Tweets Emotions Analysis of Community Activities Restriction as COVID-19 Policy in Indonesia Using Support Vector Machine

Abi Nizar Sutranggono[1] and Elly Matul Imah[2*]
[1]Mathematic Department, Universitas Negeri Surabaya
Jawa Timur 60213, Indonesia
[2]Data Science Department, Universitas Negeri Surabaya
Jawa Timur 60213, Indonesia
Email: [1]abinizar76@gmail.com, [2]ellymatul@unesa.ac.id

***Abstract*—With the rising number of COVID-19 cases in Indonesia, the government has implemented the Imposition of Restrictions on Emergency Community Activities (Pemberlakuan Pembatasan Kegiatan Masyarakat - PPKM) as Indonesia's COVID-19 policy. Several controversies and protests have colored the implementation of this emergency policy. Some netizens on Twitter voice their opinions about the policy in their tweets. Emotions in tweets can be recognized through text-based emotion detection or emotion analysis. However, text-based emotion detection is a challenging task. One of the main issues in classifying text with a machine learning-based approach deals with the feature dimensions. As a result, appropriate methods for accurately identifying emotion based on the text are required. The research studies an emotions analysis task on Indonesians' PPKM-related tweets to understand their emotional state while implementing the PPKM. The machine learning classification algorithms used are Support Vector Machine (SVM) and random forest. The total number of tweets is 4,401. The results show that SVM with linear kernel function combined with the TF-IDF and Chi-Square methods outperforms other classifiers with an accuracy of 0.7528. The accuracy value is higher than those obtained by previous studies. Moreover, the results of the emotion classification on PPKM tweets reveal that most Indonesians are unhappy with the implementation of the PPKM policy.**

***Index Terms*—Tweets, Emotions Analysis, COVID-19 Policy, Support Vector Machine**

I. Introduction

THE Coronavirus Disease 2019 (COVID-19) has circulated widely throughout the world. This outbreak was first discovered in late December 2019 in Wuhan City, Hubei Province, China. The transmission of the SARS-CoV-2 virus that causes COVID-19 is fast, so the spreading rate of the COVID-19 disease is high. For that reason, World Health Organization (WHO) has announced COVID-19 as a pandemic [1]. In response to the increasing number of COVID-19 cases, the Indonesian government has implemented an Imposition of Restrictions on Emergency Community Activities (Pemberlakuan Pembatasan Kegiatan Masyarakat - PPKM) policy. This policy is applied in Java and Bali areas, designated as red zones. The implementation of the PPKM began on July 3rd, 2021 [2]. However, implementing this PPKM policy has caused various controversies and protests from a group of citizens, specifically people who felt affected by its implementation. Moreover, some netizens have expressed their opinion or frustration towards the PPKM on social media.

Social networks evolve rapidly in tandem with technological advancement. As a result, various information is uploaded and shared by social media users [3]. That information can be in text, videos, photos, or audio. In Indonesia, several social media are used and accessed by the public, such as Twitter, Facebook, and Instagram. Twitter is one of the social media platforms with the highest total daily active users [4]. Twitter's design and features make it simple for users to share messages, known as tweets. Each tweet also contains a wide range of content. Some tweets have vents, opinions, or positive and negative criticisms. Governments, scientists, and companies also attempt to process information from Twitter to get some insight despite the varying content of the tweets.

Furthermore, during the COVID-19 pandemic, Twitter has become a place for people to express their opinions, thoughts, and feelings via tweets [5]. A phrase

Received: Feb. 10, 2022; received in revised form: June 13, 2022; accepted: July 01, 2022; available online: March 17, 2023.
*Corresponding Author

conveyed through a tweet can reflect how a person’s feelings or emotions are experienced. Emotions are human’s feelings strongly influenced by a situation or a relationship with another person [6]. The emotions in a social media post can be recognized through emotion detection, a sub-part of sentiment analysis. Sentiment analysis focuses on obtaining information by interpreting a person’s point of view or opinions and classifying it into polarities, such as positive, negative, or neutral [7]. Meanwhile, emotion detection is a subset of sentiment analysis that entails extracting a person’s opinions, thoughts, or ideas to identify emotions, such as sadness, anger, happiness, and so on [8].

Emotion detection from social media posts is a relatively new and challenging area of computational intelligence research that has caught researchers’ attention in recent years [9]. Even though there are numerous ways to communicate on social media, text remains one of the most widely used forms of communication. As a result, text-based emotion detection has become an essential part of the research. The cases discussed and the techniques employed are quite diverse. A machine learning-based approach is one of the most widely used methods for detecting emotions in text [10]. For example, previous research classifies emotions from Indian text data into five emotions: surprise, happy, sad, fear, and love, using Support Vector Machine (SVM) [11]. In-text classification, SVM is one of the machine learning algorithms that is frequently implemented [12]. In addition, machine learning can be used to detect emotions in tweets. Another example is the emotion classification system of English-language tweets using SVM, Logistic Regression (LR), random forest, and Stochastic Gradient Descent (SGD) algorithms [13]. On the other hand, K-Nearest Neighbors (KNN) and Naïve Bayes (NB) algorithms are used to identifying emotions expressed through tweets [14].

Feature engineering techniques must be implemented to analyze emotions presented in text data by adapting machine learning algorithms [15]. The word embedding method is one of the feature engineering methods commonly used in Natural Language Processing (NLP). Each word in the text can be represented numerically based on text corpora using the word2vec embedding method [16]. Nevertheless, this method has a limitation. The word2vec cannot accurately insert words into vectors if the word is not found in a corpus [17]. The FastText embedding method, on the other hand, can solve this Out of Vocabulary (OOV) issue [18]. Furthermore, another approach can be applied to represent numbers from text datasets to employ the word embedding-based method. It is the Term Frequency-Inverse Document Frequency method (TF-IDF) [19].

One of the main challenges when performing text classification tasks using a machine learning-based approach is dealing with the dimensions of the features used [20]. The dimensions of input features can affect the overall performance of the classification models, especially when the feature space has high dimensions. Hence, dimensionality reduction is critical in employing machine learning-based algorithms for the classification task. Moreover, the dimensionality reduction process is implemented to reduce the dimensions of input data. Feature extraction or feature selection can minimize the dimensionality of input data. Previous research performs a text classification task using different machine learning algorithms, including SVM, KNN, random forest, and decision tree (J-48), by implementing the feature extraction process using the Principal Component Analysis (PCA) method [21]. After applying PCA, high feature vector dimensions can be reduced. PCA can also remove irrelevant features, allowing classifiers to classify text better. It is an effective and popular feature extraction method used in various fields [22].

Moreover, the Analysis of Variance (ANOVA) feature selection method is applied to reduce the dimensions of the input data so that the performance of machine learning classifiers in analyzing Arabic-language tweets is maximized [23]. On the other hand, a previous study employes a modified Chi-Square feature selection technique to improve the performance of SVM in classifying Arabic text data [24]. In addition, a dimension reduction process is applied using the Chi-Square feature selection method to improve the classifier’s performance when executing emotion detection tasks for English text using the SVM algorithm [25].

Finding an appropriate feature representation of text data representing an entity is not straightforward [26]. Furthermore, feelings expressed through a text are not always expressed through words directly related to emotions but are frequently understood through the context or meaning of the text [27]. These factors make text-based emotion detection tasks require applying the proper method to obtain accurate emotional identification results. The research conducts a text-based emotion analysis of PPKM policy using the SVM and random forest, based on Indonesian-language Twitter text data uploaded when the policy takes place. FastText and TF-IDF methods are used to represent text data in numeric form.

Only a few studies apply the dimension reduction process to improve the performance of the classification model in identifying text-based emotion. On

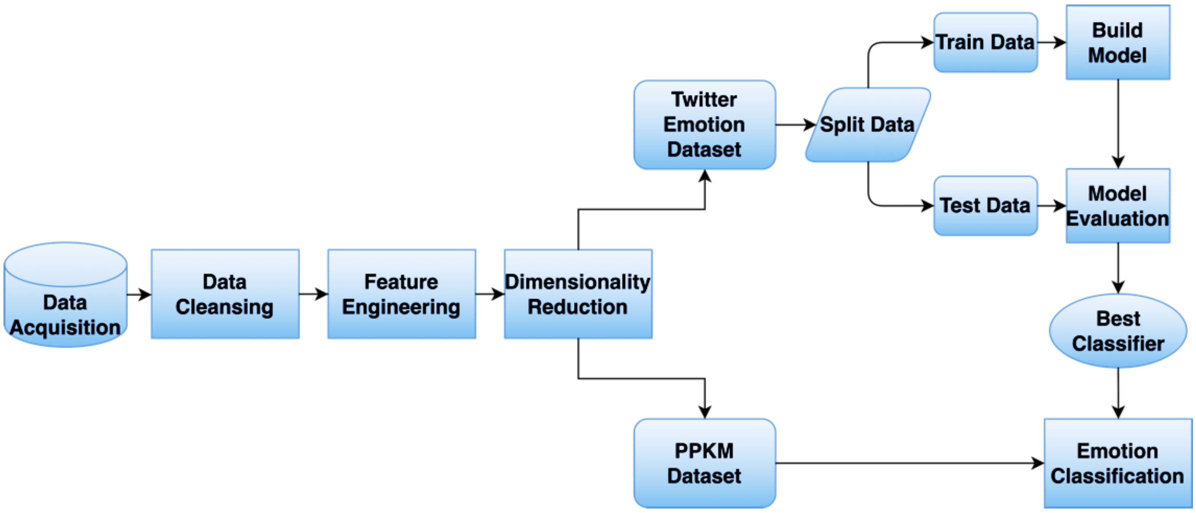

Fig. 1. The emotion classification process conducted in the research.

the other hand, most studies have adapted the dimension reduction process to improve the accuracy of the classifier in classifying text or sentiment data into certain classes [21, 23, 24]. The rest, most of the texts analyzed, are in English [25, 28]. Therefore, the research implements a dimension reduction procedure to overcome the existing limitations and improve each classifier's performance in classifying emotions based on Indonesian-language tweets. The dimensional reduction methods that will be compared are PCA, ANOVA, and Chi-Square.

II. Research Method

Text-based emotion detection combines a set of predefined emotion-labeled datasets with a data analysis algorithm to extract emotions from textual data [28]. Text-based emotion detection employs an NLP approach that combines linguistic and computational techniques to assist computers in understanding text in the form of human language [8]. Furthermore, text-based emotion detection can overview public perceptions on a particular topic, product, service, or event. In Fig. 1, the emotion detection process in the research consists of five major stages, including data acquisition, data cleansing, feature engineering, dimensionality reduction, and emotions classification. The process of data collection or data acquisition is divided into two types: data collection to build a classification model and data collection to be used as a research subject. Furthermore, data cleansing aims to prepare data for further processing by removing noise from each data. FastText and TF-IDF methods are used in the feature engineering process to convert text data to numbers. Then, the dimensionality reduction stage employs several methods, including PCA, ANOVA, and Chi-Square, to reduce the dimensions of features obtained during the feature engineering stage.

On the other hand, a classification model or classifier will be created using trained data in the emotions classification process. Each formed classifier is tested on test data, and its performance is evaluated. In addition, the best classifier will be used to classify tweets about PPKM policy into five emotion categories. The five emotion categories are anger, happiness, sadness, fear, and love.

A. Dataset

The data used to build a classification model or classifiers in the research is obtained from the Twitter emotion dataset [29]. The dataset used is a collection of Indonesian tweets collected from June 1st, 2018, to June 14th, 2018, using the streaming method via Twitter API. The total number of tweets is 4,401. In addition, each tweet has been labeled by the researchers. To determine the emotion class of the tweet, the researchers used Shaver's basic emotion theory. It is later popularized by Parrot, known as Parrot's primary emotion. This emotion theory divides emotions into six classes: surprise, love, joy, anger, sadness, and fear [30].

On the other hand, Shaver and Murdaya have updated the Indonesian language's structural definition of the emotion lexicon based on previous emotion theory but have yet to include surprise emotion [31]. Thus, the emotion lexicon for the Indonesian language used by the researchers to establish the annotation guidelines only has five emotion classes: anger, happiness,

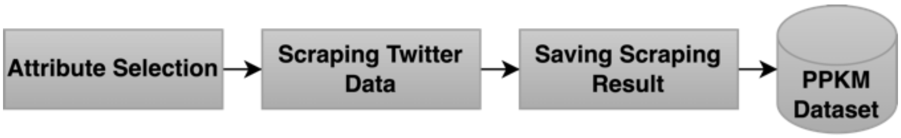

Fig. 2. The Twitter data scraping process.

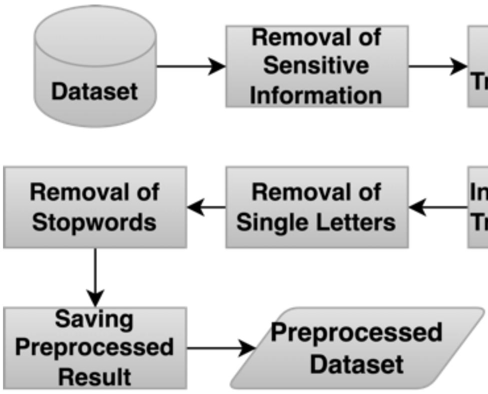

Fig. 3. Data cleansing process.

sadness, fear, and love. The annotation process is also carried out with the assistance of two professional annotators. Additionally, the researchers use Cohen Kappa’s value assessments to evaluate the agreement of the final annotation results of tweets.

B. Data Acquisition

The data used as a research subject is primary data. Twitter is used as a data collection place for the research. On the other hand, the data collection process is carried out using a scraping technique that scrapes tweets associated with the PPKM policy. Scraping is a data extraction technique applied to collect large amounts of relevant metadata, which is then stored in a database or specific files for future analysis [32]. In the research, scraping uses Python programming language with the help of scrape modules. Snscrape is one of the Python 3 libraries that can access historical data from social media platforms such as Twitter [33]. Figure 2 depicts the flow of the scraping process implemented in the research. The first scraping stage determines attributes, such as keywords, language, and the time-lapse between uploaded tweets.

Moreover, the research uses “ppkm” as the keyword. The selected tweets are in the Indonesian language. Besides that, the tweets are posted from July 3rd, 2021, to October 4th, 2021. After determining the attributes, the scraping process begins. The scraping process results are saved in a .csv format file.

C. Data Cleansing

Identifying emotion based on Twitter text data is challenging [34]. It is because Twitter text data is not structured. Nowadays, abbreviations, emojis, or particular phrases are widely used in tweets. On the other hand, a similar problem exists on Twitter Indonesia. In addition, tweets uploaded by netizens consist of various types of informal words [29]. Therefore, before performing the sentiment analysis task, Twitter text data must be prepared in advance through data cleansing or preprocessing. Text preprocessing aims to convert unstructured text into structured text data [35]. Then, data cleansing is carried out using several tools, including RegEx, NLTK, and demoji. In Fig. 3, the data cleansing process applied to the dataset is as follows.

The removal of sensitive information process is implemented to remove sensitive information, such as mentions, URLs, hashtags, and sensitive numbers on the dataset’s characteristics. Each emoticon, such as “:)”, “:’(”, and “XD” is converted into a corresponding word form in the step of emoticons transformation. Then, the punctuation in a tweet is removed in the removal of punctuation process. Next, each letter in the text is modified to lowercase in the lower casing process. The next step is to transform the informal word into formal words. For example, “pls” is replaced with “please”. Afterward, the research removes single letters in a text. Stopwords often appear in the text but do not have an informative meaning [36]. Hence, the research removes every stopword from the text. It also

TABLE I
EXAMPLES OF THE WORD TRANSFORMATION.

Symbol & slang words	Word transformation
:)	Smile
:'(	Crying, sad
XD	Laugh out loud
pls	Need help
ngakak	Laugh
ily	I love you

deletes emojis and too-short tweets from the PPKM dataset. The results of the data cleansing process are stored in a .csv format file. Table I presents examples of emoticons and informal word transformations.

D. Feature Engineering

Feature Engineering converts text data into numbers. The research uses FastText and TF-IDF as feature engineering. In sentiment analysis, word embedding is a technique for extracting features from text data [37]. Word embedding is applied to jolt a word into a vector of real numbers with N dimensions [38]. Applying the word embedding method makes the text classification process easier [39]. One of the word embedding techniques is FastText. Moreover, FastText is a development of the word2vec embedding method that can deal with the issue of OOV by using the concept of similarity or resemblance of meaning from a word [40]. For example, a group of Facebook researchers has developed FastText to study word representation efficiently. FastText considers the word structure that its numerical representation looks for something. Thus, a word with a different form studies the similarity of the word independently [41]. Moreover, FastText has been implemented in previous research using a pre-trained FastText model with 100 dimensions [29]. In particular, the pre-trained FastText model has been trained using one million tweets in the Indonesian language. The pre-trained model is used to find the numerical representation of each word in a tweet by embedding the word into a real number vector based on the phrase loaded by the pre-trained FastText model. On the other hand, that pre-trained model contains 69,465-word vectors.

Every word that its numerical representation will look for is weighted by TF-IDF. The weight given by the TF-IDF to a word reflects how important the word is in the documents. TF-IDF is often implemented as a feature engineering method in sentiment analysis tasks. TF-IDF is one of the feature engineering approaches that consider word frequency in a text [42]. In addition, Term Frequency (TF) in Eq. (1) is the rate of occurrence of a word in the documents. Then, Eq. (2) is the Inverse Document Frequency (IDF) that presents the occurrence level of words rarely encountered in the document. The TF-IDF value is obtained using Eq. (3).

$$\text{TF}(w) = \frac{\sum \text{presences of word } w \text{in document } d}{\sum \text{word in document } d}, \tag{1}$$

$$\text{IDF}(w) = \log_e \frac{\sum \text{document collections}}{\sum \text{document consisting word } w}, \tag{2}$$

$$\text{TF} - \text{IDF}(w) = \text{TF}(w) \times \text{IDF}(w). \tag{3}$$

E. Dimensionality Reduction

The FastText method produces features with dimensions of 4401×100, while the TF-IDF method produces features with dimensions of 4401×15386. Hence, the dimensionality reduction step is implemented using the PCA feature extraction method and the feature selection method with ANOVA and Chi-Square techniques to reduce the features' dimensions.

PCA approach is used to reduce the dimensions of the input data into smaller dimensions by considering the variance of the data. Furthermore, PCA is applied to find a form of linear transformation that states a coordinate framework from input data to new orthonormal coordinates [43]. On the other hand, the linear transformation of input data into new data is called Principal Components (PCs). Retrieval from PCs represents the size of the data stored. By using PCA, the feature extraction algorithm is as follows.

1) Normalization of input data.
2) Calculate the covariance matrix.

$$cov(xy) = \frac{\sum xy}{n} - (\bar{x})(\bar{y}) \tag{4}$$

3) Compute the eigenvalues and the eigenvectors.

$$(A - \lambda I) = 0 \tag{5}$$

4) Sort eigenvalues descending and specifying PCs.

$$[A - \lambda I][X] = [0] \tag{6}$$

Next, the ANOVA feature selection technique ranks a feature based on the feature's linkage to a label. The benchmark used to evaluate the rating of each feature is called the f ratio [44]. The f ratio can be found using Eq. (7), σ^2_{cl} denoting the variance between classes and σ^2_{err} denoting a variance within the class. Moreover, σ^2_{cl} is calculated through Eq. (8), and the variance within the class is calculated by Eq (9).

$$fratio = \frac{\sigma_{cl}^2}{\sigma_{err}^2}, \quad (7)$$

$$\sigma_{cl}^2 = \frac{\sum(\bar{xi} - \bar{x})^2}{k - l}, \quad (8)$$

$$\sigma_{err}^2 = \frac{(\sum(x_{ij} - \bar{x})^2) - (\sum(\bar{x}_i - \bar{x})^2 n_i)}{N - k}. \quad (9)$$

Additionally, $\bar{xi}$ represents the average of the ith class. The $\bar{x}$ is the overall average, x_{ij} represents the ith measurement of jth class, and n_i is the number of ith class measurements. The highest f ratio indicates the more relevant a feature to the label or class.

Chi-Square (χ^2) is a statistical test that measures the divergence from the expected distributions if the appearance of a feature is assumed to be independent of its classes [45]. The Chi-Square test also considers dependencies among stochastic variables and is well-defined. Therefore, features that are not dependent on their class will be eliminated because they are considered irrelevant or essential for a classification process. It shows r as the number of different values in the feature, c as the number of various values in the class, n_{js} as the frequency of the j-th class element with the sth class, μ_{js} and n_{j*} as the frequency of the j-th element, and n_{*s} as the total element with the s-th class. Each feature is assigned a value using Eqs. (10) and (11). In addition, the highest Chi-Square value indicates that a feature is very informative or helpful for a classification process [46].

$$\mu_{js} = \frac{n_{*s} n_{j*}}{n}, \quad (10)$$

$$\chi_f^2 = \sum_{j=1}^{r} \sum_{s=1}^{c} \frac{(n_{js} - \mu_{js})^2}{\mu_{js}}. \quad (11)$$

F. Emotion Classification

The research classifies tweets into five emotions: anger, happiness, sadness, fear, and love. On the other hand, the classification algorithms involved are SVM and random forest. Non-Linear SVM performs data classification by determining the best hyperplane constructed from the data set that separates each class. This data set is the closest point to a hyperplane called a support vector [47]. Furthermore, the best hyperplane can be measured by calculating the value of the maximum margin, where the margin represents the distance between a class to another class. On the other hand, the performance of SVM can be affected by the selected parameters. The parameter in question is the kernel functions adapted. There are several kernel functions. Equations (12)–(14) define the linear kernel, Radial Basis Function (RBF) kernel, and polynomial kernel, respectively. It shows x and y as input data, c as a constant value, d as a polynomial degree, and σ as the standard deviation. Suppose a dataset has been labeled into two classes: $x_i \in R^d$, $y_i \in -1, +1$ with $i = N$, $d > 1$, and hyperplane $g(x) = (w, x) + c$. The best hyperplane is recognized by maximizing the margin value of Eq. (15). The applicable provision is in Eq. (16). Additionally, the sign decision $f(x)$ in Eq. (17) is calculated to determine the class of the data, with m as the support vector, α_i as the weight of each data, and $K(x_i, x)$ as the kernel function.

$$K(x, y) = (x^T y), \quad (12)$$

$$K(x, y) = exp\left\{\frac{\|x - y\|^2}{2\sigma^2}\right\}, \quad (13)$$

$$K(x, y) = (x^T y + c)^d, \quad (14)$$

$$\frac{1}{2}\|W\|^2 = \frac{1}{2}(W_1^2 + W_2^2), \quad (15)$$

$$(w_1 x_i + w_2 x_i + c) \geq 1, \quad (16)$$

$$f(x_d) = \sum_{i=1}^{m} \alpha_i yi K(x_i, x) + y. \quad (17)$$

Random forest is a classification algorithm that uses bootstrap aggregation (bagging) and random feature selection methods [48]. Furthermore, random forest works by analyzing trees grown and becoming forests. The determination of classification results using the random forest algorithm is obtained based on the majority voting results from each tree [49]. Here are the stages involved in forming a forest consisting of random trees:

1) By replacing the data groups (bootstrap), choose N cases randomly.
2) With the example of bootstrap, design a tree to grow to its maximum size without even being pruned. A number of m attributes are selected at random for each tree.
3) Repeat steps 1 and 2 at k times until a forest of k trees has grown.

Suppose that $mg(X, Y)$ is a margin function. The margin function computes the average voting results of each tree at (X, Y) with X as the predictor vector and Y as the classification target. Equation (18) defines the margin function that υ and ν are formulated in Eqs. (19)–(20).

Besides determining the majority voting's result, The margin function represents the level of certainty of the classification's final result. Moreover, random forest uses the Generalization Error (PE^*) parameter

TABLE II
DATA DISTRIBUTIONS FROM EACH EMOTION CLASS.

Class	Training data	Testing data	Total
Anger	992	109	1,101
Happy	929	88	1,017
Sadness	894	103	997
Fear	577	69	649
Love	568	72	637
Total	3,960	441	4,401

to measure the accuracy characteristics in Eq. (21).

$$mg(X,Y) = \upsilon - \max_{j\neq Y}(\nu), \quad (18)$$
$$\upsilon = a\nu_k I(h_k(X) = Y), \quad (19)$$
$$\nu = a\nu_k I(h_k(X) = j), \quad (20)$$
$$PE^* = p_{x,y}(mg(X,Y)) < 0. \quad (21)$$

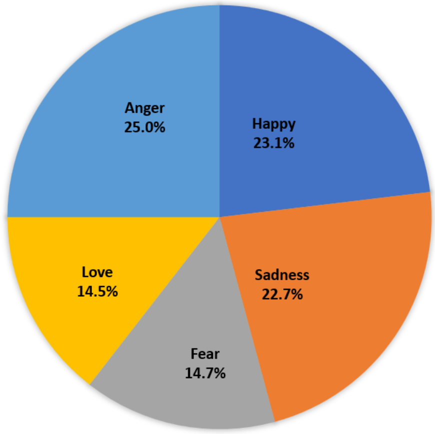

Fig. 4. Comparison of the number of tweets from each emotion class on the Twitter emotion dataset.

III. RESULTS AND DISCUSSION

In the research, Indonesian tweets are classified into five emotions: anger, happiness, sadness, fear, and love. The machine learning-based approach is employed to execute text-based emotion detection. The machine learning-based approach solves the problem of text-based emotion detection by utilizing machine learning algorithms to classify text into specific emotional categories [8]. The research uses SVM and random forest. Each algorithm uses two different feature types: FastText and TF-IDF. Additionally, the research uses SVM with linear, RBF, and polynomial kernels with three degrees. The research also conducts several experiments with and without dimensional reduction. It uses PCA as feature extraction and ANOVA and Chi-Square as feature selection.

Table II shows the tweet distributions from each emotion category in the dataset used after partitioning. Moreover, the Twitter emotion dataset is partitioned into 90% training data and 10% evaluation data [50]. The number of tweets is 3,960 for training data and 441 for evaluation data. On the other hand, data testing used as a research subject is Indonesian tweets related to the PPKM policy. The dataset is obtained through the scraping process.

Figure 4 compares the number of tweets from each emotion class on the Twitter Emotion Dataset. In particular, the difference in the number of tweets from each category of emotions is not statistically significant. Then, from Fig. 5, it can be seen that most of the tweets collected were uploaded in July 2021, totaling 187,678 tweets. Moreover, there are 322,864 tweets in total. After implementing the data cleansing process, the number of tweets on the PPKM dataset is reduced to 191,017.

The f-ratio values given to the features selected by using ANOVA in the research are as follows. The lowest f-ratio of the selected FastText feature is 31.667. Meanwhile, the lowest f-ratio of the selected TF-IDF feature is 1.4779. The highest f-ratio of the selected FastText feature is 248.126, whereas the highest f-ratio value of the selected TF-IDF feature is 412.523. Moreover, the lowest Chi-Square value of the selected FastText feature is 3.6185, while the lowest Chi-Square value of the selected TF-IDF feature is 1.0415. Then, the highest Chi-Square values given to the selected FastText and TF-IDF features are 34.520 and 247.061, respectively.

The performance of each classifier is evaluated using accuracy, precision, recall, and F1 score evaluation metrics. In addition, the time it takes classifiers to classify emotions or the testing time is used as a benchmark for evaluating the performance of each classification model. Figure 6 depicts an evaluation of the processing time in each classifier to classify emotion classes from the testing data. The dimensions of the TF-IDF feature are more significant than the FastText feature. Moreover, the testing time of classifiers with the TF-IDF feature is longer than the FastText feature.

Details of each classifier's evaluation results are presented in Table III. If the process of dimensional reduction is not employed, the highest accuracy achieved by SVM with linear kernel function is built using the TF-IDF feature. The accuracy of the classifier is 0.7052. The precision, recall, and F1 score achieved are 0.74, 0.71, and 0.718, respectively.

Next, PCA is applied to reduce the dimensions of input data. The FastText feature dimensions of 4401×100 are reduced to 4401×80. In addition, the

Cite this article as: A. N. Sutranggono and E. M. Imah, “Tweets Emotions Analysis of Community Activities Restriction as COVID-19 Policy in Indonesia Using Support Vector Machine”, CommIT Journal 17(1), 13–25, 2023.

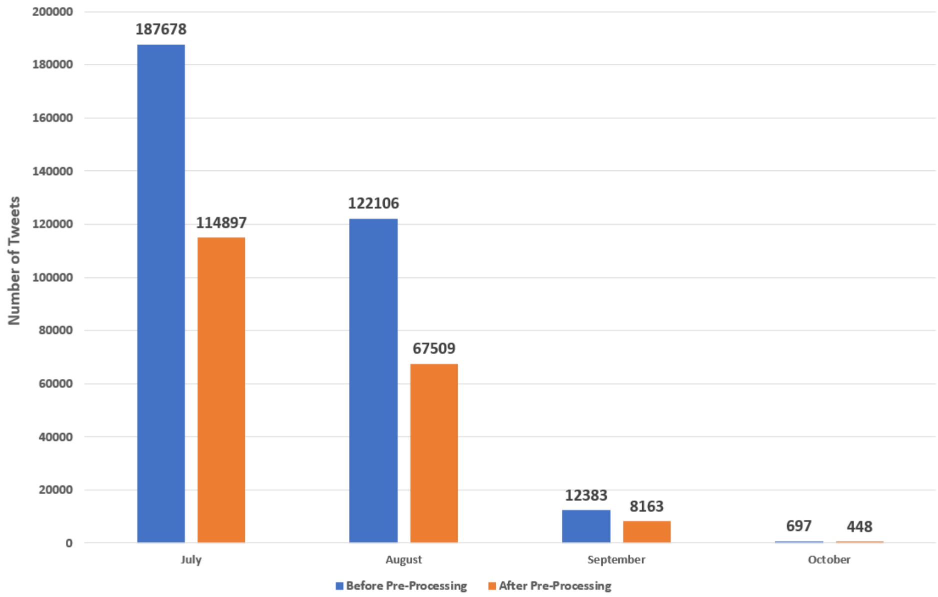

Fig. 5. Number of tweets per month on PPKM datasets.

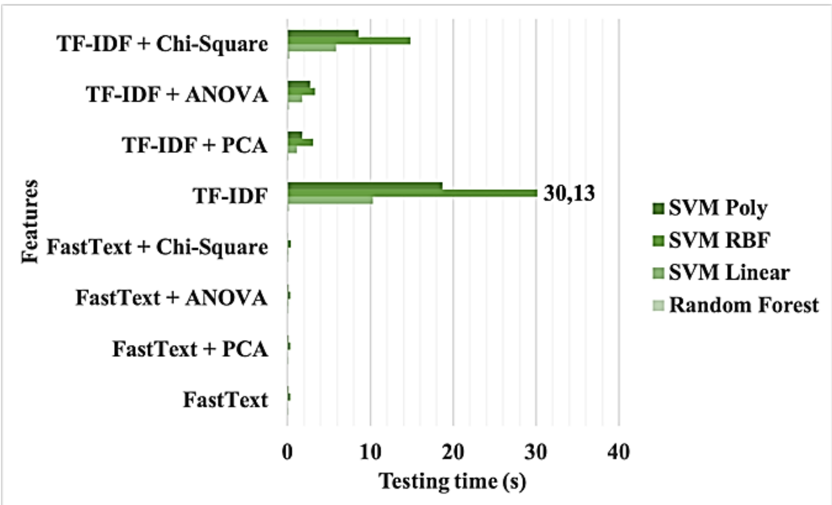

Fig. 6. Testing time of each classifier.

TF-IDF feature with dimensions of 4401×15386 is reduced to the size of 4401×2200. After applying PCA, the highest accuracy is obtained using the SVM RBF classifier built with the FastText feature. The value is 0.7052. Furthermore, that classifier has precision, recall, and F1 score values of 0.746, 0.6980, and 0.7120, respectively. On the other hand, PCs used by the SVM-RBF kernel function using the FastText feature contain 94,24% of input data characteristics.

The dimension of TF-IDF features is reduced to 4401×2500 using the ANOVA feature selection method. On the other hand, the dimensions of the FastText feature are decreased to 4401×75. The SVM classifier achieves the highest accuracy after applying the ANOVA feature selection technique with linear kernel function, compiled using the TF-IDF feature at 0.7506. The classifier’s precision, recall, and F1 score are 0.772, 0.7480, and 0.7540, respectively. In particular, if the dimensions of each feature are reduced again using ANOVA feature selection, the accuracy attained is no better than 0.7506.

Then, the Chi-Square feature selection is implemented to minimize the dimensions of the FastText feature to 4401×70. Meanwhile, the dimensions of

Cite this article as: A. N. Sutranggono and E. M. Imah, “Tweets Emotions Analysis of Community Activities Restriction as COVID-19 Policy in Indonesia Using Support Vector Machine”, CommIT Journal 17(1), 13–25, 2023.

TABLE III
THE RESULTS OF EACH CLASSIFIER EVALUATION IN CLASSIFYING EMOTIONS.

Classifier	Feature	Accuracy	Precision	Recall	F1 Score	Classifier	Feature	Accuracy	Precision	Recall	F1 Score
Random Forest	FastText	0.5737	0.5860	0.5680	0.5680	SVM Linear	FastText	0.6440	0.6600	0.6460	0.6520
	TF-IDF	0.6803	0.7040	0.6820	0.6880		TF-IDF	0.7052	0.7400	0.7100	0.7180
	FastText + PCA	0.6236	0.6700	0.5980	0.6160		FastText + PCA	0.6599	0.6840	0.6600	0.6680
	TF-IDF + PCA	0.5125	0.5820	0.4980	0.5200		TF-IDF + PCA	0.6848	0.7260	0.6880	0.7000
	FastText + ANOVA	0.5510	0.5780	0.5540	0.5440		FastText + ANOVA	0.5986	0.5980	0.6020	0.5920
	TF-IDF + ANOVA	0.6871	0.7200	0.6920	0.7020		TF-IDF + ANOVA	0.7506	0.7720	0.7480	0.7540
	FastText + Chi-Square	0.5896	0.5880	0.5700	0.5720		FastText + Chi-Square	0.6145	0.6660	0.6000	0.6180
	TF-IDF + Chi-Square	0.7120	0.7420	0.7280	0.7300		TF-IDF + Chi-Square	0.7528	0.7800	0.7580	0.7660
SVM RBF	FastText	0.6463	0.6680	0.6520	0.6540	SVM Poly	FastText	0.6599	0.6760	0.6640	0.6640
	TF-IDF	0.6576	0.7020	0.6480	0.6580		TF-IDF	0.3356	0.6340	0.2640	0.2100
	FastText + PCA	0.7052	0.7460	0.6980	0.7120		FastText + PCA	0.6780	0.7160	0.6700	0.6780
	TF-IDF + PCA	0.6961	0.7520	0.6800	0.7020		TF-IDF + PCA	0.4580	0.6540	0.3820	0.3640
	FastText + ANOVA	0.6190	0.6260	0.6300	0.6180		FastText + ANOVA	0.6259	0.6300	0.6320	0.6240
	TF-IDF + ANOVA	0.7392	0.7680	0.7320	0.7460		TF-IDF + ANOVA	0.5760	0.6820	0.5340	0.5280
	FastText + Chi-Square	0.6508	0.6740	0.6480	0.6560		FastText + Chi-Square	0.5780	0.5780	0.5860	0.5800
	TF-IDF + Chi-Square	0.7324	0.7740	0.7280	0.7440		TF-IDF + Chi-Square	0.4762	0.6740	0.3940	0.3700

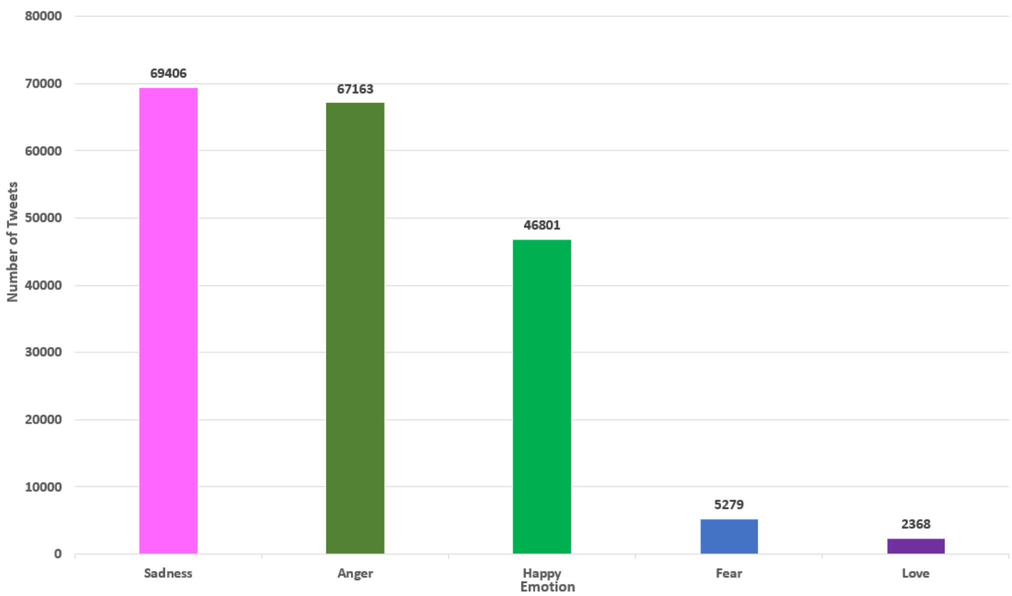

Fig. 7. Emotions classification results on tweets related to the PPKM.

the TF-IDF feature have been reduced to 4401×9000. After implementing Chi-Square, the highest accuracy value is attained by the SVM classifier with linear kernel function and TF-IDF feature, which is 0.7528. On the other hand, the classifier’s precision, recall, and F1 score values are consecutively 0.7800, 0.7580, and 0.7660. If the dimensions of each feature are reduced again, the results obtained are not optimal.

Experiments that have been carried out show that the performance of the SVM with linear kernel function designed using the TF-IDF feature outperforms other classifiers after applying the Chi-Square feature selection method. Although the classifier takes a long time to classify emotions, the accuracy is quite good, 0.7528. As a result, combining TF-IDF and Chi-Square feature selection methods and selecting the linear kernel functions of the SVM algorithm are the best approach for improving the performance of the SVM classifier in identifying emotions from the text. The method proposed in the research can provide better accuracy than previous studies [25, 51, 52]. The previous study combines the Part of Speech (POS) tagger method, Chi-Square feature selection, and SVM to identify emotions in text. This combination yields the best accuracy result of 0.7243 [25]. On the other hand, combining the Bag-of-Words (BoW), TF-IDF, and SVM methods achieves an accuracy of 0.7260 in classifying emotions in English texts [51]. Previous research has an accuracy of 0.4734 when performing an emotional analysis on a tweet using the Naïve Bayes method [52].

Emotions analysis on the PPKM dataset is conducted using the best classifier. Figure 7 shows the emotions classification results on the PPKM dataset. It can be seen that the sadness emotion classes have the most tweets, with 69,406 tweets classified in the sadness class out of the 191,017 tweets related to the PPKM dataset. In addition, the number of tweets from anger emotion classes is almost the same as in sadness classes. The proportion of tweets classified as sadness class is 36.33%, and anger is 35.16%.

IV. Conclusion

The research analyzes the emotional condition of the Indonesians during the PPKM policy during the COVID-19 pandemic. It is based on Indonesian language text data from Twitter uploaded when the PPKM policy is implemented. The research uses a classification algorithm based on machine learning: SVM and random forest. Text-based emotion detection, on the other hand, is a challenging task. Dealing with feature dimensions is one of the most difficult parts of classifying text using a machine learning-based approach. As a result, the research uses a combination of methods to achieve better results.

The results show that the TF-IDF method is the best method for representing Twitter text data in numerical form based on the experimental results. The accuracy of the classifier designed using the TF-IDF feature is overall better than the FastText feature. However, when viewed in terms of the speed of the classifier in analyzing emotions, the execution time of the classifier with the TF-IDF feature is longer than the FastText feature as the dimensions of the TF-IDF feature are larger. Next, classifier performance optimization is carried out through a dimension reduction process. The dimension reduction method is the PCA as the feature extraction method and the Chi-Square and ANOVA as the feature selection methods. After implementing the dimension reduction process, several classifiers' execution time or testing time in classifying an emotion becomes faster.

On the other hand, after applying the Chi-Square feature selection technique, the performance of the SVM classifier with linear kernel functions using the TF-IDF feature outperforms the different classifiers. The accuracy obtained is 0.7528. This accuracy value is better than the accuracy obtained by previous studies. The accuracy of similar previous studies is in the range of 0.4734 to 0.7260. Combining the TF-IDF method, Chi-Square feature selection, and linear kernel function selection in the SVM algorithm can improve SVM's performance in identifying text-based emotion.

In addition, on the classification of emotions on Twitter text data related to the PPKM policy, the Indonesian people's tweets on Twitter during the PPKM policy belong to the emotion classes: sadness at 36.33% and anger at 35.16%. Suppose the results of the classification are interpreted. In that case, it can be understood that most Indonesian people are annoyed with the implementation of the PPKM by the government.

Limitations in text-based emotion identification depend on the characteristics of the data to be tested. Several factors related to the text data analyzed can limit text-based emotion identification. The length and complexity of the text can impact the result of emotion identification accuracy. Short texts, such as tweets or headlines, may not contain enough information to capture the writers' intended emotions accurately. In contrast, longer texts may contain multiple emotions or conflicting sentiments that are difficult to disentangle. In addition, using slang, colloquialisms, and regional dialects can also pose a challenge for emotion identification methods. Cultural and contextual factors can also impact the accuracy of emotion identification. Different cultures may have different norms and conventions for expressing emotions, which a generic emotion identification model may not capture. Text data that are incomplete, ambiguous, or poorly written also can be complex for automated methods to interpret accurately. The research can be used as a reference for further research by implementing a dimension reduction stage to minimize irrelevant features. Therefore, the classifier's performance in identifying text-based emotion can become more optimal.

References

[1] D. Cucinotta and M. Vanelli, “WHO declares COVID-19 a pandemic,” *Acta Bio Medica: Atenei*

Parmensis, vol. 91, no. 1, pp. 157–160, 2020.

[2] Menteri Dalam Negeri Republik Indonesia, “Instruksi Menteri Dalam Negeri Nomor 15 Tahun 2021,” 2021. [Online]. Available: https://bit.ly/3lZlTmK

[3] Z. Drus and H. Khalid, “Sentiment analysis in social media and its application: Systematic literature review,” *Procedia Computer Science*, vol. 161, pp. 707–714, 2019.

[4] P. Monachesi and S. Witteborn, “Building the sustainable city through Twitter: Creative skilled migrants and innovative technology use,” *Telematics and Informatics*, vol. 58, pp. 1–10, 2021.

[5] K. Garcia and L. Berton, “Topic detection and sentiment analysis in Twitter content related to COVID-19 from Brazil and the USA,” *Applied Soft Computing*, vol. 101, pp. 1–15, 2021.

[6] M. Baali and N. Ghneim, “Emotion analysis of Arabic tweets using deep learning approach,” *Journal of Big Data*, vol. 6, no. 1, pp. 1–12, 2019.

[7] S. Kusal, S. Patil, K. Kotecha, R. Aluvalu, and V. Varadarajan, “AI based emotion detection for textual big data: Techniques and contribution,” *Big Data and Cognitive Computing*, vol. 5, no. 3, pp. 1–45, 2021.

[8] F. A. Acheampong, C. Wenyu, and H. Nunoo-Mensah, “Text-based emotion detection: Advances, challenges, and opportunities,” *Engineering Reports*, vol. 2, no. 7, pp. 1–24, 2020.

[9] M. Z. Asghar, F. Subhan, M. Imran, F. M. Kundi, S. Shamshirband, A. Mosavi, P. Csiba, and A. R. Várkonyi-Kóczy, “Performance evaluation of supervised machine learning techniques for efficient detection of emotions from online content,” *Preprint*, pp. 1–30, 2019.

[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, pp. 1–19, 2021.

[11] R. Jayakrishnan, G. N. Gopal, and M. S. Santhikrishna, “Multi-class emotion detection and annotation in Malayalam novels,” in *2018 International Conference on Computer Communication and Informatics (ICCCI)*. IEEE, 2018, pp. 1–5.

[12] P. Saigal and V. Khanna, “Multi-category news classification using support vector machine based classifiers,” *SN Applied Sciences*, vol. 2, no. 3, pp. 1–12, 2020.

[13] L. De Bruyne, O. De Clercq, and V. Hoste, “Lt3 at semeval-2018 task 1: A classifier chain to detect emotions in tweets,” in *Proceedings of the 12th International Workshop on Semantic Evaluation*, 2018, pp. 123–127.

[14] M. Suhasini and B. Srinivasu, “Emotion detection framework for Twitter data using supervised classifiers,” in *Data engineering and communication technology. Advances in intelligent systems and computing*. Springer, 2020.

[15] J. Carrillo-de Albornoz, J. Rodriguez Vidal, and L. Plaza, “Feature engineering for sentiment analysis in e-health forums,” *PLOS ONE*, vol. 13, no. 11, pp. 1–25, 2018.

[16] A. Jaffe, Y. Kluger, O. Lindenbaum, J. Patsenker, E. Peterfreund, and S. Steinerberger, “The spectral underpinning of word2vec,” *Frontiers in Applied Mathematics and Statistics*, vol. 6, pp. 1–10, 2020.

[17] N. Kulkarni, R. Vaidya, and M. Bhate, “A comparative study of word embedding techniques to extract features from text,” *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, vol. 12, no. 12, pp. 3550–3557, 2021.

[18] O. Kwon, D. Kim, S. R. Lee, J. Choi, and S. Lee, “Handling out-of-vocabulary problem in hangeul word embeddings,” in *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, 2021, pp. 3213–3221.

[19] M. Chiny, M. Chihab, O. Bencharef, and Y. Chihab, “LSTM, VADER and TF-IDF based hybrid sentiment analysis model,” *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 7, pp. 265–275, 2021.

[20] D. Endalie and T. Tegegne, “Designing a hybrid dimension reduction for improving the performance of Amharic news document classification,” *PLOS ONE*, vol. 16, no. 5, pp. 1–14, 2021.

[21] D. A. Eisa, A. I. Taloba, and S. S. I. Ismail, “A comparative study on using principle component analysis with different text classifiers,” *International Journal of Computer Applications*, vol. 180, no. 31, pp. 1–6, 2018.

[22] U. A. M. E. Ali, M. A. Hossain, and M. R. Islam, “Analysis of PCA based feature extraction methods for classification of hyperspectral image,” in *2019 2nd International Conference on Innovation in Engineering and Technology (ICIET)*. IEEE, 2019, pp. 1–6.

[23] M. Alassaf and A. M. Qamar, “Improving sentiment analysis of Arabic tweets by One-Way ANOVA,” *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 6, pp. 2849–2859, 2022.

[24] S. Bahassine, A. Madani, M. Al-Sarem, and M. Kissi, “Feature selection using an improved Chi-square for Arabic text classification,” *Journal*

of King Saud University-Computer and Information Sciences, vol. 32, no. 2, pp. 225–231, 2020.

[25] L. Singh, S. Singh, and N. Aggarwal, “Two-stage text feature selection method for human emotion recognition,” in *Proceedings of 2nd International Conference on Communication, Computing and Networking: ICCCN 2018, NITTTR Chandigarh, India*. Springer, 2018, pp. 531–538.

[26] C. H. Cheng and H. H. Chen, “Sentimental text mining based on an additional features method for text classification,” *PLOS ONE*, vol. 14, no. 6, pp. 1–17, 2019.

[27] A. R. Murthy and K. A. Kumar, “A review of different approaches for detecting emotion from text,” in *IOP Conference Series: Materials Science and Engineering*, vol. 1110. IOP Publishing, 2021, pp. 1–23.

[28] N. Alswaidan and M. E. B. Menai, “A survey of state-of-the-art approaches for emotion recognition in text,” *Knowledge and Information Systems*, vol. 62, pp. 2937–2987, 2020.

[29] M. S. Saputri, R. Mahendra, and M. Adriani, “Emotion classification on Indonesian Twitter dataset,” in *2018 International Conference on Asian Language Processing (IALP)*. IEEE, 2018, pp. 90–95.

[30] E. Cambria, A. Livingstone, and A. Hussain, “The hourglass of emotions,” in *Cognitive behavioural systems. Lecture notes in computer science*. Springer, 2012.

[31] P. R. Shaver, U. Murdaya, and R. C. Fraley, “Structure of the Indonesian emotion lexicon,” *Asian Journal of Social Psychology*, vol. 4, no. 3, pp. 201–224, 2001.

[32] I. Dongo, Y. Cadinale, A. Aguilera, F. Martínez, Y. Quintero, and S. Barrios, “Web scraping versus Twitter API: A comparison for a credibility analysis,” in *Proceedings of the 22nd International Conference on Information Integration and Web-Based Applications & Services*, 2020, pp. 263–273.

[33] K. M. Ridhwan and C. A. Hargreaves, “Leveraging Twitter data to understand public sentiment for the COVID-19 outbreak in Singapore,” *International Journal of Information Management Data Insights*, vol. 1, no. 2, pp. 1–15, 2021.

[34] B. Gaind, V. Syal, and S. Padgalwar, “Emotion detection and analysis on social media,” 2019. [Online]. Available: http://arxiv.org/abs/1901.08458

[35] H. T. Duong and T. A. Nguyen-Thi, “A review: Preprocessing techniques and data augmentation for sentiment analysis,” *Computational Social Networks*, vol. 8, no. 1, pp. 1–16, 2021.

[36] S. Sarica and J. Luo, “Stopwords in technical language processing,” *PLOS ONE*, vol. 16, no. 8, pp. 1–13, 2021.

[37] H. Young, “The digital language divide: How does the language you speak shape your experience of the Internet?” [Online]. Available: http://labs.theguardian.com/digital-language-divide/

[38] R. Chandra and A. Krishna, “COVID-19 sentiment analysis via deep learning during the rise of novel cases,” *PLOS ONE*, vol. 16, no. 8, pp. 1–26, 2021.

[39] Y. Li and T. Yang, “Word embedding for understanding natural language: A survey,” *Guide to Big Data Applications*, pp. 83–104, 2018.

[40] 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*, vol. 100, pp. 1–18, 2019.

[41] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, “Enriching word vectors with subword information,” *Transactions of the Association for Computational Linguistics*, vol. 5, pp. 135–146, 2017.

[42] A. Madasu and S. Elango, “Efficient feature selection techniques for sentiment analysis,” *Multimedia Tools and Applications*, vol. 79, pp. 6313–6335, 2020.

[43] E. J. Cushion, J. Warmenhoven, J. S. North, and D. J. Cleather, “Principal component analysis reveals the proximal to distal pattern in vertical jumping is governed by two functional degrees of freedom,” *Frontiers in Bioengineering and Biotechnology*, vol. 7, pp. 1–11, 2019.

[44] K. J. Johnson and R. E. Synovec, “Pattern recognition of jet fuels: Comprehensive GC$\times$ GC with ANOVA-based feature selection and principal component analysis,” *Chemometrics and Intelligent Laboratory Systems*, vol. 60, no. 1-2, pp. 225–237, 2002.

[45] G. Forman, “An extensive empirical study of feature selection metrics for text classification,” *Journal of Machine Learning Research*, vol. 3, pp. 1289–1305, 2003.

[46] S. Chatterjee, K. Chakrabarti, A. Garain, F. Schwenker, and R. Sarkar, “JUMRv1: A sentiment analysis dataset for movie recommendation,” *Applied Sciences*, vol. 11, no. 20, pp. 1–24, 2021.

[47] E. M. Imah, F. Al Afif, M. I. Fanany, W. Jatmiko, and T. Basaruddin, “A comparative study on Daubechies Wavelet Transformation, Kernel PCA and PCA as feature extractors for arrhythmia

detection using SVM,” in *TENCON 2011-2011 IEEE Region 10 Conference*. IEEE, 2011, pp. 5–9.

[48] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, pp. 5–32, 2001.

[49] V. Y. Kullarni and P. K. Sinha, “Random forest classifier: a survey and future research directions,” *International Journal of Advanced Computing*, vol. 36, no. 1, pp. 1144–1156, 2013.

[50] M. Allouch, A. Azaria, R. Azoulay, E. Ben-Izchak, M. Zwilling, and D. A. Zachor, “Automatic detection of insulting sentences in conversation,” in *2018 IEEE International Conference on the Science of Electrical Engineering in Israel (ICSEE)*. IEEE, 2018, pp. 1–4.

[51] M. Z. Ali, K. Javed, E. U. Haq, and A. Tariq, “Sentiment and emotion classification of epidemic related bilingual data from social media,” 2021. [Online]. Available: http://arxiv.org/abs/2105.01468

[52] K. Sailunaz and R. Alhajj, “Emotion and sentiment analysis from twitter text,” *Journal of Computational Science*, vol. 36, pp. 1–18, 2019.