Sentiment Analysis of Covid-19 Handling in Indonesia Based on Lexicon Weighting

Usman Ependi[1]*, Sabeli Aliya [2], Ari Wibowo[3]
Informatics Study Program, Universitas Bina Darma[1], [3]
Management Study Program, Universitas Bina Darma[2]
Palembang, Indonesia
u.ependi@binadarma.ac.id [1], sably@binadarma.ac.id [2], ariwibowo@students.binadarma.ac.id [3]

***Abstract*— Covid-19, a contagious disease, has been classified as a global pandemic. Indonesia, as one of the ASEAN countries, has taken various measures to combat the spread of this disease. One of the government's initiatives to tackle the pandemic is the PeduliLindungi application, through which the public provides feedback on government policies. However, analyzing and comprehending public opinions in a non-subjective manner poses a challenge in objectively evaluating government services. This study aims to address this issue by conducting a sentiment analysis of Covid-19 handling in Indonesia, using a lexicon-based weighting system that includes SentiStrengthID and InSet. The decision tree (DT) machine learning algorithm is utilized to evaluate the polarity results provided by the lexicon. The results indicate that the sentiment polarity towards Covid-19 handling in Indonesia is negative based on both SentiStrengthID and InSet weights. Evaluating machine learning performance with the SentiStrengthID lexicon, the DT-entropy and DT-gini models achieved an accuracy of 82% and 83%, respectively. Similarly, evaluating machine learning performance with the InSet lexicon, the DT-entropy and DT-gini models achieved an accuracy of 81% and 82%, respectively.**

Keywords*— *sentiment analysis, covid-19, lexicon weight, machine learning

I. INTRODUCTION

Coronavirus, or COVID-19, is an infectious disease caused by SARS-CoV-2 that primarily attacks the respiratory system and can lead to interstitial pneumonia and acute respiratory distress syndrome (ARDS) [1]. As of March 2022, there have been 435 million cases and 5.95 million deaths worldwide due to this virus, according to Google statistics. In ASEAN countries, Indonesia has reported 5.54 million cases and 148,000 deaths, while Malaysia has reported 3.42 million cases and 32,674 deaths, and Thailand has reported 2.89 million cases and 22,933 deaths.

The death tolls in ASEAN countries, particularly Indonesia, Malaysia, and Thailand, show a significant discrepancy in the ratio of cases to deaths for each country. The Indonesian government has taken various measures to address the spread of COVID-19, as reflected in the government's response through Presidential Decree No. 7 of 2020 and the 265 government regulations at the level of the president, ministry, or other institutions that have been implemented until March 2022.

Controlling and managing COVID-19 is crucial to address the pandemic. In Indonesia, the PeduliLindungi application is being used to monitor community activities, while Malaysia is using the MySejahtera application, and Thailand is using the MorChana application. This emphasizes the importance of handling COVID-19, particularly in Indonesia. However, it is not enough for the government to make efforts alone, and public perception of COVID-19 management needs to be taken into account.

Sentiment analysis can be used to assess public opinion. This process involves learning public opinion about an entity and can be used for opinion mining. Sentiment analysis allows for information mining from social media platforms, which are used for interaction and become a treasure trove of information [3]. The use of social media by the Indonesian government, such as the PeduliLindungi application, presents an opportunity for deeper information mining related to public sentiment regarding government policies.

However, one of the main challenges in sentiment analysis is understanding the words and phrases used in public opinion about COVID-19 management. Proper understanding is crucial to determine sentiment polarity accurately, which can be positive, neutral, or negative. Many sentiment analysis studies related to COVID-19 have been conducted, such as public sentiment towards vaccination, public opinion on vaccines, pros and cons of vaccines, and public sentiment regarding COVID-19 [4-7]. However, determining polarity in these studies is usually done manually and validated by language experts, which is subjective and influenced by personal knowledge.

The goal of this study is to perform sentiment analysis on the COVID-19 handling in Indonesia using a lexicon weighting approach. Lexicon, which refers to the storage of words in long-term memory related to grammar in composing phrases and sentences, contains information such as speech parts [8]. In the context of sentiment analysis, lexicon serves as a comparator in determining sentiment results through predetermined word weights. This research employs two types of lexicons: Lexicon SentiStrengthID [9] and InSet lexicon [10] to determine sentiment. The polarity classification is achieved using a decision tree machine learning algorithm. The data used to assess public sentiment regarding COVID-19 management is derived from reviews (opinions) on the PeduliLindungi application.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v12i1.1615, Copyright ©2023
Submitted : February 15, 2023, Revised : March 8, 2023, Accepted : March 10, 2023, Published : March 14, 2023

II. METHODS

The research is divided into six stages: data preparation, lexicon weighting, sentiment exploration, machine learning classification, and performance evaluation. These steps are outlined in Figure 1, and each method employed at every stage is explained below.

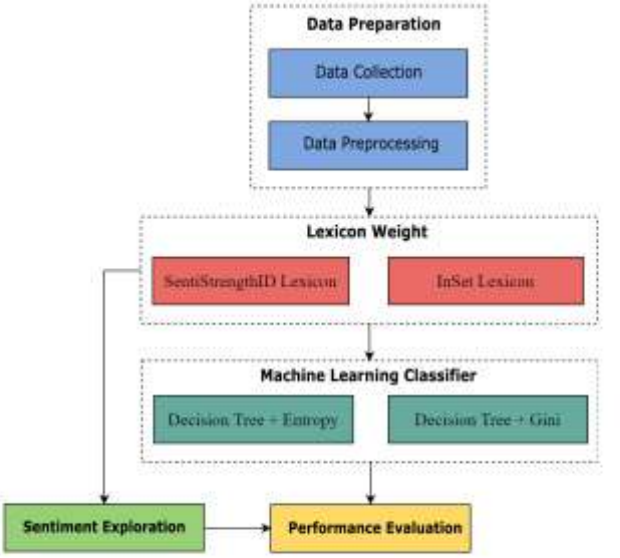


Fig. 1. Research Framework

A. Data Preparation

Data preparation is the initial process in conducting research. This stage consists of two tasks, namely data collection and data preprocessing. Here is an explanation of each stage.

1) Data Collection

We employed the scraping technique to collect data for this research. This technique is commonly used to gather data from social media or application reviews [11]. We collected opinion data from Indonesia's COVID-19 management application, PeduliLindungi, which is available on Google PlayStore. The application ID is com.telkom.tracencare. During the scraping process, we used data filters for the Indonesian language (id) and the newest comment category (review). The result of the scraping process yielded 52,680 raw data, with a focus on the "content" column.

2) Data Preprocessing

Data preprocessing involves preparing data to make it easier for computers to process and understand [12]. The data preprocessing techniques used in this research included tag and duplicate removal, case folding, stemming, stop words removal, and normalization. We utilized the Sastrawi library for stemming and stop words handling [13], while normalization utilized 17,320 normalized words. As illustrated in Figure 2, the data preprocessing phase resulted in 41,086 rows of data from the 52,680 raw data successfully crawled.

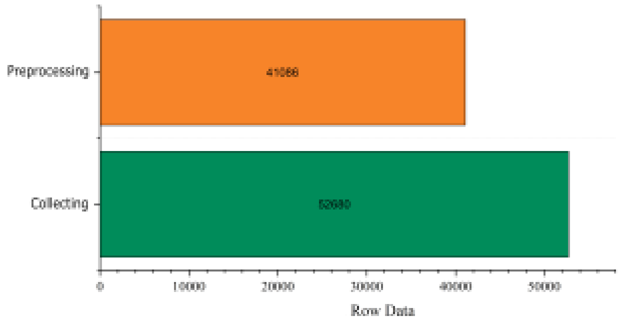


Fig. 2. Data Collection and Preprocessing Distribution

B. Lexicon Weight

We utilized two lexicons, namely SentiStrengthID (Table 1) [9] and InSet (Table 2) [10], for the lexicon weighting process. This was done by comparing the words in the lexicons with those in the row data and accumulating the weights to obtain a final score for each row data. The polarity of the row data was determined based on the condition 1. If the weighting result is greater than zero (> 0), it is classified as positive; if the weighting result is less than zero (< 0), it is classified as negative; and if the weighting result is equal to zero (= 0), it is classified as neutral.

TABLE I. LEXICON WORD WEIGHTING OF SENTISTRENGTHID

Number	Words	Weight
1	abadi	2
2	absen	-2
3	abu-abu	-2
4	acuh	-1
5	adil	2
...	...	...
1569	waspadalah	-2
1570	wow	3
1571	wtf	-3
1572	yakin	1
1573	yatim	-2

$$\text{sentiment}\begin{cases}\text{positive} = \text{bobot} > 0\\ \text{negative} = \text{bobot} < 0\\ \text{neutral} = \text{bobot} = 0\end{cases} \quad (1)$$

TABLE II. LEXICON WORD WEIGHTING OF INSET

Number	Words	Weight
1	putus tali gantung	-2
2	gelebah	-2
3	gobar hati	-2
4	tersentuh (perasaan)	-1
5	isak	-5
...	...	...
10214	melarikan	3
10215	vakansi	3
10216	spesial	4
10217	asrama	3
10218	orisinal	3

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v12i1.1615, Copyright ©2023
Submitted : February 15, 2023, Revised : March 8, 2023, Accepted : March 10, 2023, Published : March 14, 2023

C. *Sentiment Exploration*

We conducted sentiment exploration using exploratory data analysis (EDA). EDA is an evaluative process that aims to uncover insights from a given set of data [14]. It can also be seen as the art of reviewing a set of data that needs to be understood [15]. In this research, we utilized EDA to gain a better understanding of the public sentiment towards the management of COVID-19 through the PeduliLindungi application. EDA also helped us in understanding the sentiment weighting and data labeling.

D. *Machine Learning Classifier*

In this study, we employed the decision tree algorithm for machine learning classification. Decision tree is a diagrammatic structure where each node represents data in an attribute. The fundamental assumption of this classification is to identify similar object identifiers across variables [16]. Decision tree is a popular data mining technique for determining the decision of a case. This algorithm is capable of solving large cases and does not require previous knowledge management processes [17]. The decision tree is formed by separating nodes (criterion) based on one feature of the dataset with a set of if-then-else decision rules. Based on its capabilities, our decision tree is deemed suitable for sentiment polarity classification of Covid-19 management through the PeduliLindungi application.

We utilized decision tree machine learning classification with criteria consisting of Entropy and Gini Impurity. The feature extraction utilized TFIDF. The decision tree and criteria were developed using scikit-learn [18]. Entropy represents the sequence of randomness. In decision trees, it helps the model in selecting features for separation by measuring the purity of the separation at a node. The following provisions were applied:

"The value of Entropy = 0 indicates a pure separation, where all instances belong to the same class. On the other hand, when Entropy = 1, the separation is completely impure, which can happen, for instance, when the same occurrence has an equal percentage of both classes at the node, causing extreme disturbance. Equation (2) describes how to calculate the Entropy (Hi) value".

$$H_i = -\sum_{\substack{k=1 \\ p_{i,k} \neq 0}}^{n} p_{i,k} \log_2 (p_{i,k}) \tag{2}$$

P(i,k) represents the probability of positive and negative instances in class i at a specific node, where n is the number of distinct class values at that node. The entropy range H varies between 0 and 1. Gini Impurity also calculates the purity of the split in the decision tree node. Equation (3) provides the mathematical computation of the Gini attribute *(Gi)* at the *i-th* node.

$$G_i = 1 - \sum_{k=1}^{n} {p_{i,k}}^2 \tag{3}$$

p(i,k) is the ratio of class k among the training data within the node.

E. *Performance Evaluation*

Performance evaluation is the process of reviewing the results of a classification that has been performed. In this study, we used a confusion matrix to assess whether the classification results achieved good or poor performance. The confusion matrix includes the following provisions: true negative (TN), false positive (FP), true positive (TP), and false negative (FN) [19]. Based on these values, we evaluate performance using metrics such as accuracy, precision, recall, and F1-score. The calculation for each performance evaluation is shown in equations 4 to 7 [20].

$$\text{Accuracy} = \frac{(TP+TN)}{(TP + FP + FN + TN)} \tag{4}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{5}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{6}$$

$$\text{F-Measure} = 2 \text{ x } \frac{\text{Precision x Recall}}{\text{Precision + Recall}} \tag{7}$$

III. RESULTS AND DISCUSSION

According to the findings from the experiments that analyzed public opinions regarding the management of COVID-19 in Indonesia via the PeduliLindungi application, the following observations can be made.

A. *Sentiment*

The sentiment determination is greatly influenced by the words found in the opinion data row. Figure 3 illustrates the words that have a significant impact on the sentiment weight based on the words in the lexicon. There are discrepancies between the SentiStrengthID and InSet lexicons in terms of the frequency of word occurrences.

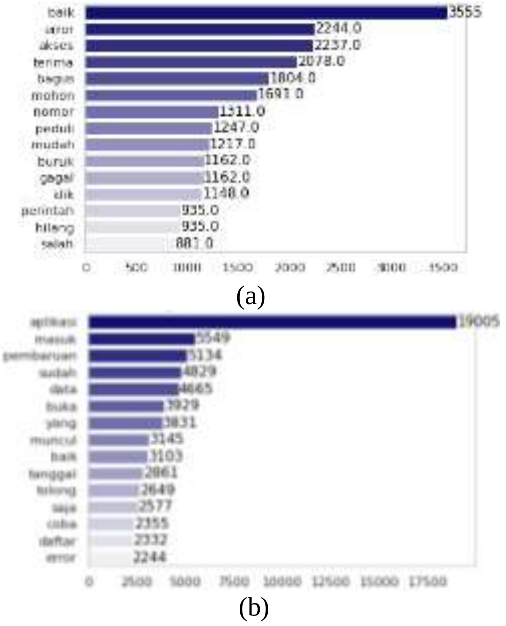


Fig. 3. Top fifteen frequently occurring words: (a) SentiStrengthID, (b) InSet

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v12i1.1615, Copyright ©2023
Submitted : February 15, 2023, Revised : March 8, 2023, Accepted : March 10, 2023, Published : March 14, 2023

Figure 3 reveals that the word frequency greatly affects the sentiment determination for each lexicon. The SentiStrengthID lexicon is dominated by the word "baik" with a frequency of 3555 times, while for the InSet lexicon, the most frequent word is "aplikasi" with a frequency of 19005 times. Such a difference has a significant impact on the sentiment weight assigned to each lexicon, as illustrated in Figure 4. The sentiment weighting results of each lexicon differ significantly. For instance, in the first row, both lexicons assign a negative weight, but with different magnitudes. On the other hand, the differences in the second and third rows are quite contrasting.

	text	sentiment
0	aplikasi mengizinkan pendaftaran keliru menyelesaikan bidang selesai bidang disorot mencoba nomo...	-6.0
1	mantap biar pandemi	0.0
2	perubahan yang esensial versi tolong diwajibkan pembaharuan perangkat lunak	2.0
3	pembaruan handphone instal dan instal ulang nomor handphone aplikasi verifikasi diperbarui mener...	3.0
4	aplikasi tidak	0.0
5	hasil cari negatif aplikasinya hitam tunggu isoman mohon diperbaiki konfirmasi security wajib scan	0.0
6	tidak jelas pembaruan mulu	2.0
7	lihat sertifikat vaksin	0.0
8	aplikasi bagus	2.0
9	surem user not found versi memasukkan no handphone langsung login daftar nik nama sesuai kartu t...	0.0

(a)

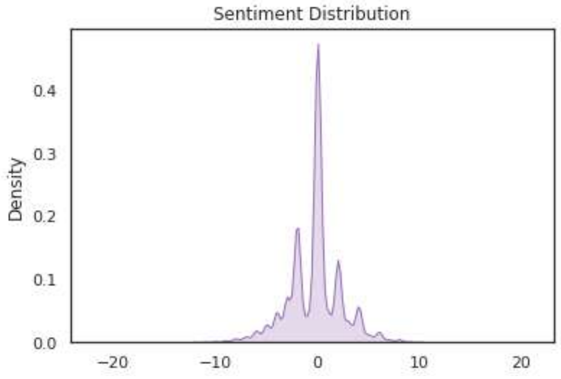
	text	sentiment
0	aplikasi mengizinkan pendaftaran keliru menyelesaikan bidang selesai bidang disorot mencoba nomo...	-22
1	mantap biar pandemi	-5
2	perubahan yang esensial versi tolong diwajibkan pembaharuan perangkat lunak	-4
3	pembaruan handphone instal dan instal ulang nomor handphone aplikasi verifikasi diperbarui mener...	3
4	aplikasi tidak	4
5	hasil cari negatif aplikasinya hitam tunggu isoman mohon diperbaiki konfirmasi security wajib scan	-12
6	tidak jelas pembaruan mulu	6
7	lihat sertifikat vaksin	3
8	aplikasi bagus	-6
9	surem user not found versi memasukkan no handphone langsung login daftar nik nama sesuai kartu t...	-2

(b)

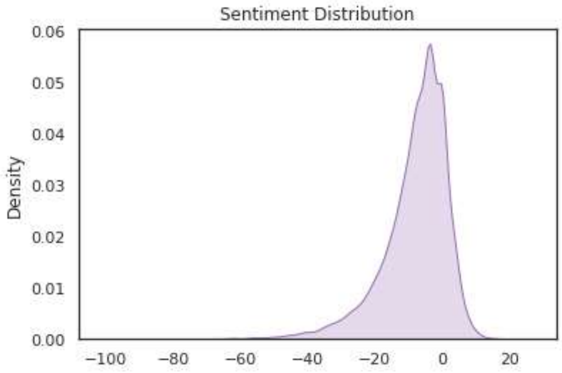
Fig. 4. Lexicon weighting results: (a) SentiStrengthID, (b) InSet

The study found that the sentiment weighting results, which are illustrated in Figure 4, had an impact on the distribution of sentiment. Specifically, the use of different sentiment lexicons led to different patterns in the sequence of neutral, negative, and positive sentiment weights. When the SentiStrengthID lexicon was used, the weights were often neutral, negative, and positive in sequence. In contrast, when the InSet lexicon was used, the weights were more likely to be negative, followed by positive and neutral weights. This difference in sentiment sequence had a significant impact on the overall sentiment distribution, as seen in Figure 5.

Moreover, the type of lexicon used also had a significant influence on the weights themselves, as shown in Figure 6. The SentiStrengthID lexicon generally provided negative weights ranging from -5 to -22 and positive weights ranging from 2 to 20, while the InSet lexicon provided negative weights ranging from -2 to -99 and positive weights ranging from 3 to 23. These findings suggest that the choice of sentiment lexicon can have a substantial impact on sentiment analysis results, and researchers should be mindful of these differences when interpreting their findings.

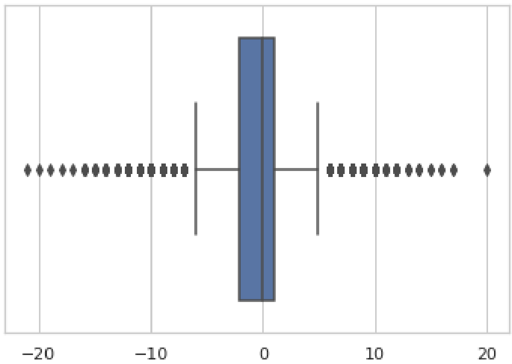


(a)

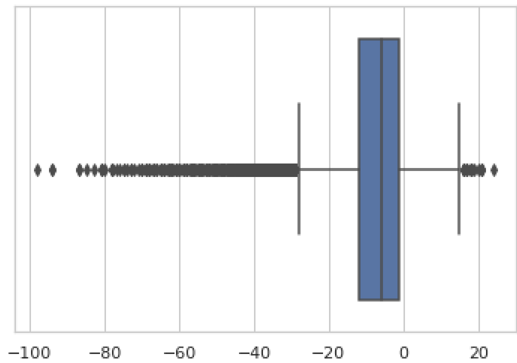


(b)

Fig. 5. Sentiment Distribution: (a) SentiStrengthID, (b) InSet



(a)



(b)

Fig. 6. Sentiment Weight Distribution (a) SentiStrengthID, (b) InSet

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v12i1.1615, Copyright ©2023
Submitted : February 15, 2023, Revised : March 8, 2023, Accepted : March 10, 2023, Published : March 14, 2023

After applying the sentiment weighting using SentiStrengthID and InSet, the next step in the sentiment analysis process was to assign labels to the data. This labeling process was carried out based on specific rules outlined in equation 1, which helped to ensure consistency and accuracy in the labeling of the data. By following these rules, each data point was assigned a label based on the sentiment weight derived from the lexicon weighting step.

The resulting distribution of labels can be seen in Figure 7, which provides an overview of the sentiment distribution of the analyzed data. This figure shows the frequency of each sentiment label assigned to the data points, indicating the overall sentiment polarity of the dataset. These labels provide a useful summary of the sentiment content of the data, and can be used to gain insights into the overall sentiment of a particular topic or domain. Overall, the use of sentiment labeling based on lexicon weighting helps to provide a systematic and objective approach to sentiment analysis, which can be useful in a variety of research and business applications.

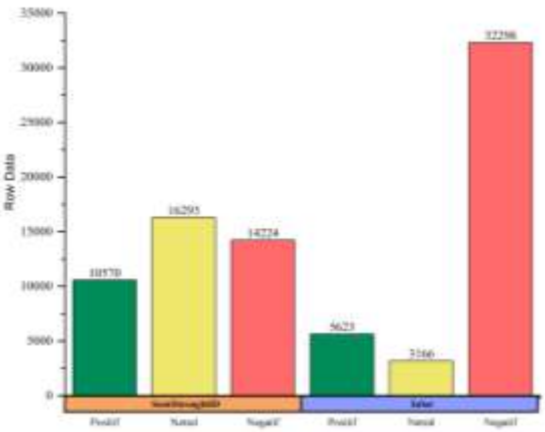


Fig. 7. Class Distribution of Sentiment

B. Performance Evaluation

The use of sentiment lexicons and machine learning models for sentiment analysis has become increasingly popular in recent years. In this study, the researchers examined the impact of two different sentiment lexicons, SentiStrengthID and InSet, on the performance of a decision tree machine learning model. The data was labeled according to the lexicons and two different criteria, entropy and Gini, were used in the training process. The results showed that the choice of sentiment lexicon and criteria can significantly influence the performance of the machine learning model in sentiment analysis.

The researchers found that the SentiStrengthID lexicon tended to produce more stable results, while the InSet lexicon was dominant in the negative class or label. This suggests that the choice of sentiment lexicon can have a significant impact on the performance of the machine learning model. Moreover, the use of different criteria can also influence the accuracy and stability of the model. By understanding these relationships, researchers can make more informed decisions about which lexicon and criteria to use for sentiment analysis, depending on their specific goals and requirements.

The training results using labels assigned by the SentiStrengthID lexicon showed significant results for each class and criterion. The f1-score for the Entropy criterion obtained an average value of 81.3%, while the Gini criterion obtained an average value of 83%. On the other hand, training using labels assigned by the InSet lexicon yielded varied results for each criterion, but tended to be the same for each class. The f1-score for the Entropy criterion obtained an average value of 63%, while the Gini criterion obtained an average value of 65%. These results demonstrate the importance of carefully selecting the sentiment lexicon and criteria for sentiment analysis.

The findings of this study have practical implications for researchers and businesses that rely on sentiment analysis. By understanding the impact of different sentiment lexicons and criteria on the performance of machine learning models, researchers can improve the accuracy and stability of their sentiment analysis models. Furthermore, businesses can use this knowledge to optimize their sentiment analysis tools and gain a deeper understanding of consumer sentiment, which can inform their marketing and branding strategies.

Finally, the choice of sentiment lexicon and criteria can have a significant impact on the performance of machine learning models in sentiment analysis. The findings of this study demonstrate the importance of carefully selecting these factors to optimize the accuracy and stability of sentiment analysis models. By understanding the relationships between sentiment lexicons, criteria, and machine learning models, researchers and businesses can improve their sentiment analysis tools and gain valuable insights into consumer sentiment.

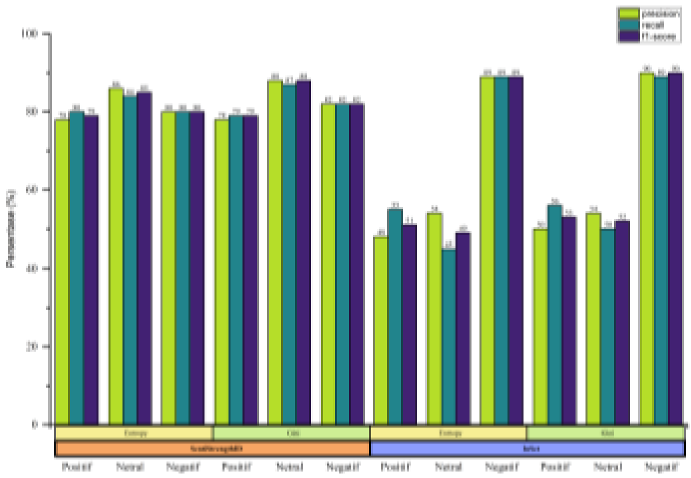


Fig. 8. precission, recall, and f1-score

Based on the f1-score values for each training process, it can be seen that the percentage obtained is greatly influenced by the amount of data for each label assigned by the type of lexicon. This condition also affects the accuracy of training and testing, as shown in Figure 9. Where the amount of data for each class has a correlation between the f1-score percentage, training accuracy, and testing accuracy. The training results using labels assigned by the SentiStrengthID lexicon tend to obtain the same values for f1-score, training accuracy, and testing accuracy. Meanwhile, the training results using labels assigned by the InSet lexicon obtained different values for f1-score, training

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v12i1.1615, Copyright ©2023
Submitted : February 15, 2023, Revised : March 8, 2023, Accepted : March 10, 2023, Published : March 14, 2023

accuracy, and testing accuracy.

These findings demonstrate the importance of carefully selecting and preparing the data for sentiment analysis. The amount and quality of data for each class can have a significant impact on the performance of the machine learning model, as evidenced by the variations in f1-score, training accuracy, and testing accuracy observed in this study. Researchers and practitioners must take into account the potential biases and limitations of the selected lexicon and the data used to train and test the model.

In addition, the results of this study suggest that the choice of lexicon can also affect the stability and consistency of the machine learning model. The SentiStrengthID lexicon tends to yield more stable and consistent results across different training processes, while the InSet lexicon is more prone to variations and fluctuations in performance. Researchers and practitioners must carefully evaluate the trade-offs between stability, accuracy, and generalizability when selecting a lexicon for sentiment analysis.

Overall, the findings of this study provide important insights into the factors that influence the performance of machine learning models for sentiment analysis. By carefully selecting the lexicon and preparing the data for training and testing, researchers and practitioners can improve the accuracy and stability of their models and generate more reliable and informative results. Further research is needed to explore the potential of other lexicons and machine learning techniques for sentiment analysis, and to evaluate their performance in different domains and contexts.

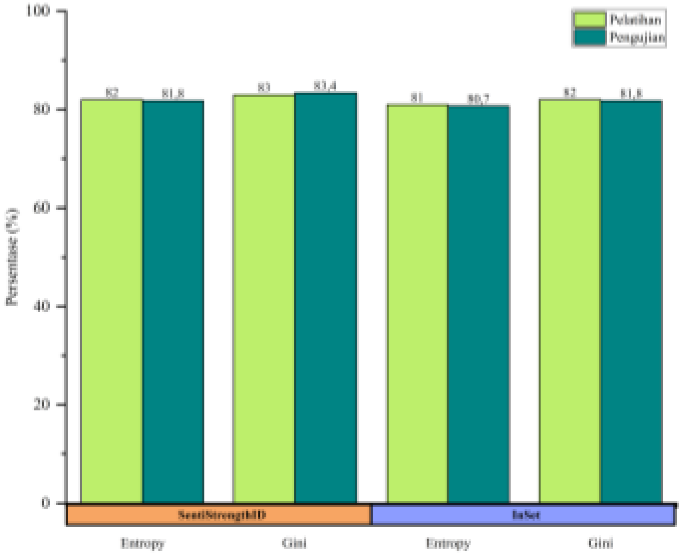


Fig. 9. Classification performance with Decision Tree

The use of lexicons to determine sentiment weights, both SentiStrengthID and InSet, can greatly affect the data training process. The number of words and weights in the lexicon does not necessarily result in balanced label assignment, as demonstrated by the InSet lexicon which produced imbalanced classes. On the other hand, the SentiStrengthID lexicon can provide more balanced labels and better performance, as shown in the two types of decision tree training using Entropy and Gini criteria. The f1-score values for each training process indicate that the percentage obtained is greatly influenced by the amount of data for each label given by the type of lexicon. This condition also affects the accuracy of training and testing, as shown in Figure 9, where the amount of data for each class has a correlation between the f1-score percentage, training accuracy, and testing accuracy.

The use of lexicons is an important factor to consider in sentiment analysis, as it can greatly affect the performance of machine learning models. The imbalance in label assignment can cause bias in the training data and lead to inaccurate results. Therefore, it is crucial to carefully select the appropriate lexicon and ensure the balance of label assignment. In this study, the SentiStrengthID lexicon was found to be more effective in providing balanced labels and better performance in decision tree training.

Overall, the results of this study demonstrate the importance of choosing the right sentiment lexicon and criteria for sentiment analysis. The performance of machine learning models can be greatly influenced by these factors, and understanding their relationships can help researchers optimize the accuracy of their results. Further research can be done to explore the effectiveness of different lexicons and criteria in different contexts and domains.

IV. Conclusion And Future Work

Based on the results of the experiments conducted to examine public sentiment towards the management of COVID-19 in Indonesia, several conclusions can be drawn. Firstly, Public sentiment shows a predominantly negative outlook (when compared between positive and negative) using both SentiStrengthID and InSet lexicon weighting. However, the SentiStrengthID lexicon also includes neutral sentiment. Secondly, The classification results are highly influenced by the label data provided by each type of lexicon, whether it is SentiStrengthID or InSet. Thirdly, the number of words and weights in the lexicon has a significant impact on determining labels, but it does not necessarily guarantee an increase in classification performance.

Further research is necessary to improve classification performance using different approaches such as deep learning optimization. It is also important to focus on addressing the issue of imbalanced classes.

Acknowledgement

I would like to express my sincere gratitude to Bina Darma University for providing the financial support that made this research possible. This project would not have been successful without the resources and facilities provided by the university. Lastly, I would like to extend my thanks to all the participants who contributed their time and effort to this study. Their willingness to participate was critical to the success of this research.

References

[1] F. Landi *et al.*, "Post-COVID-19 global health strategies: the need for an interdisciplinary approach,"

Aging Clin. Exp. Res., vol. 32, no. 8, pp. 1613–1620, 2020, doi: 10.1007/s40520-020-01616-x.

[2] B. Liu, "Social Network Analysis," in *Web data mining: Exploring hyperlinks, contents, and usage data*, vol. 1, Heidelberg: Springer, 2011, pp. 269–309. doi: 10.1007/978-3-642-19460-3.

[3] D. Sharma, M. Sabharwal, V. Goyal, and M. Vij, "Sentiment analysis techniques for social media data: a review," in *First international conference on sustainable technologies for computational intelligence*, 2020, pp. 75–90.

[4] B. Laurensz and Eko Sediyono, "Analisis Sentimen Masyarakat terhadap Tindakan Vaksinasi dalam Upaya Mengatasi Pandemi Covid-19," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 10, no. 2, pp. 118–123, 2021, doi: 10.22146/jnteti.v10i2.1421.

[5] W. Yulita *et al.*, "Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma Naïve Bayes Classifier," *J. Data Min. dan Sist. Inf.*, vol. 2, no. 2, pp. 1–9, 2021.

[6] F. F. Rachman and S. Pramana, "Analisis Sentimen Pro dan Kontra Masyarakat Indonesia tentang Vaksin COVID-19 pada Media Sosial Twitter," *Heal. Inf. Manag. J.*, vol. 8, no. 2, pp. 100–109, 2020, [Online]. Available: https://inohim.esaunggul.ac.id/index.php/INO/article/view/223/175

[7] A. L. Fairuz, R. D. Ramadhani, and N. A. F. Tanjung, "Analisis Sentimen Masyarakat Terhadap COVID-19 Pada Media Sosial Twitter," *J. Dinda Data Sci. Inf. Technol. Data Anal.*, vol. 1, no. 1, pp. 42–51, 2021, doi: 10.20895/dinda.v1i1.180.

[8] A. Prihatini, "Semantic network of the word association in the field of law," *Litera*, vol. 18, no. 3, pp. 430–446, 2019.

[9] P. Putra and S. B. Yudhoatmojo, "User Perception Analysis of Online Learning Platform 'Zenius' During the Coronavirus Pandemic Using Text Mining Techniques," *J. Sist. Inf.*, vol. 17, no. 2, pp. 33–47, 2021, doi: 10.21609/jsi.v17i2.1065.

[10] F. Koto and G. Y. Rahmaningtyas, "InSet Lexicon: Evaluation of a Word List for Indonesian Sentiment Analysis in Microblogs," *Proc. 2017 Int. Conf. Asian Lang. Process. IALP 2017*, vol. 2018-Janua, pp. 391–394, 2018, doi: 10.1109/IALP.2017.8300625.

[11] A. Saiful, "Prediksi Harga Rumah Menggunakan Web Scrapping dan Machine Learning Dengan Algoritma Linear Regression," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 8, no. 1, pp. 41–50, 2021, doi: 10.35957/jatisi.v8i1.701.

[12] V. Chitraa and D. A. S. Davamani, "A Survey on Preprocessing Methods for Web Usage Data," *Int. J. Comput. Sci. Inf. Secur.*, vol. 7, no. 3, pp. 78–83, 2010, doi: 10.48550/arXiv.1004.1257.

[13] T. B. T. Dewi, N. A. Indrawan, I. Budi, A. B. Santoso, and P. K. Putra, "Community Understanding of the Importance of Social Distancing Using Sentiment Analysis in Twitter," in *2020 3rd International Conference on Computer and Informatics Engineering, IC2IE 2020*, 2020, pp. 336–341. doi: 10.1109/IC2IE50715.2020.9274589.

[14] C. R. Panigrahi, B. Panda, and B. Pati, "Exploratory Data Analysis and Sentiment Analysis of Drug Reviews," *Comput. y Sist.*, vol. 26, no. 3, pp. 1181–1189, 2022, doi: 10.13053/cys-26-3-4093.

[15] R. K. Pearson, *Exploratory data analysis using R*. New York: Taylor & Francis, 2018. doi: 10.1201/9781315382111.

[16] N. W. Prabowo, V. P. Siregar, and S. B. Agus, "Klasifikasi Habitat Bentik Berbasis Objek Dengan Algoritma Support Vector Machines Dan Decision Tree Menggunakan Citra Multispektral Spot-7 Di Pulau Harapan Dan Pulau Kelapa," *J. Ilmu dan Teknol. Kelaut. Trop.*, vol. 10, no. 1, pp. 123–134, 2018, doi: 10.29244/jitkt.v10i1.21670.

[17] S. Rizky Ade Putranto, Triastuti Wuryandari, "Perbandingan Analisis Klasifikasi Antara Decision Tree Dan Support Vector Machine Multiclass," *J. Gaussian*, vol. 4, no. 4, pp. 1007–1016, 2015

[18] A. Géron, *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. " O'Reilly Media, Inc.," 2022.

[19] P. R. Togatorop and A. Fauzi, "Klasifikasi Penggunaan Masker Wajah Menggunakan Squeezenet," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 1, pp. 397–406, 2022, doi: 10.35957/jatisi.v9i1.642.

[20] S. Devella, Y. Yohannes, and F. N. Rahmawati, "Implementasi Random Forest Untuk Klasifikasi Motif Songket Palembang Berdasarkan SIFT," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 7, no. 2, pp. 310–320, 2020, doi: 10.35957/jatisi.v7i2.289.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v12i1.1615, Copyright ©2023
Submitted : February 15, 2023, Revised : March 8, 2023, Accepted : March 10, 2023, Published : March 14, 2023