*, "Twitter data sentiment analysis of tourism in Thailand during the COVID-19 pandemic using machine learning," *Heliyon*, vol. 8, no. 10, 2022, doi: 10.1016/j.heliyon.2022.e10894.

- [21] Y. al Amrani, M. Lazaar, and K. E. el Kadir, "Random forest and support vector machine based hybrid approach to sentiment analysis," in *Procedia Computer Science*, 2018, vol. 127, pp. 511–520. doi: 10.1016/j.procs.2018.01.150.
- [22] M. G. Sousa, K. Sakiyama, L. D. S. Rodrigues, P. H. Moraes, E. R. Fernandes, and E. T. Matsubara, "BERT for stock market sentiment analysis," in *Proceedings - International Conference on Tools with Artificial Intelligence, ICTAI*, Nov. 2019, 2019, pp. 1597–1601. doi: 10.1109/ICTAI.2019.00231.
- [23] M. Pota, M. Ventura, H. Fujita, and M. Esposito, "Multilingual evaluation of pre-processing for BERT-based sentiment analysis of tweets," *Expert Syst Appl*, vol. 181, 2021, doi: 10.1016/j.eswa.2021.115119.
- [24] M. Behzadi, I. G. Harris, and A. Derakhshan, "Rapid Cyber-bullying detection method using Compact BERT Models," in *Proceedings - 2021 IEEE 15th International Conference on Semantic Computing, ICSC 2021*, 2021, pp. 199–202. doi: 10.1109/ICSC50631.2021.00042.
- [25] Apriantoni Apriantoni, D. Purwitasari, and A. B. Raharjo, "Indonesian tweets dataset for identifying emotion changes among Twitter users following the onset of the COVID-19," 2022. <https://data.mendeley.com/datasets/x8t4gn6mt6> (accessed Feb. 25, 2023).
- [26] Dionisius Darryl Hermansyah, "COVID-19 Indonesian tweets," 2021. <https://www.kaggle.com/dionisiusdh/covid19-indonesian-tweets-sentiment> (accessed Feb. 25, 2023).
- [27] Buitinck et al, "API design for machine learning software: experiences from the scikit-learn project," 2013.
- [28] "BERT." [https://huggingface.co/docs/transformers/model\\_doc/bert](https://huggingface.co/docs/transformers/model_doc/bert) (accessed Feb. 08, 2023).
- [29] J. He, L. Xu, Y. Fan, Z. Xu, M. Yan, and Y. Lei, "Deep learning based valid bug reports determination and explanation," in *Proceedings - International Symposium on Software Reliability Engineering, ISSRE*, Oct. 2020, 2020, pp. 184–194. doi: 10.1109/ISSRE5003.2020.00026.
- [30] A. Spicer, "Organizational culture and COVID-19," *Journal of Management Studies*, vol. 57, no. 8. Blackwell Publishing Ltd, pp. 1737–1740, 2020. doi: 10.1111/joms.12625.
- [31] J. M. Baena-Díez, M. Barroso, S. I. Cordeiro-Coelho, J. L. Díaz, and M. Grau, "Impact of COVID-19 outbreak by income: hitting hardest the most deprived," *Journal of Public Health (United Kingdom)*, vol. 42, no. 4, pp. 698–703, 2020, doi: 10.1093/pubmed/fdaa136.
- [32] M. Kowal et al., "Who is the most stressed during the COVID-19 pandemic? data from 26 countries and areas," *Appl Psychol Health Well Being*, vol. 12, no. 4, pp. 946–966, 2020, doi: 10.1111/aphw.12234.
- [33] F. Glowacz and E. Schmits, "Psychological distress during the COVID-19 lockdown: the young adults most at risk," *Psychiatry Res*, vol. 293, 2020, doi: 10.1016/j.psychres.2020.113486.
- [34] P. Jiang, J. J. Klemeš, Y. van Fan, X. Fu, and Y. M. Bee, "More is not enough: a deeper understanding of the covid-19 impacts on healthcare, energy and environment is crucial," *International Journal of Environmental Research and Public Health*, vol. 18, no. 2. MDPI AG, pp. 1–22, 2021. doi: 10.3390/ijerph18020684.
- [35] S. P. J. M. Horbach, "Pandemic publishing: Medical journals strongly speed up their publication process for covid-19," *Quantitative Science Studies*, vol. 1, no. 3, pp. 1056–1067, 2020, doi: 10.1162/qss\_a\_00076.
- [36] S. Gautam and L. Hens, "COVID-19: impact by and on the environment, health and economy," *Environment, Development and Sustainability*, vol. 22, no. 6. Springer, pp. 4953–4954, 2020. doi: 10.1007/s10668-020-00818-7.

**Publisher's Note:** Publisher stays neutral with regard to jurisdictional claims in published maps and institutional affiliation