Sentiment Analysis of Covid-19 Vaccination using Support Vector Machine in Indonesia

Majid Rahardi[1]*, Afrig Aminuddin[2], Ferian Fauzi Abdulloh[3], Rizky Adhi Nugroho[4]
Faculty of Computer Science, Universitas Amikom Yogyakarta, Yogyakarta, Indonesia

***Abstract*—Along with the development of the Covid-19 pandemic, many responses and news were shared through social media. The new Covid-19 vaccination promoted by the government has raised pros and cons from the public. Public resistance to covid-19 vaccination will lead to a higher fatality rate. This study carried out sentiment analysis about the Covid-19 vaccine using the Support Vector Machine (SVM). This research aims to study the public response to the acceptance of the vaccination program. The research result can be used to determine the direction of government policy. Data collection was taken via Twitter in the year 2021. The data then undergoes the preprocessing methods. Afterward, the data is processed using SVM classification. Finally, the result is evaluated by a confusion matrix. The experimental result shows that SVM produces 56.80% positive, 33.75% neutral, and 9.45% negative. The highest model accuracy was obtained by RBF kernel of 92%, linear and polynomial kernels obtained 90% accuracy, and sigmoid kernel obtained 89% accuracy.**

Keywords*—*Covid-19; vaccination; support vector machine; twitter

I. INTRODUCTION

World Health Organization (WHO) announced that the Covid-19 virus has spread to countries since 2019. WHO officially set Covid-19 as a global pandemic on March 11, 2020 [1][2]. Various steps were taken to overcome this pandemic. One of them is the manufacture of vaccines. In Indonesia, President Joko Widodo inaugurated Perpres Nomor 99 Tahun 2020 About Vaccine Procurement and Vaccination Implementation In the Context of Combating the Corona Virus Pandemic Disease 2019 (Covid-19) [3].

The pros and cons of vaccination have attracted various groups to express opinions. Social media is a medium that is easy and fast to access. So, it is not uncommon for people to express their views on social media. According to APJII, around 51.5% of internet users in Indonesia use social media daily [4]. One of the social media that is often used in Indonesia is Twitter. Twitter has 152 million registered users worldwide and more than 500 million unregistered users per month[5].

The number of opinions and the ease of accessing social media allow researchers to research cyberspace. One of these studies is sentiment analysis. Sentiment analysis is a classification process to classify the text in the document into positive, negative, and neutral classes [6][7]. The results from sentiment analysis can be used for various purposes [8].

Based on that problem, this research will compare the sentiment analysis results regarding the Indonesian people's point of view towards Covid-19 vaccination activities from Twitter. First, the data are taken from Twitter in Bahasa. Afterward, the data is preprocessed. Then, sentiment analysis is carried out using the Support Vector Machine.

*Corresponding Author.

II. LITERATURE REVIEW

A. Sentiment Analysis

The opinion is a point of view or attitude of humans to a situation, entity, and others. The statement of each individual has a subjective nature. So that it can provide different points of view [9]–[12]. Differences of opinion can then be investigated, giving rise to a generalization of the field of study in the form of sentiment analysis. Sentiment analysis is an area of study that studies opinions, sentiments, evaluations, judgments, attitudes, and emotions towards an entity such as an organization, event, person, and goods [13].

B. Preprocessing

Preprocessing is an activity that is carried out before further analysis of the data. In this process, entities and unnecessary information are removed. Preprocessing is done through several stages: cleaning, case folding, tokenizing, normalization, stopwords, and stemming[14]. Preprocessing aims to convert raw data into more structured data to be recognizable to the machine[15], [16].

C. TF-IDF

TF-IDF consists of two interrelated matrices, namely TF and IDF. TF or Term-Frequency is a matrix that counts the number of times a word appears in a document [17]. IDF or Inverse Document Frequency is a matrix that counts the frequency of a word that appears throughout the document [17]. The purpose of TF-IDF is to calculate the importance level of words in a document [18][19]. The equation in TF-IDF can be seen in (1).

$$TFIDF(t,d) = T(t,d) \text{ X } IDF(t) \quad (1)$$

t = the number of a term or word in the document.

d = the number of document.

D. Synthetic Minority Oversampling Technique

The total of each dataset class determines the level of validity and accuracy of a model. The quality of a good dataset can be obtained with consistency and good trustworthiness. The imbalanced class dataset can be overcome by using SMOTE [20]. This technique utilizes the concept of K-Nearest Neighbor, which then SMOTE plays a role in making synthetic data from minority classes [21]. A new minority sample is created using this method by linear interpolation between two minority samples [22].

E. Support Vector Machine

Support Vector Machine is a method of machine learning used for data classification. This method is included in supervised learning. Therefore, labeling is required in the data [23][24]. Support Vector Machine works to separate data by finding the best hyperplane and margin maximum [25]. The hyperplane is a plane separator or differentiator between two classes, while the margin is the distance between the outermost samples of the class or which is called the Support Vector. The equation of Support Vector Machine can be seen in (2) [24].

$$w.x - b = 0 \tag{2}$$

where *w* represents the weight vector, *x* denotes the input vector, and *b* indicates the bias.

F. Performance Evaluation Measure

Performance Evaluation Measure is a process for evaluating the performance or capability of a classification system or model [26]. Evaluations are displayed in a table with several entities called the confusion matrix. These entities include accuracy, precision, recall, and F1-Score.

Precision is the match value between the answers system and the user's information. The recall is the value of the accuracy of the unit of data with a previously called up team of information. Accuracy results from a comparison between correct information by the whole information. F1-Score is the value obtained from the weighted average between precision and recall. The equation of performance evaluation measure can be seen in (3), (4), (5), and (6) [27]–[29].

$$Precision = \frac{TP}{TP+FP} \tag{3}$$

$$Recall = \frac{TP}{TP+FN} \tag{4}$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{5}$$

$$F1\ Score = \frac{2(Precision)(Recall)}{(Precision+Recall)} \tag{6}$$

The table is known as the Confusion Matrix used to describe PEM. This table contains the results of the dataset test with the model that has been made in the prediction class and actual class. The table of Confusion Matrix can be seen in Table I [30].

True Positive indicates that the model previously predicted true proved true. True Negative suggests that the previous model was predicted wrong and proved wrong. False Positive shows that the previously predicted model true is proven false. False Negative indicates that the previously predicted model was proven wrong correct [27]–[29].

TABLE I. CONFUSION MATRIX

	Actual True	**Actual False**
True Predicted	True Positive (TP)	False Positive (FP)
False Predicted	False Negative (FN)	True Negative (TN)

III. METHODS

A. System Overview

The system takes data through Twitter in Bahasa Indonesia with a total of 2000 tweets and the timeline between October 19, 2021, and October 22, 2021. After that, the data is manually labeled after cleaning the tweet. Then, do the preprocessing to clean up data noise and simplify the classification process. After that, word weighting was done using the TF-IDF method and sentiment dataset class balancing using the SMOTE method. Then, divide the dataset into data training and data testing. Next, classification sentiment is done using the Support Vector Machine and Evaluating the value indicated through Confusion Matrix. The system of this sentiment analysis process can be seen in Fig. 1.

B. Data Collection and Cleaning

Data collection is done through Twitter using the Twitter API. Twitter API consists of Access Token, Access Token Secret, API Key, and API Key Secret. This process is written in Python using Tweepy library. Researchers use the keywords "vaksin" and "vaksinasi" as a reference for searching and used in a CSV file or Comma Separated Values. Results of the Crawling Tweets process can be seen in Table II.

The next process that needs to be done is cleaning to simplify the labeling process. Also, removes URLs, punctuation, and numeric, including in the cleaning process. In addition, eliminating mentions of users by @ and deleting duplicate tweets are included in the cleaning process. The results of the cleaning process can be seen in Table III.

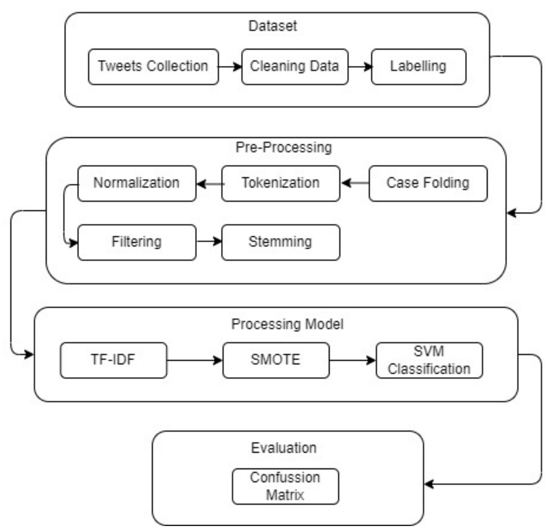

Fig. 1. System Overview of Sentiment Analysis.

TABLE II. RESULTS OF THE CRAWLING TWEETS

Date and Time	**User**	**Tweet**
2021-10-19 22:58:43	COVID PASTI BERLALU	Para pelajar menyampaikan rasa terima kasih atas program vaksinasi yang terus dilakukan oleh pemerintah. #AyoVaksin #KaltaraSuksesVaksinasi #VaksinasiPulihkanRI #IndonesiaSehat #IndonesiaHebat https://t.co/h5x41UO3tF

TABLE III. RESULTS OF THE CLEANING PROCESS

Tweet before Cleaning	Tweet After Cleaning
Para pelajar menyampaikan rasa terima kasih atas program vaksinasi yang terus dilakukan oleh pemerintah. #AyoVaksin #KaltaraSuksesVaksinasi #VaksinasiPulihkanRI #IndonesiaSehat #IndonesiaHebat https://t.co/h5x41UO3tF	Para pelajar menyampaikan rasa terima kasih atas program vaksinasi massal yang terus dilakukan oleh pemerintah

C. Data Labelling

The labeling process needs to be done before the sentiment analysis process. This process is done manually conducted by researchers with several parameters. The parameter used is the positive sentiment with the number two symbol, a neutral sentiment with the number symbol one, and negative sentiment with zero. Results of the labeling process can be seen in Table IV.

TABLE IV. RESULTS OF THE LABELLING PROCESS

Tweet after Cleaning	Label
Para pelajar menyampaikan rasa terima kasih atas program vaksinasi massal yang terus dilakukan oleh pemerintah	2

D. Data Preprocessing

This stage is one of the most important carried out in sentiment analysis. At this stage, data is cleaned of noise to form the data according to the standard. This process makes it easier for the system to recognize and analyze a word.

1) Case Folding: This process is changing each letter in the sentence to lowercase. The result of Case Folding process can be seen in Table V.

TABLE V. RESULTS OF CASE FOLDING PROCESS

Tweet before Case Folding	Tweet after Case Folding
Para pelajar menyampaikan rasa terima kasih atas program vaksinasi massal yang terus dilakukan oleh pemerintah	para pelajar menyampaikan rasa terima kasih atas program vaksinasi massal yang terus dilakukan oleh pemerintah

2) Tokenizing: This process is splitting sentences into a piece of the word called tokens. The result of the Tokenizing process can be seen in Table VI.

3) Normalization: This step is converting each word or token into a common phrase. The standard used in this case is the standard Indonesian word. This method was based on research published in the journal by Salsabila et al. [31]. The results of the Normalization process can be seen in Table VII.

4) Stopwords or filtering: This process removes words or tokens that are unnecessary and unrelated for further analysis. The result of the Stopwords process can be seen in Table VIII.

TABLE VI. RESULTS OF THE TOKENIZING PROCESS

Tweet before Tokenizing	Tweet after Tokenizing
para pelajar menyampaikan rasa terima kasih atas program vaksinasi massal yang terus dilakukan oleh pemerintah	[‘para’,’pelajar’,’menyampaikan’,’rasa’,’terima’,’kasih’,’atas’,’program’,’vaksinasi’,’massal’,’yang’,’terus’,’dilakukan’,’oleh’,’pemerintah’]

TABLE VII. RESULTS OF THE NORMALIZATION PROCESS

Tweet Tokens	Normalization
[‘betul’,’vaksinasi’,’itu’,’penting’,’sbg’,’benteng’,’pertahanan’,’diri’,’dr’,’dalam’]	[‘betul’,’vaksinasi’,’itu’,’penting’,’sebagai’,’benteng’,’pertahanan’,’diri’,’dari’,’dalam’]

TABLE VIII. RESULTS OF THE STOPWORDS PROCESS

Normalization	Stopwords
[‘para’,’pelajar’,’menyampaikan’,’rasa’,’terima’,’kasih’,’atas’,’program’,’vaksinasi’,’massal’,’yang’,’terus’,’dilakukan’,’oleh’,’pemerintah’]	[‘pelajar’,’menyampaikan’,’rasa’,’terima’,’kasih’,’atas’,’program’,’vaksinasi’,’massal’,’terus’,’dilakukan’]

5) Stemming: This process changes words or tokens into a word form base. The results of the stemming process can be seen in Table IX.

TABLE IX. RESULTS OF THE STEMMING PROCESS

Stopwords	Stemming
[‘pelajar’,’menyampaikan’,’rasa’,’terima’,’kasih’,’atas’,’program’,’vaksinasi’,’massal’,’terus’,’dilakukan’]	[‘ajar’,’sampai’,’rasa’,’terima’,’kasih’,’atas’,’program’,’vaksinasi’,’massal’,’terus’,’laku’]

E. Feature Extraction

Feature extraction or TF-IDF will calculate each word by the value against the number of documents used at this stage. This process aims to carry out the classification process easily and recognize words in the form of numbers. The results of the implementation of TF-IDF are as follows:

- Document 1: Ayo dukung program vaksinasi masal capai target herd immunity nasional untuk Indonesia sehat.
- Document 2: Mau sehat Ayo vaksinasi Indonesia bangkit.

The calculation of TF-IDF can be seen in Table X.

TABLE X. RESULTS OF TF-IDF PROCESS

Word	TF		IDF	TF-IDF Score	
	D1	D2		D1	D2
Ayo	$\frac{1}{13}$	$\frac{1}{13}$	$\log\left(\frac{2}{2+1}\right) = 0.18$	0.013	0.013
Dukung	$\frac{1}{13}$	0	$\log\left(\frac{2}{1+1}\right) = 0$	0	0
Program	$\frac{1}{13}$	0	$\log\left(\frac{2}{1+1}\right) = 0$	0	0
Vaksinasi	$\frac{1}{13}$	$\frac{1}{13}$	$\log\left(\frac{2}{2+1}\right) = 0.18$	0.013	0.013
Masal	$\frac{1}{13}$	0	$\log\left(\frac{2}{1+1}\right) = 0$	0	0
Capi	$\frac{1}{13}$	0	$\log\left(\frac{2}{1+1}\right) = 0$	0	0
Target	$\frac{1}{13}$	0	$\log\left(\frac{2}{1+1}\right) = 0$	0	0
Herd	$\frac{1}{13}$	0	$\log\left(\frac{2}{1+1}\right) = 0$	0	0

Immunity	$\frac{1}{13}$	0	$\log\left(\frac{2}{1+1}\right) = 0$	0	0
Nasional	$\frac{1}{13}$	0	$\log\left(\frac{2}{1+1}\right) = 0$	0	0
Untuk	$\frac{1}{13}$	0	$\log\left(\frac{2}{1+1}\right) = 0$	0	0
Indonesia	$\frac{1}{13}$	$\frac{1}{13}$	$\log\left(\frac{2}{2+1}\right) = 0.18$	0.013	0.013
Sehat	$\frac{1}{13}$	$\frac{1}{13}$	$\log\left(\frac{2}{2+1}\right) = 0.18$	0.013	0.013
Mau	0	$\frac{1}{13}$	$\log\left(\frac{2}{1+1}\right) = 0$	0	0
Bangkit	0	$\frac{1}{13}$	$\log\left(\frac{2}{1+1}\right) = 0$	0	0

F. *Synthetic Minority Oversampling Technique*

SMOTE will balance the dataset by taking advantage of the TF-IDF value. Balancing dataset class is a must because there is a gap in data between positive, neutral, and negative classes. SMOTE will create class neutral and negative belonging to the minority class synthetic data to match the majority class, positive class. Before and after the implementation of SMOTE can be seen in Fig. 2 and 3.

```
[ ] df["Label"].value_counts()

    2    1136
    1     675
    0     189
    Name: Label, dtype: int64
```

Fig. 2. Before Implementation of SMOTE.

```
[52] from imblearn.over_sampling import SMOTE
     sm_combine = SMOTE(random_state=10)
     x_vect, y = sm_combine.fit_resample(x_vect,df['Label'])
     y.value_counts()

     1    1136
     2    1136
     0    1136
     Name: Label, dtype: int64
```

Fig. 3. After Implementation of SMOTE.

G. *Support Vector Machine*

This research tested the model using four Support Vector Machine kernels, i.e., linear, polynomial, RBF, and sigmoid. Various kernels are utilized to change the level of data dimensions to be different depending on the kernel used. The equation of each Support Vector Machine kernel can be seen in Table XI [32].

TABLE XI. SUPPORT VECTOR MACHINE KERNEL EQUATION

Kernel	**Equation**	
Linear	$K(x_i, x) = x_i^T x$	
Polynomial	$K(x_i, x) = (\gamma . x_i^T x + r)^p$	p = degree r = coef()
RBF	$K(x_i, x) = exp(-\gamma \lvert x_i^T x \rvert^2)$	γ = gamma
Sigmoid	$K(x_i, x) = tanh(\gamma . x_i^T x + r)$	r = coef()

IV. RESULT

A. *Sentiment Analysis*

Researchers analyze sentiment by manually labeling each tweet. The labeled sentiment is done after the cleaning process. As a result, there are 1136 tweets (56.80%) with positive sentiment, 675 tweets (33.75%) with the neutral sentiment, and 189 (9.45%) tweets with negative sentiment. In the case of the imbalanced dataset, researchers implement SMOTE to balance the minority class and majority class.

B. *Performance Evaluation Measure*

The evaluation of the model in this research is presented in a confusion matrix table which consists of prediction class and actual class. The comparison of each kernel also becomes an evaluation parameter in this research.

1) Linear kernel: Parameters used in the test of the linear kernel, i.e., the complexity and maximum iterations. Increased complexity causes accuracy to decrease, and increasing the maximum iteration value tends to give a constant value with low volatility. The best accuracy result is 90% with C = 1 and max iteration = "default". The confusion matrix table of the linear kernel can be seen in Table XII, and the performance evaluation measure can be seen in Table XIII.

2) RBF Kernel: Parameters used in the test of RBF kernel, i.e., complexity and gamma. Testing the complexity value of more than five tends to produce high accuracy values constant, and increases in the value of gamma cause accuracy to decrease. The best accuracy result is 92% with C = 1 and gamma = 1. The confusion matrix table of the RBF kernel can be seen in Table XIV, and the performance evaluation measure can be seen in Table XV.

TABLE XII. CONFUSION MATRIX OF LINEAR KERNEL

Actual	**Predicted**		
	Negative	**Neutral**	**Positive**
Negative	88	0	0
Neutral	4	117	13
Positive	3	15	101

TABLE XIII. PERFORMANCE EVALUATION MEASURE OF LINEAR KERNEL

Sentiment	**Precision**	**Recall**	**F1-Score**
Negative	0.93	1.00	0.96
Neutral	0.86	0.87	0.88
Positive	0.89	0.85	0.87

TABLE XIV. CONFUSION MATRIX OF RBF KERNEL

Actual	**Predicted**		
	Negative	**Neutral**	**Positive**
Negative	88	0	0
Neutral	2	115	17
Positive	1	8	110

TABLE XV. PERFORMANCE EVALUATION MEASURE OF RBF KERNEL

Sentiment	Precision	Recall	F1-Score
Negative	0.97	1.00	0.98
Neutral	0.93	0.86	0.89
Positive	0.87	0.92	0.89

3) Polynomial Kernel: Parameters used in the test of Polynomial kernel, i.e., complexity, gamma, and degree. Testing on three parameters causes a decrease in accuracy as the value increases. The best accuracy result is 90% with C = 1, gamma = “scale”, and degree = 1. The confusion matrix table of the Polynomial kernel can be seen in Table XVI, and the performance evaluation measure can be seen in Table XVII.

TABLE XVI. CONFUSION MATRIX OF POLYNOMIAL KERNEL

Actual	Predicted		
	Negative	**Neutral**	**Positive**
Negative	88	0	0
Neutral	4	118	12
Positive	3	16	100

TABLE XVII. PERFORMANCE EVALUATION MEASURE OF POLYNOMIAL KERNEL

Sentiment	Precision	Recall	F1-Score
Negative	0.93	1.00	0.96
Neutral	0.88	0.88	0.88
Positive	0.89	0.84	0.87

4) Sigmoid Kernel: Parameters used in the test of the sigmoid kernel, i.e., complexity and gamma. Increasing the value of complexity and gamma produces an accuracy value that tends to decrease. The best accuracy result is 89%, with C = 1 and gamma = “scale”. The confusion matrix table of the sigmoid kernel can be seen in Table XVIII, and the performance evaluation measure can be seen in Table XIX.

TABLE XVIII. CONFUSION MATRIX OF SIGMOID KERNEL

Actual	Predicted		
	Negative	**Neutral**	**Positive**
Negative	88	0	0
Neutral	7	114	13
Positive	5	13	101

TABLE XIX. PERFORMANCE EVALUATION MEASURE OF SIGMOID KERNEL

Sentiment	Precision	Recall	F1-Score
Negative	0.88	1.00	0.94
Neutral	0.90	0.85	0.87
Positive	0.89	0.85	0.87

V. CONCLUSION

Sentiment class classification generates positive sentiment 1136 tweets (56.80%), neutral sentiment 675 tweets (33.75%), and negative sentiment 189 tweets (9.45%). The dataset class number gap is balanced through the SMOTE method, which generates 1136 tweets in all class sentiments. The classification results using the Support Vector Machine have the best accuracy result obtained on the RBF kernel by 92%, followed by Linear kernel and Polynomial kernel by 90%; lastly, the Sigmoid kernel by 89%.

REFERENCES

[1] Cucinotta and M. Vanelli, “WHO declares COVID-19 a pandemic,” Acta Biomed., vol. 91, no. 1, pp. 157–160, 2020, doi: 10.23750/abm.v91i1.9397.

[2] H. Chen et al., “Clinical characteristics and intrauterine vertical transmission potential of COVID-19 infection in nine pregnant women: a retrospective review of medical records,” Lancet, vol. 395, no. 10226, pp. 809–815, 2020, doi: 10.1016/S0140-6736(20)30360-3.

[3] Presiden Republik Indonesia, “Peraturan Presiden No. 99 Tahun 2020 tentang Pengadaan Vaksin dan Pelaksanaan Vaksinasi dalam Rangka Penanggulangan Pandemi Corona Virus Disease 2019 (Covid-19)],” Peratur. Pres., vol. 2019, no. 039471, pp. 1–13, 2020.

[4] Asosiasi Penyelenggara Jasa Internet Indonesia, “Laporan Survei Internet APJII 2019 – 2020,” Asos. Penyelenggara Jasa Internet Indonesia, 2020. https://apjii.or.id/survei.

[5] K. Garcia and L. Berton, “Topic detection and sentiment analysis in Twitter content related to COVID-19 from Brazil and the USA,” Appl. Soft Comput., vol. 101, p. 107057, 2021, doi: 10.1016/j.asoc.2020.107057.

[6] A. J. Nair, G. Veena, and A. Vinayak, “Comparative study of Twitter Sentiment on COVID - 19 Tweets,” Proc. - 5th Int. Conf. Comput. Methodol. Commun. ICCMC 2021, no. Iccmc, pp. 1773–1778, 2021, doi: 10.1109/ICCMC51019.2021.9418320.

[7] M. S. Akhtar, D. Gupta, A. Ekbal, and P. Bhattacharyya, “Feature selection and ensemble construction: A two-step method for aspect based sentiment analysis,” Knowledge-Based Syst., vol. 125, pp. 116–135, 2017, doi: 10.1016/j.knosys.2017.03.020.

[8] L. C. Cheng and S. L. Tsai, “Deep learning for automated sentiment analysis of social media,” Proc. 2019 IEEE/ACM Int. Conf. Adv. Soc. Networks Anal. Mining, ASONAM 2019, pp. 1001–1004, 2019, doi: 10.1145/3341161.3344821.

[9] M. Wongkar and A. Angdresey, “Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler: Twitter,” Proc. 2019 4th Int. Conf. Informatics Comput. ICIC 2019, pp. 1–5, 2019, doi: 10.1109/ICIC47613.2019.8985884.

[10] L. Kurniasari and A. Setyanto, “Sentiment Analysis using Recurrent Neural Network,” J. Phys. Conf. Ser., vol. 1471, no. 1, Mar. 2020, doi: 10.1088/1742-6596/1471/1/012018.

[11] E. Y. Sari, A. D. Wierfi, and A. Setyanto, “Sentiment Analysis of Customer Satisfaction on Transportation Network Company Using Naive Bayes Classifier,” 2019 Int. Conf. Comput. Eng. Network, Intell. Multimedia, CENIM 2019 - Proceeding, vol. 2019-November, Nov. 2019, doi: 10.1109/CENIM48368.2019.8973262.

[12] M. Rosanensi, M. Madani, R. T. P. Wanggono, A. Setyanto, A. A. Selameto, and S. N. Wahyuni, “Analysis sentiment and tourist response to rinjani mountain tour based on comments from photo upload in instagram,” Proc. - 2018 3rd Int. Conf. Inf. Technol. Inf. Syst. Electr. Eng. ICITISEE 2018, pp. 184–188, Jul. 2018, doi: 10.1109/ICITISEE.2018.8720960.

[13] B. Liu, “Sentiment Analysis and Opinion Mining,” Synth. Lect. Hum. Lang. Technol., vol. 5, no. 1, pp. 1–167, May 2012, doi: 10.2200/S00416ED1V01Y201204HLT016.

[14] J. H. Jaman and R. Abdulrohman, “Sentiment Analysis of Customers on Utilizing Online Motorcycle Taxi Service at Twitter with the Support Vector Machine,” ICECOS 2019 - 3rd Int. Conf. Electr. Eng. Comput.

Sci. Proceeding, pp. 231–234, 2019, doi: 10.1109/ICECOS47637.2019.8984483.

[15] U. Makhmudah, S. Bukhori, J. A. Putra, and B. A. B. Yudha, “Sentiment Analysis of Indonesian Homosexual Tweets Using Support Vector Machine Method,” Proc. - 2019 Int. Conf. Comput. Sci. Inf. Technol. Electr. Eng. ICOMITEE 2019, pp. 183–186, 2019, doi: 10.1109/ICOMITEE.2019.8920940.

[16] H. Hairani, M. Innuddin, and M. Rahardi, “Accuracy Enhancement of Correlated Naive Bayes Method by Using Correlation Feature Selection (CFS) for Health Data Classification,” 2020 3rd Int. Conf. Inf. Commun. Technol. ICOIACT 2020, pp. 51–55, Nov. 2020, doi: 10.1109/ICOIACT50329.2020.9332021.

[17] R. Hassan and M. R. Islam, “Impact of Sentiment Analysis in Fake Online Review Detection,” 2021 Int. Conf. Inf. Commun. Technol. Sustain. Dev. ICICT4SD 2021 - Proc., pp. 21–24, 2021, doi: 10.1109/ICICT4SD50815.2021.9396899.

[18] R. Ahuja, A. Chug, S. Kohli, S. Gupta, and P. Ahuja, “The impact of features extraction on the sentiment analysis,” Procedia Comput. Sci., vol. 152, pp. 341–348, 2019, doi: 10.1016/j.procs.2019.05.008.

[19] M. Rathi, A. Malik, D. Varshney, R. Sharma, and S. Mendiratta, “Sentiment Analysis of Tweets Using Machine Learning Approach,” in 2018 Eleventh International Conference on Contemporary Computing (IC3), Aug. 2018, pp. 1–3, doi: 10.1109/IC3.2018.8530517.

[20] G. Yang and L. Qicheng, “An over Sampling Method of Unbalanced Data Based on Ant Colony Clustering,” IEEE Access, vol. 9, pp. 130990–130996, 2021, doi: 10.1109/ACCESS.2021.3114443.

[21] E. dwi nurindah Sari, “Analisis Sentimen Nasabah Pada Layanan Perbankan Menggunakan Metode Regresi Logistik Biner , Naïve Bayes Classifier (NBC), dan Support Vector Machine (SVM),” J. Sains Dan Seni Its, vol. 8, no. 2, p. 177, 2019.

[22] F. Wang, J. Zhang, Y. Bu, and B. Chen, “Research on imbalanced data set preprocessing based on deep learning,” Proc. - 2021 Asia-Pacific Conf. Commun. Technol. Comput. Sci. ACCTCS 2021, pp. 75–79, 2021, doi: 10.1109/ACCTCS52002.2021.00023.

[23] K. R. Kavitha, A. Gopinath, and M. Gopi, “Applying improved SVM classifier for leukemia cancer classification using FCBF,” 2017 Int. Conf. Adv. Comput. Commun. Informatics, ICACCI 2017, vol. 2017-Janua, pp. 61–66, 2017, doi: 10.1109/ICACCI.2017.8125817.

[24] S. Zahoor and R. Rohilla, “Twitter Sentiment Analysis Using Machine Learning Algorithms: A Case Study,” Proc. - 2020 Int. Conf. Adv. Comput. Commun. Mater. ICACCM 2020, pp. 194–199, 2020, doi: 10.1109/ICACCM50413.2020.9213011.

[25] S. Rani and S. Bhatt, “Sentiment Analysis on twitter data using Machine Learning,” J. Xidian Univ., vol. 14, no. 12, pp. 1–4, Dec. 2020, doi: 10.37896/jxu14.12/039.

[26] A. Imron, “Analisis Sentimen Terhadap Tempat Wisata di Kabupaten Rembang Menggunakan Metode Naive Bayes Classifier,” pp. 10–13, 2019.

[27] K. Kusrini, E. T. Luthfi, M. Muqorobin, and R. W. Abdullah, “Comparison of naive bayes and K-NN method on tuition fee payment overdue prediction,” 2019 4th Int. Conf. Inf. Technol. Inf. Syst. Electr. Eng. ICITISEE 2019, vol. 6, pp. 125–130, 2019, doi: 10.1109/ICITISEE48480.2019.9003782.

[28] F. Ernawan, A. Aminuddin, D. Nincarean, M. F. A. Razak, and A. Firdaus, “Three Layer Authentications with a Spiral Block Mapping to Prove Authenticity in Medical Images,” Int. J. Adv. Comput. Sci. Appl., vol. 13, no. 4, pp. 211–223, 2022, doi: 10.14569/IJACSA.2022.0130425.

[29] A. Aminuddin and F. Ernawan, “AuSR1: Authentication and self-recovery using a new image inpainting technique with LSB shifting in fragile image watermarking,” J. King Saud Univ. - Comput. Inf. Sci., Feb. 2022, doi: 10.1016/J.JKSUCI.2022.02.009.

[30] A. C. Emcha, Widyawan, and T. B. Adji, “Quotation extraction from Indonesian online news,” 2019 Int. Conf. Inf. Commun. Technol. ICOIACT 2019, pp. 408–412, 2019, doi: 10.1109/ICOIACT46704.2019.8938558.

[31] N. Aliyah Salsabila, Y. Ardhito Winatmoko, A. Akbar Septiandri, and A. Jamal, “Colloquial Indonesian Lexicon,” Proc. 2018 Int. Conf. Asian Lang. Process. IALP 2018, pp. 226–229, 2019, doi: 10.1109/IALP.2018.8629151.

[32] H. C. Husada and A. S. Paramita, “Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM),” Teknika, vol. 10, no. 1, pp. 18–26, 2021, doi: 10.34148/teknika.v10i1.311.