Sentiment Classification of COVID-19 Issues on Twitter Using Indonesia Sentiment Lexicon and Machine Learning-Based Approaches (Case Study: Indonesia, March-May 2020)

N A Daulay[1], **R Ramadhan**[1], **L H Suadaa** [1,2,*]

[1] Department of Statistical Computing, Politeknik Statistika STIS, Jakarta, Indonesia
[2] BPS-Statistics Indonesia, Jakarta, Indonesia

*Corresponding author's e-mail: lya@stis.ac.id

Abstract. This study examines sentiment analysis related to COVID-19 in Indonesia (March-May 2020) using InSet Lexicon as training data in supervised machine learning models. The dataset comprises 7,967 tweets, divided into 90% training data and 10% testing data. The results reveal that Support Vector Machine (SVM) and Random Forest (RF) are the most effective methods, achieving accuracy above 80%, with SVM reaching 87% and RF at 86%. InSet Lexicon itself attains an accuracy of 75%, a macro average of 69%, and a weighted average of 74%, making it an effective alternative for large-scale data labeling. Research recommendations support further development of InSet Lexicon for sentiment classification and expansion of the lexicon for foreign languages to enhance sentiment analysis accuracy in a global context. This study provides valuable insights into understanding public sentiment regarding crucial issues such as COVID-19 in Indonesia.

1. Introduction

The COVID-19 pandemic, which emerged in early 2020, was declared a global pandemic by the World Health Organization (WHO) [1]. As a global pandemic, COVID-19 has become not only a public health issue ([2], [3], [4]) but has also disrupted various aspects of social life [5], the economy [6], and politics ([7], [8]). Indonesia, as one of the countries significantly affected by this pandemic ([9], [10], [11]), faces significant challenges in managing the spread of the virus [12], mitigating its impact [13], and responding to public concerns and perceptions [14] related to the issue.

To break the chain of COVID-19 transmission, the Indonesian government has implemented various policies, such as restrictions on educational activities [15], limitations on work activities [16], and prohibitions on gatherings [17]. Due to restrictions on direct contact, many people have turned to social media for socializing [18]. Compared to 2019, the number of social media users increased by 25 million in 2020 [18].

Twitter, as one of the social media platforms, has a substantial user base in Indonesia [19]. Twitter has become a platform for Indonesian citizens to share information, communicate, and express their sentiments regarding the COVID-19 issue [20]. The increased activity on social media during the pandemic [18] reflects the vital role of this platform in creating a virtual public space that enables widespread interactions among individuals, information exchange, as well as the expression of emotions and viewpoints on critical issues [19].

The heightened social media activity during the pandemic [18] not only reflects the public's eagerness to engage in open discussions but also serves as a valuable source of data for understanding public sentiment [21]. Studying how the public responds to the COVID-19 pandemic on social media can provide deep insights into the attitudes, concerns, and perceptions that evolve within society.

One way to comprehend public sentiment is through sentiment analysis. Sentiment analysis involves Natural Language Processing (NLP), text analysis, and computational capabilities to extract insights from a collection of texts ([22]-[25]). Various sectors and disciplines have adopted it for various purposes. Sentiment analysis has been widely utilized to understand public opinions ([26], [27], [28]), early warning or crisis detection ([29], [30]), information filtering [31], customer satisfaction measurement ([32]-[35]), and marketing advertisement evaluation ([36], [37], [38]). One crucial aspect of sentiment analysis is sentiment classification ([39], [40], [41]), which allows us to categorize texts into sentiment categories such as positive, negative, or neutral [42]. These classification methods can be used to categorize public views on a specific topic or issue. Two common methods used in sentiment classification are lexicon-based ([43]-[47]), where texts are analyzed based on dictionaries of words with specific sentiments, and manual labeling ([48], [49], [50]), where humans perform direct classification.

A Lexicon is a dictionary of words classified based on specific sentiments, such as positive, negative, or neutral ([51], [52], [53]). When used in sentiment analysis, this dictionary is employed to identify words in the text and associate them with the corresponding sentiment values. However, there are several limitations to be noted with lexicons. The first limitation is the lexicon's inability to deal with words or phrases that are not in its dictionary ([54], [55]). This results in many text contents that may not be classified correctly or may be ignored. Another limitation is the lexicon's inability to handle complex contexts and word meanings that vary depending on the context ([56], [57]). Therefore, lexicons often lack the precision required for in-depth sentiment analysis.

In addition to lexicons, manual labeling methods are also used to classify sentiment in text. In this method, humans manually classify text as positive, negative, or neutral. However, this approach has several drawbacks, including being time-consuming, resource-intensive, susceptible to individual biases, and challenging to use on a large scale.

In the context of sentiment analysis, a promising approach is machine learning. Machine learning is a branch of artificial intelligence that allows computers to "learn" from data and develop the ability to make predictions or decisions without being explicitly programmed ([58], [59]). The main advantage of machine learning is its ability to handle complexity and uncertainty in text data, process large volumes of data efficiently, and automatically update its model to improve accuracy ([60], [61]).

In sentiment analysis, machine learning holds great potential [62]. Using this technique, computational models can learn from existing labeled data to automatically classify text on a large scale and with higher accuracy. Therefore, this research considers the application of machine learning to classify sentiment in texts related to the COVID-19 issue on Twitter. We intend to use the results of sentiment prediction based on lexicon as training data to train our machine learning model to produce the best model in our study case.

Several studies have carried out sentiment classification using machine learning. [64] tried to classify sentiment about documents with some machine learning and found that naive bayes had the best performance. Then, [65] applied some machine learning to text filtering and found that the Rocchio method worked well on the data used. Then, [66] carried out an analysis of Genetic Algorithm with several machine learning and found that K-Nearest Neighbor had the best accuracy in tumor classification. Another study, [67], tried to compare several machine learning methods for predicting compressive strength of concrete and found that the Support Vector Machine was better and more stable. Then, [68] detected the anomaly with some machine learning and found that Random Forest was superior. Therefore, we are interested in applying these methods to the data we have so that the best method is obtained.

Thus, this study aims to: (1) analyze the sentiment of Twitter users in Indonesia related to the COVID-19 pandemic from March to May 2020 using Naive Bayes, Rocchio Classification, K-Nearest Neighbor, Support Vector Machine (SVM), and Random Forest classification modeling methods. (2) Compare the performance of Naive Bayes, Rocchio Classification, K-Nearest Neighbor, Support Vector Machine (SVM), and Random Forest classification modeling methods in analyzing the sentiment of Twitter users in Indonesia regarding the COVID-19 pandemic. This research is expected to yield the best model and provide deeper insights into the complexities of this global pandemic situation.

2. Methodology

2.1. Data Used

The research was conducted to implement sentiment analysis in identifying and analyzing reactions and public opinion on social media Twitter about COVID-19. In this study, the method used for sentiment analysis is lexicon-based and various learning supervised methods. The data used is data on Twitter collected from March 20 to May 20, 2020, of 7,967 tweets that have been taken based on crawling [62].

2.2. Data Preprocessing

The data obtained from the data collection phase still contains noise. Noise is an unusable element in research. Noise can be symbols, numbers, links, and so on. This can cause problems when further analysis is carried out. Therefore, data preprocessing needs to be done, as well as data cleaning. These phases are intended to transform unstructured data into structured, ready-to-analyze data. In this study, case folding was performed to convert text into a lowercase or small letter form. Then, tokenizing to break down a word into a single word (unigram), as well as stemming to change the equivocal words into their basic form. Figure 1 shows the raw data before preprocessing and after it.

	id	created_at	source	original_text	clean_text	lang
0	1266039172315230209	Thu May 28 16:10:35 +0000 2020	Twitter for Android	RT @ridwankamil: Minggu ini 60% Kota/Kab Jabar sudah di zona Biru (level 2) dan 40 % di zona kuning (level 3]. Sudah tdk ada lagi yg di zon...	kota kabupaten jabar zona biru level zona kuning level tidak zon	in
1	1266039166350983169	Thu May 28 16:10:33 +0000 2020	Twitter for Android	RT @kompascom: Pembukaan mal tidak berarti akan berdampak pada masyarakat kecil. Hal itu karena mal-mal lebih menyasar kepada kalangan mene...	pembukaan mal tidak berdampak masyarakat kecil mal mal menyasar kalangan mene	in

Figure 1. Data Before and After Cleaning

2.3. Inset Lexicon-Based and Free-Handed Labelling

After the data preprocessing is done, every review tweet is labeled. This labeling is used as target data that uses machine learning methods to learn about the sentiment category of existing data so that it includes supervised learning. In this study, two labeling methods were used: Inset Lexicon-Based on data training and Free-Handed Labeling on data testing with the aim of seeing the variation of labeling. Inset lexicon-based uses the previously developed Indonesian list dictionary [https://github.com/fajri91/InSet], which consists of a group of words (bag of words) weighted positive and negative. Each word will be counted in its weight so that if the review has a positive weight, then it will be categorized as positive, and so on for the negative category. In comparison, the neutral category has a weight of 0. Free-handed labeling is done on data testing by labeling each review manually; in this case, all the researchers in this study contribute to doing manual labeling.

2.4. Modeling

The data used was 7,967 tweets, with the division of 90% training data and 10% testing data. Classification models used are Naive Bayes, Rocchio Classification, K-Nearest Neighbour, Support Vector Machine (SVM), and Random Forest methods. The k-fold cross-validation process is also applied with the k value of 5. The models performed by the grid search process are K-Nearest Neighbours (KNN), Support Vector Machine (SVM), and Random Forest (RF). Table 1 illustrates the range of hyperparameter tuning used.

Table 1. Hyperparameter Tuning Range for Classification Model

Model	Hyperparameter/Parameter	Description
K-Nearest Neighbour	Neighbors: [3,5,7,9,13,15,17,19,21,23,25,27,29]	29 is the square root of the test data.
Support Vector Machine	C: [0.1, 1, 10, 100, 1000] Gamma: [1,0.1,0.01,0.001,0.0001], Kernel: [linear, rbf]	The kernels used are linear and radial basic functions. (RBF).
Random Forest	Tree: [100,300,500,1000]	The maximum number of trees is 100 with the maximum depth tree is 300

A Naive Bayes classifier is a basic probabilistic model that uses Bayes' rule and a strong assumption of independence. The Naive Bayes model simplifies things by assuming that, for a given class (positive or negative), the words are independent of each other. While this assumption doesn't have a major impact on text classification accuracy, it enables the use of highly efficient classification algorithms for the task. In their 2003 paper, [69] evaluated how Naïve Bayes performs in text classification tasks.

In our situation, the expression that provides the probability of a word belonging to a specific class based on maximum likelihood is as follows:

$$P(x_i|c) = \frac{count\ of\ x_i\ in\ documents\ of\ class\ c}{Total\ no\ of\ words\ in\ documents\ of\ class\ c} \quad (1)$$

During the training stage, the word frequency counts are stored in hash tables. According to Bayes' Rule, the probability of a specific document being associated with a class ci is expressed as follows:

$$P(c_i|d) = \frac{P(d|c_i)\ x\ P\ (c_i)}{P(d)} \quad (2)$$

When we apply the simplifying conditional independence assumption, which states that, given a class (positive or negative), the words are considered conditionally independent from each other, this model is referred to as 'naïve' due to this simplification.

$$P(c_i|d) = \frac{(\prod P(x_i|c_j))\ x\ P(c_j)}{P(d)} \quad (3)$$

In this context, the xi represent the individual words within the document. The classifier determines the class with the highest posterior probability. Furthermore, we eliminate duplicate words from the document since they do not contribute additional information. This variant of the Naïve Bayes algorithm is known as Bernoulli Naïve Bayes. Focusing solely on the presence or absence of a word, rather than its count, has been observed to yield slight performance improvements, especially in cases with a large number of training examples.

Rocchio Classification is a document classification method based on vector space that uses the average or center of mass of documents in each class as prototypes or centroids. This method assumes that documents relevant to a query will be close to the corresponding class centroid, while irrelevant documents will be far from that centroid. The method classifies new documents by calculating their distances to class centroids and then assigning them to the nearest class. This method is simple and efficient, but it may not be accurate when classes are not spherical and have similar radii [70].

Pattern classification is a crucial task within the domains of big data, data science, and machine learning. The K-nearest neighbor (KNN) algorithm is among the earliest, simplest, and most accurate methods for pattern classification and regression models. Originally introduced in 1951 by [71] and subsequently refined by [72], KNN has earned recognition as one of the top ten techniques in data mining [73]. As a result, KNN has undergone extensive research and widespread application across various fields [74]. Consequently, KNN serves as the foundational classifier in numerous pattern classification problems, including pattern recognition [75], text categorization [76], ranking models [77], object recognition [78], and event recognition [79] applications.

Support Vector Machine (SVM) implements algorithms by searching for hyperplans in dimensional spaces with as many features as possible so that classifications can be produced well and clearly. A good classification result is to find a field that has the maximum margin or maximum distance between two classes of data points. One of the advantages of this SVM is that it uses a kernel that can work in both linear and non-linear situations. The use of a linear kernel is the same as regression in general, whereas nonlinear kernels can be used such as polynomial, sigmoid, and rbf.

The last algorithm used is a random forest that uses an ensemble method or a combination of many tree-based modeling or weak learner methods. Using this method can produce better predictions than single models. Random forests are created from subsets of data, and the final output is based on average or majority ranking; hence the problem of overfitting is taken care of. Random forest randomly selects observations, builds a decision tree, and takes the average result.

2.5. Evaluation

The study uses several measurements of evaluation: precision, recall, F1-score, accuracy, macro average, and weighted average [83][84]. Precision takes into account all predicted positive predictions being positive, while recall is a measure of how well the model can find all the positive cases. Then the F1-Score takes into account both aspects of model performance, precision and recall in a single number, so we can get a more complete picture of the model's performance. While accuracy takes into account the total number of true predictions divided by the total amount of predictions made for a set of data. Macro Average, as the name suggests, operates on a class-by-class basis, independently calculating performance metrics such as Precision, Recall, F1-Score, and Accuracy for each class [80]. This approach does not take into account the imbalance in class distribution and assigns equal weight to every class [80]. Subsequently, the Macro Average is calculated by taking the unweighted mean of these class-specific metrics. It offers an unbiased evaluation of a model's performance, treating each class equally and thus mitigating the bias caused by class imbalance. In contrast, Weighted Average is tailored to address the issue of class imbalance directly. It assigns different weights to each class proportional to the number of samples in that class. By doing so, it provides a more accurate representation of a model's performance, giving more importance to the classes with more instances. Performance metrics, such as Weighted Precision, Weighted Recall, Weighted F1-Score, and Weighted Accuracy, are computed by taking these weighted contributions into account. Weighted Average serves as a pragmatic choice in scenarios where class imbalance is a significant concern, as it offers a practical reflection of a model's utility in real-world applications [81][82].

Table 2. Confusion matrix

Confusion Matrix		Actual Classes	
		Yes	No
Predicted Classes	Yes	True Positive (TP)	False Positive (FP)
	No	False Negative (FN)	True Negative (TN)

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

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

$$F1 - Score = \frac{2\ x\ precision\ x\ recall}{precision + recall} \tag{6}$$

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP} \tag{7}$$

3. Result and Discussion

3.1. Data Analysis and Exploration

In this section on data analysis and exploration, we will introduce the results of the analysis and data exploration that have been conducted. These results were obtained through classification using InSet Lexicon, a collection of words in the Indonesian language, and their sentiment scores. Using this method, we will discuss interesting findings related to public sentiment toward the issue under investigation, providing a deep understanding of the dynamics of public opinion over a specific time frame.

The data set includes a total of 7,967 Indonesian language tweets distributed in 18 different columns. All columns essentially contain string data types, with exceptions such as id, the number of likes, the number of retweets, the word count, and sentiment scores, which have more structured data types. As part of the analysis process, the dataset has undergone the necessary cleaning and preprocessing steps to ensure data quality and consistency. One key step in this phase is the removal of negation words from each tweet, enabling a more accurate sentiment analysis.

Figure 2 presents a visualization depicting the most frequently used words by Twitter users in this collected dataset. As shown in the word cloud, words like "zona," "mal," and "level" stand out with a larger appearance, indicating the highest frequency of occurrence compared to other words. This strongly suggests that these three words play a central role in the conversations on this platform when negation words are excluded from the calculations. Thus, this visualization provides an initial overview of topics or issues that may be highly relevant in the context under investigation.

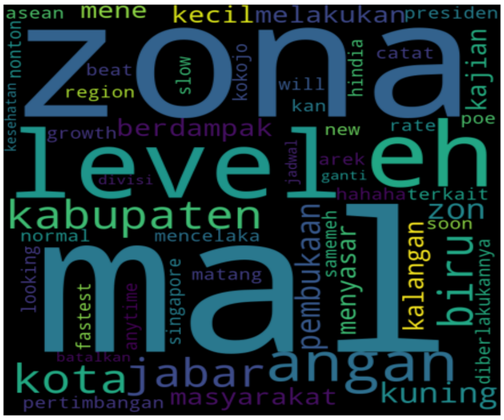

Figure 2. Wordcloud

In this analysis, each token found in every tweet is assigned a weighted score based on its contribution to the overall sentiment, and these weights are then summed up. The result of this summation process is referred to as the sentiment score. Figure 3 provides a visual representation of the distribution of sentiment scores within the dataset used. As shown in Figure 3, the density curve shows a significant tendency towards negative sentiment. This means that the majority of the data tends to have negative sentiment values. This finding is crucial in understanding public sentiment towards the investigated issue, indicating a dominance of dissatisfaction or displeasure in the observed context.

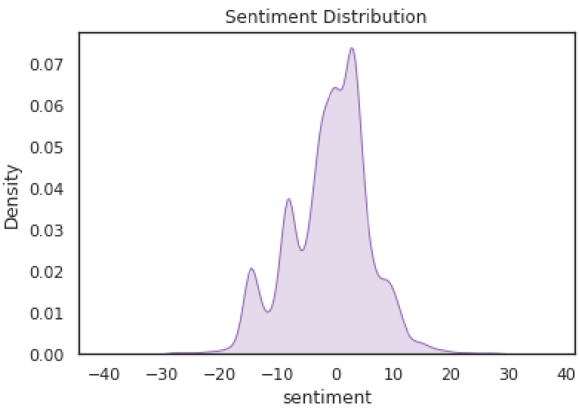

Figure 3. Sentiment Distribution

In Figure 4, we can observe the presence of a prominent boxplot that distinctly indicates the existence of a number of outliers in the negative sentiment category. This phenomenon provides important insights that the distribution of sentiment tends to be skewed towards the negative side in this dataset. In other words, the significant outliers in negative sentiment suggest the presence of a number of tweets or content with sentiment much more negative than the majority of the data, which may reflect critical or controversial aspects in this sentiment analysis.

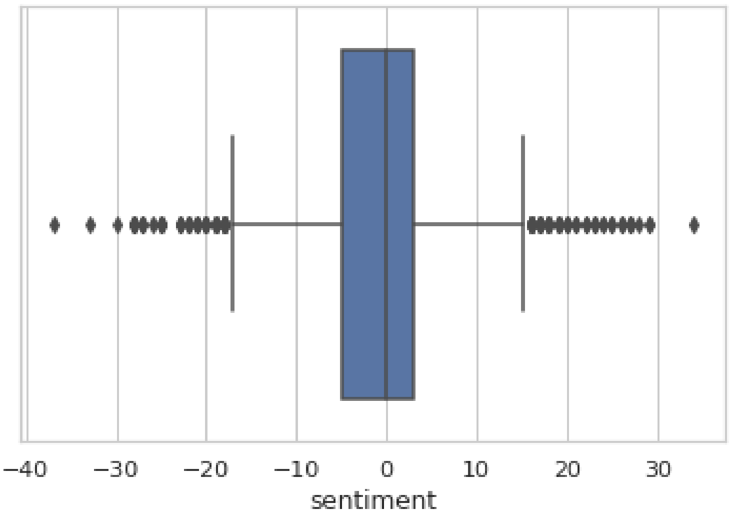

Figure 4. Boxplot Sentiment

The sentiment scores obtained through this analysis are then categorized based on sentiment types to better understand the public's perception of the issue under investigation. In determining sentiment types, if the sentiment score is greater than 0, it is considered positive sentiment, reflecting supportive or happy responses. Conversely, if the sentiment score is less than 0, it is considered negative sentiment, indicating dissatisfaction or disagreement. If the sentiment score equals 0, it is considered neutral sentiment, reflecting a neutral or sentiment-less attitude.

Figure 5, presented in the analysis, vividly depicts the dominance of sentiment types in the collected data. It can be clearly observed that most of the data tends to have negative sentiment. This finding provides a significant insight into the public's perspective on the observed issue. It indicates that in the context under investigation, the public tends to have a critical or unsatisfied viewpoint.

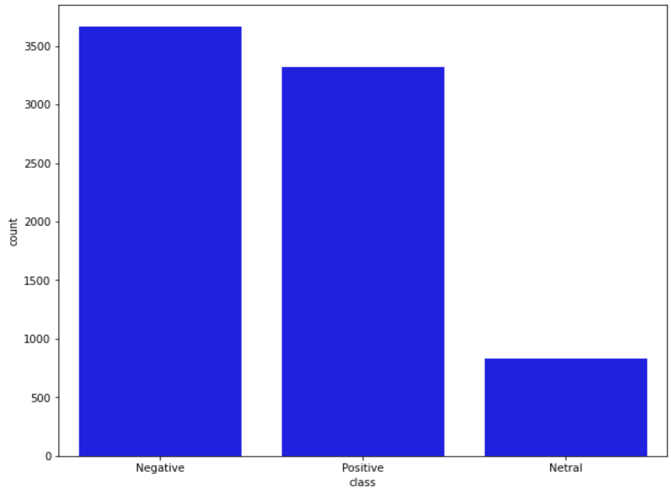

Figure 5. Barplot of Sentiment Types

This analysis also reveals a significant correlation among several words in the dataset. This correlation reflects how often specific words appear together in the context of the same tweets. In fact, there are as many as 156 words that consistently appear together in every observed tweet. In other words, these words have a strong bond and a significant connection in the conversations on the platform.

Figure 6, presented in the analysis, provides a visual overview of some words with significant correlations. In this figure, we can clearly see a number of words that frequently appear together in most of the analysed tweets. This finding indicates that these words play an important role in the observed context and may refer to specific topics or issues that consistently capture the public's attention.

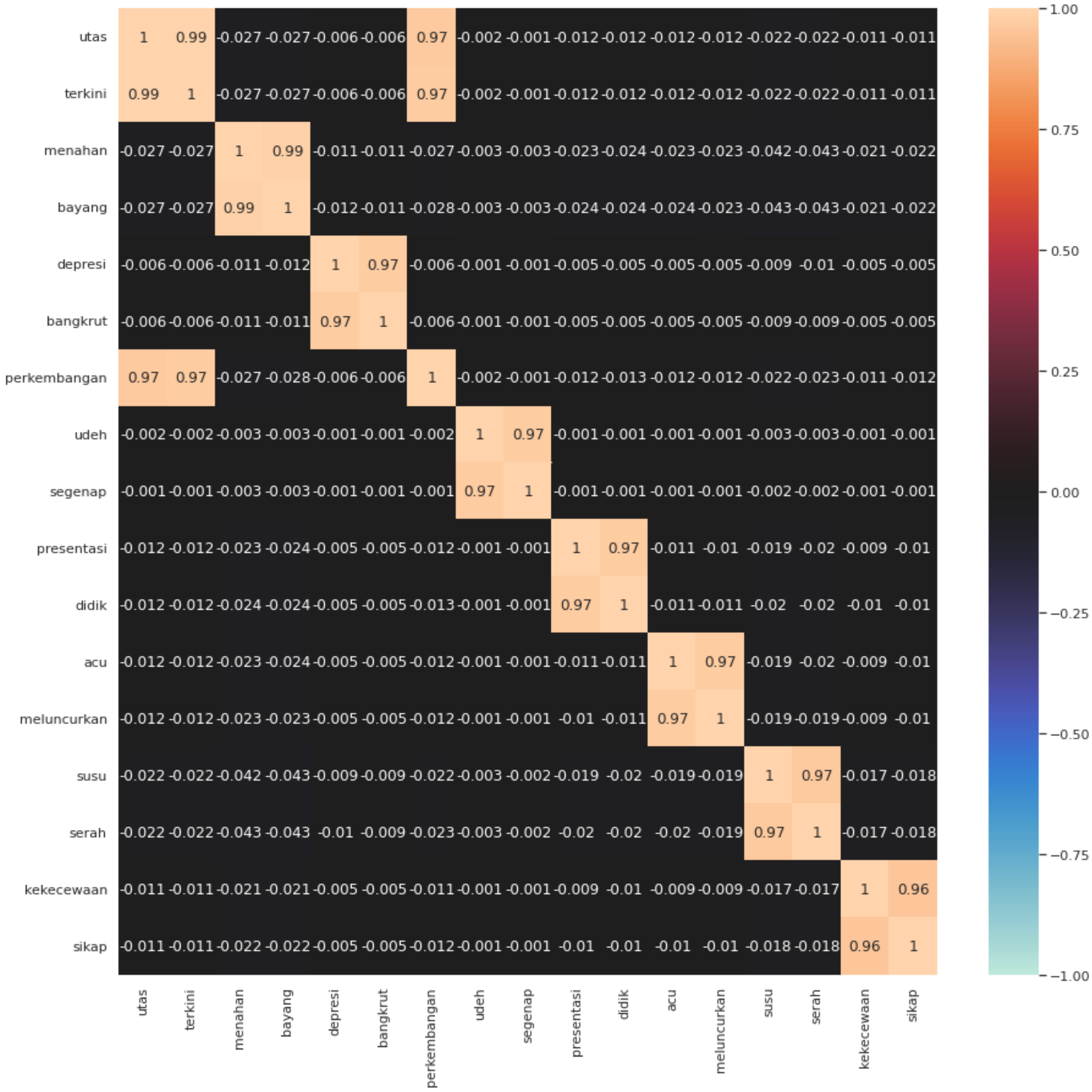

Figure 6. Correlation of Words that Often Appear Together

Figure 6 shows several words frequently appearing together in a single tweet. The correlation between these words measures how closely spaced these two words are in each tweet. The higher the correlation value, the closer the distance between these two words in the observed tweets. For example, consider the words "perkembangan" and "terkini" which have a correlation value of 0.97. This correlation value indicates that most tweets containing the word "perkembangan" are usually directly followed by the word "terkini", while the words "depresi" and "kekecewaan" have a correlation value of around -0.005, which means that the more often the word "depresi" appears, the less frequently "kekecewaan" appears. Furthermore, this research also highlights the most frequently appearing words in the observed tweets. These specific words have been included in the lexicon dictionary. Figure 7 displays the top 15 words that appear most frequently in the dataset. In this case, the word "tidak" dominates with a frequency of 1949 appearances, followed by the word "normal" with 1,018 appearances. This information provides valuable insights into the words most commonly used in the context under investigation and underscores the importance of these words in understanding public sentiment regarding relevant issues.

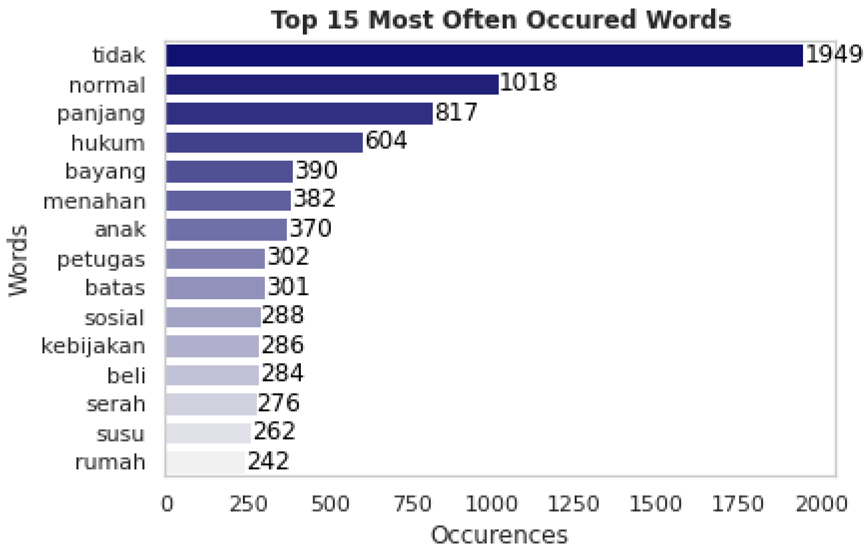

Figure 7. Most Frequently Appearing Words

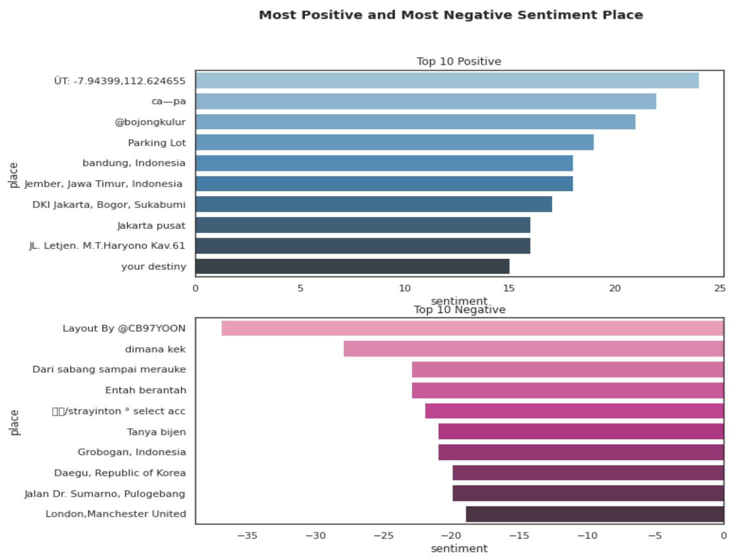

Figure 8. Locations Generating the Most Positive and Most Negative Sentiments

In this analysis, we discovered something interesting about users who list "Indonesia" as their location on Twitter. Although most tweets originate from users located in "Indonesia," the average sentiment score of these tweets tends to be closer to negative sentiment, approximately -0.986. In other words, overall, tweets generated by users who list "Indonesia" as their location tend to have more negative sentiments than positive ones. A clearer picture is shown in Figure 9.

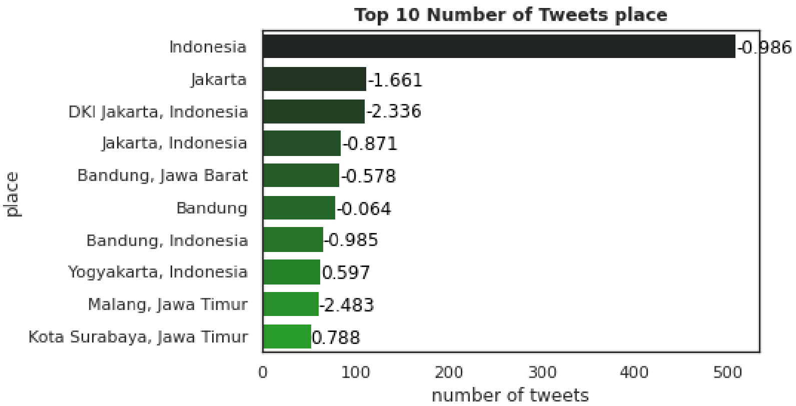

Figure 9. Locations Generating the Most Tweets

3.2. Evaluation Results of Classification with InSet Lexicon

Table 3 shows classification results using the testing data labeled by the InSet Lexicon (Indonesia Sentiment Lexicon) compared to manual labeling (hand-labeling).

Table 3. Evaluation Results of Classification with Inset-Lexicon Method on Testing Data

Sentiment	*Precision*	*Recall*	*F1-Score*	*Support*
Negative	0.67	0.75	0.71	285
Neutral	0.60	0.46	0.52	116
Positive	0.84	0.83	0.83	382
Accuracy		75%		
Macro Average	70%	68%	69%	-
Weighted Average	74%	74%	74%	-

Based on Table 3, classification using the Lexicon Inset method yields a 75% accuracy of classification when compared to the handed-labeling method. If each class of sentiment in the classification has an equally important contribution to the evaluation calculation, then the macro average value is 69%. In comparison, using a weighted average yield of 74% which has given a greater contribution to the class of sentiments whose data is more in the test data. Thus, based on the results of accuracy, macro averages, and weighted averages, the method of classification with InSet Lexicon can be used as the classification of sentiment data on Twitter in this study.

3.3. Evaluation Results of Classification with Machine Learning

Tables 4 and 5 show the evaluation of classification results using the machine learning method with the training data using InSet Lexicon labeling. The sentiment prediction results compared to free hand-labeling on the testing data.

Table 4. Evaluation Results of Classification Using Machine Learning on Testing Data

Method	Negative			Positive			Neutral		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
Naïve Bayes	0.57	0.72	0.64	0.71	0.72	0.72	0.63	0.19	0.29
Rocchio Classifier	0.55	0.50	0.52	0.86	0.34	0.49	0.21	0.66	0.31
K-Nearest Neighbours	0.59	0.69	0.63	0.74	0.70	0.72	0.42	0.31	0.36
Support Vector Machine	0.61	0.75	0.67	0.75	0.74	0.75	0.60	0.29	0.39
Random Forest	0.61	0.74	0.67	0.80	0.67	0.73	0.46	0.47	0.46

Table 5. Accuracy, Macro Average, and Weighted Average Results Using Machine Learning on Testing Data

Metode	*Accuracy*	*Macro Average*	*Weighted Average*
Naive Bayes	81%	73%	80%
Rocchio Classification	58%	56%	63%
K-Nearest Neighbour	84%	79%	84%
Support Vector Machine	87%	83%	87%
Random Forest	86%	83%	87%

The Naive Bayes method is good at studying sentiment classifications on InSet Lexicon training data, which has shown promising results on his testing data with an accuracy of 81%, a macro average of 73%, and a weighted average of 80%. Then, the Rocchio Classification method is only capable of producing 58% of the accurate classification with its macro averages of 56% and the weighted mean of 63%. Furthermore, the K-Nearest Neighbours method can produce a good classification accuracy of 84% with a macro average of 79% and a weighted average of 84%. The Support Vector Machine (SVM) method with the optimum Radial Basic Function (RBF) kernel can produce an excellent classification accuracy of 87%, with a macro average of 83% and a weighted average of 87%. Thus, the support vector machine method is excellent in classifying sentiment classes on the data used that are higher than the Naive Bayes, Rocchio Classification, and K-Nearest Neighbours methods. Finally, the Random Forest method can produce a good classification accuracy of 86% with a macro average of 83% and a weighted average of 87%, so it is good at classifying the sentiment classes.

4. Conclusion

Based on data visualization using word cloud, words like "zona," "mal," and "level" stand out with a larger appearance, indicating the highest frequency of occurrence compared to other words from the tweets. Based on the sentiment distribution by density curve shows a significant tendency towards negative sentiment for covid-19 issues in Indonesia. By a total of 7967 tweet data as well as a division of 90% training data and 10% testing data obtained that the best modeling techniques in the sentiment classification of COVID-19 issues in Indonesia from March 20, 2020 to May 20, 2020 are the Support Vector Machine (SVM) and Random Forest methods. (RF). It is proven that the results of the accuracy, macro average, and weighted average of the two methods against the labeling of InSet Lexicon result in classification precision above 80% and only a 1% difference between the methods (accuracy SVM

(87%) greater than RF (86%)). Inset Lexicon itself produced accuracy of 75%, macro averages of 69%, and weighted averages of 74%, so the result of classification is also quite good and can be used as a substitute for hand labeling in large data. Thus, the two methods, SVM and RF, can be used to classify sentiments about COVID-19 in Indonesia, and InSet Lexicon can be utilized as a labeling technique in the classification of large amounts of data. The author's suggestion for further research is that the InSet Lexicon is more developed in the classification of sentiments or opinions. Later, a Lexicon of a foreign language can also be used to predict sentiments that are not using the Indonesian language.

References

[1] Saputra C, Putra ID. Pemberdayaan Penanggulangan Covid-19 Bagi Petugas Kesehatan. JCES (Journal of Character Education Society). 2020 Jul 16;3(2):320-8.
[2] Tsai J, Wilson M. COVID-19: a potential public health problem for homeless populations. The Lancet Public Health. 2020 Apr 1;5(4):e186-7.
[3] Ward JK, Alleaume C, Peretti-Watel P, Seror V, Cortaredona S, Launay O, Raude J, Verger P, Beck F, Legleye S, L'Haridon O. The French public's attitudes to a future COVID-19 vaccine: The politicization of a public health issue. Social science & medicine. 2020 Nov 1;265:113414.
[4] Abuelgasim E, Saw LJ, Shirke M, Zeinah M, Harky A. COVID-19: Unique public health issues facing Black, Asian and minority ethnic communities. Current problems in cardiology. 2020 Aug 1;45(8):100621.
[5] Kajdy A, Feduniw S, Ajdacka U, Modzelewski J, Baranowska B, Sys D, Pokropek A, Pawlicka P, Kaźmierczak M, Rabijewski M, Jasiak H. Risk factors for anxiety and depression among pregnant women during the COVID-19 pandemic: A web-based cross-sectional survey. Medicine. 2020 Jul 7;99(30).
[6] Hevia C, Neumeyer A. A conceptual framework for analyzing the economic impact of COVID-19 and its policy implications. UNDP Lac COVID-19 Policy Documents Series. 2020 Mar 20;1:29.
[7] Ward JK, Alleaume C, Peretti-Watel P, Seror V, Cortaredona S, Launay O, Raude J, Verger P, Beck F, Legleye S, L'Haridon O. The French public's attitudes to a future COVID-19 vaccine: The politicization of a public health issue. Social science & medicine. 2020 Nov 1;265:113414.
[8] Yoosefi Lebni J, Abbas J, Moradi F, Salahshoor MR, Chaboksavar F, Irandoost SF, Nezhaddadgar N, Ziapour A. How the COVID-19 pandemic effected economic, social, political, and cultural factors: A lesson from Iran. International Journal of Social Psychiatry. 2021 May;67(3):298-300.
[9] Tanjung AA, Ruslan D, Lubis I, Pratama I. Stock Market Responses to Covid-19 Pandemic and Monetary Policy in Indonesia: Pre and Post Vaccine. Cuadernos de Economía. 2022 Mar 31;45(127):120-9.
[10] Devi S, Warasniasih NM, Masdiantini PR, Musmini LS. The impact of COVID-19 pandemic on the financial performance of firms on the Indonesia stock exchange. Journal of Economics, Business, & Accountancy Ventura. 2020 Aug;23(2):226-42.
[11] Setiati S, Azwar MK. COVID-19 and Indonesia. Acta Medica Indonesiana. 2020 Jan 1;52(1):84-9.
[12] Sutantri S, Putri AF, Ismiyati R. “Getting caught unprepared”: A phenomenological study of Indonesian nurses dealing with difficulties when caring for patients with coronavirus disease (COVID-19). Nursing & Health Sciences. 2022 Jun;24(2):469-78.
[13] Masudin I, Safitri NT. Food cold chain in Indonesia during the Covid-19 pandemic: a current situation and mitigation. Jurnal Rekayasa Sistem Industri. 2020 Jul 27;9(2):99-106.
[14] Wulandari LP, Khan M, Probandari A, Batura N, Ferdiana A, Mashuri YA, Wibawa T, Daraninggar D, Dewi BK, Day R, Jan S. “We face the same risk as the other health workers”: Perceptions and experiences of community pharmacists in Indonesia during the COVID-19 pandemic. PLOS global public health. 2022 Jul 1;2(7):e0000606.

[15] Wiranti W, Sriatmi A, Kusumastuti W. Determinan kepatuhan masyarakat Kota Depok terhadap kebijakan pembatasan sosial berskala besar dalam pencegahan COVID-19. Jurnal kebijakan kesehatan indonesia: JKKI. 2020 Sep;9(3):117-24
[16] Ilyas F. Analisis swot kebijakan pembatasan sosial berskala besar (psbb) dan pemberlakuan pembatasan kebijakan masyarakat (ppkm) terhadap dampak ekonomi di tengah upaya menekan laju pandemi covid-19. Akrab Juara: Jurnal Ilmu-ilmu Sosial. 2021 Aug 5;6(3):190-8.
[17] Jamaluddin MM. PONDOK PESANTREN DARUSSALAM DAN IMPLEMENTASI KEBIJAKAN PENANGGULANGAN COVID 19. InPRO SEJAHTERA (Prosiding Seminar Nasional Pengabdian kepada Masyarakat) 2023 (Vol. 5, No. 1).
[18] Candela M, Luconi V, Vecchio A. Impact of the COVID-19 pandemic on the Internet latency: A large-scale study. Computer Networks. 2020 Dec 9;182:107495.
[19] Alita D, Priyanta S, Rokhman N. Analysis of emoticon and sarcasm effect on sentiment analysis of Indonesian language on Twitter. Journal of Information Systems Engineering and Business Intelligence. 2019 Oct;5(2):100-9
[20] Jaman JH, Abdulrohman R, Suharso A, Sulistiowati N, Dewi IP. Sentiment analysis on utilizing online transportation of Indonesian customers using tweets in the normal era and the pandemic covid-19 era with support vector machine. Adv. Sci. Technol. Eng. Syst. 2020;5(5):389-94.
[21] Xue J, Chen J, Hu R, Chen C, Zheng C, Su Y, Zhu T. Twitter discussions and emotions about the COVID-19 pandemic: Machine learning approach. Journal of medical Internet research. 2020 Nov 25;22(11):e20550.
[22] Hussein DM. A survey on sentiment analysis challenges. Journal of King Saud University-Engineering Sciences. 2018 Oct 1;30(4):330-8.
[23] Wankhade M, Rao AC, Kulkarni C. A survey on sentiment analysis methods, applications, and challenges. Artificial Intelligence Review. 2022 Oct;55(7):5731-80.\
[24] Wankhade M, Rao AC, Kulkarni C. A survey on sentiment analysis methods, applications, and challenges. Artificial Intelligence Review. 2022 Oct;55(7):5731-80.
[25] Bensoltane R, Zaki T. Aspect-based sentiment analysis: an overview in the use of Arabic language. Artificial Intelligence Review. 2023 Mar;56(3):2325-63.
[26] Yin H, Song X, Yang S, Li J. Sentiment analysis and topic modeling for COVID-19 vaccine discussions. World Wide Web. 2022 May;25(3):1067-83.
[27] Hidayat TH, Ruldeviyani Y, Aditama AR, Madya GR, Nugraha AW, Adisaputra MW. Sentiment analysis of twitter data related to Rinca Island development using Doc2Vec and SVM and logistic regression as classifier. Procedia Computer Science. 2022 Jan 1;197:660-7.
[28] Hassan MK, Hudaefi FA, Caraka RE. Mining netizen's opinion on cryptocurrency: sentiment analysis of Twitter data. Studies in Economics and Finance. 2022 Apr 22;39(3):365-85.
[29] Babu NV, Kanaga EG. Sentiment analysis in social media data for depression detection using artificial intelligence: a review. SN Computer Science. 2022 Jan;3:1-20.
[30] Sufi FK, Khalil I. Automated disaster monitoring from social media posts using AI-based location intelligence and sentiment analysis. IEEE Transactions on Computational Social Systems. 2022 Mar 18.
[31] Kozik R, Kula S, Choraś M, Woźniak M. Technical solution to counter potential crime: Text analysis to detect fake news and disinformation. Journal of Computational Science. 2022 Apr 1;60:101576.
[32] Andrian B, Simanungkalit T, Budi I, Wicaksono AF. Sentiment analysis on customer satisfaction of digital banking in Indonesia. International Journal of Advanced Computer Science and Applications. 2022;13(3).
[33] Adak A, Pradhan B, Shukla N. Sentiment analysis of customer reviews of food delivery services using deep learning and explainable artificial intelligence: Systematic review. Foods. 2022 May 21;11(10):1500.

[34] Punetha N, Jain G. Game theory and MCDM-based unsupervised sentiment analysis of restaurant reviews. Applied Intelligence. 2023 Mar 31:1-22.

[35] Diekson ZA, Prakoso MR, Putra MS, Syaputra MS, Achmad S, Sutoyo R. Sentiment analysis for customer review: Case study of Traveloka. Procedia Computer Science. 2023 Jan 1;216:682-90.

[36] Ramos CM. Business Intelligence Approach and Sentiment Analysis as a Management Strategy Applied to Study Customer Satisfaction in the Hospitality Sector. InAdvances in Tourism, Technology and Systems: Selected Papers from ICOTTS 2021, Volume 2 2022 May 28 (pp. 537-547). Singapore: Springer Nature Singapore.

[37] Mehmood S, Ahmad I, Khan MA, Khan F, Whangbo T. Sentiment Analysis in Social Media for Competitive Environment Using Content Analysis. Computers, Materials & Continua. 2022 Jun 1;71(3).

[38] Wu L, Dodoo NA, Wen TJ, Ke L. Understanding Twitter conversations about artificial intelligence in advertising based on natural language processing. International Journal of Advertising. 2022 May 19;41(4):685-702.

[39] Ramos CM. Business Intelligence Approach and Sentiment Analysis as a Management Strategy Applied to Study Customer Satisfaction in the Hospitality Sector. InAdvances in Tourism, Technology and Systems: Selected Papers from ICOTTS 2021, Volume 2 2022 May 28 (pp. 537-547). Singapore: Springer Nature Singapore.

[40] Mehmood S, Ahmad I, Khan MA, Khan F, Whangbo T. Sentiment Analysis in Social Media for Competitive Environment Using Content Analysis. Computers, Materials & Continua. 2022 Jun 1;71(3).

[41] Wu L, Dodoo NA, Wen TJ, Ke L. Understanding Twitter conversations about artificial intelligence in advertising based on natural language processing. International Journal of Advertising. 2022 May 19;41(4):685-702.

[42] Fitri VA, Andreswari R, Hasibuan MA. Sentiment analysis of social media Twitter with case of Anti-LGBT campaign in Indonesia using Naïve Bayes, decision tree, and random forest algorithm. Procedia Computer Science. 2019 Jan 1;161:765-72.

[43] Wunderlich F, Memmert D. Innovative approaches in sports science—lexicon-based sentiment analysis as a tool to analyze sports-related Twitter communication. Applied sciences. 2020 Jan 7;10(2):431.

[44] Nahar KM, Jaradat A, Atoum MS, Ibrahim F. Sentiment analysis and classification of arab jordanian facebook comments for jordanian telecom companies using lexicon-based approach and machine learning. Jordanian Journal of Computers and Information Technology. 2020 Sep 1;6(3).

[45] Munoz S, Iglesias CA. A text classification approach to detect psychological stress combining a lexicon-based feature framework with distributional representations. Information Processing & Management. 2022 Sep 1;59(5):103011.

[46] Mashuri M. Sentiment analysis in twitter using lexicon based and polarity multiplication. In2019 International Conference of Artificial Intelligence and Information Technology (ICAIIT) 2019 Mar 13 (pp. 365-368). IEEE.

[47] Schmidt T, Dangel J, Wolff C. Senttext: A tool for lexicon-based sentiment analysis in digital humanities.

[48] Sousa MG, Sakiyama K, de Souza Rodrigues L, Moraes PH, Fernandes ER, Matsubara ET. BERT for stock market sentiment analysis. In2019 IEEE 31st international conference on tools with artificial intelligence (ICTAI) 2019 Nov 4 (pp. 1597-1601). IEEE.

[49] Dey S, Wasif S, Tonmoy DS, Sultana S, Sarkar J, Dey M. A comparative study of support vector machine and Naive Bayes classifier for sentiment analysis on Amazon product reviews.

In2020 International Conference on Contemporary Computing and Applications (IC3A) 2020 Feb 5 (pp. 217-220). IEEE.
[50] Naf'an MZ, Bimantara AA, Larasati A, Risondang EM, Nugraha NA. Sentiment analysis of cyberbullying on instagram user comments. Journal of Data Science and Its Applications. 2019 Apr 12;2(1):38-48.
[51] Patil S, Lokesha V. Live twitter sentiment analysis using streamlit framework. Available at SSRN 4119949. 2022 May 25.
[52] Le CC, Prasad PW, Alsadoon A, Pham L, Elchouemi A. Text classification: Naïve bayes classifier with sentiment Lexicon. IAENG International journal of computer science. 2019 May 27;46(2):141-8.
[53] Rajeswari AM, Mahalakshmi M, Nithyashree R, Nalini G. Sentiment analysis for predicting customer reviews using a hybrid approach. In2020 Advanced Computing and Communication Technologies for High Performance Applications (ACCTHPA) 2020 Jul 2 (pp. 200-205). IEEE.
[54] Al-Shabi M. Evaluating the performance of the most important Lexicons used to Sentiment analysis and opinions Mining. IJCSNS. 2020 Aug;20(1):1.
[55] Bakar MF, Idris N, Shuib L, Khamis N. Sentiment analysis of noisy Malay text: state of art, challenges and future work. IEEE Access. 2020 Jan 23;8:24687-96.
[56] Fesseha A, Xiong S, Emiru ED, Diallo M, Dahou A. Text classification based on convolutional neural networks and word embedding for low-resource languages: Tigrinya. Information. 2021 Jan 25;12(2):52.
[57] Zavarrone E, Forciniti A. CSR & Sentiment Analysis: A New Customized Dictionary. InInternational Conference on Deep Learning Theory and Applications 2023 Jul 13 (pp. 466-479). Cham: Springer Nature Switzerland.
[58] Aggarwal K, Mijwil MM, Al-Mistarehi AH, Alomari S, Gök M, Alaabdin AM, Abdulrhman SH. Has the future started? The current growth of artificial intelligence, machine learning, and deep learning. Iraqi Journal for Computer Science and Mathematics. 2022 Jan 30;3(1):115-23.
[59] Bharadiya JP. Leveraging Machine Learning for Enhanced Business Intelligence. INTERNATIONAL JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY. 2023 Jul 6;7(1):1-9.
[60] Kunnathuvalappil Hariharan N. Predictive model building for driver-based budgeting using machine learning. InPredictive Model Building for Driver-Based Budgeting Using Machine Learning: Kunnathuvalappil Hariharan, Naveen 2021. [Sl]: SSRN.
[61] Zhang Y, Shi X, Zhang H, Cao Y, Terzija V. Review on deep learning applications in frequency analysis and control of modern power system. International Journal of Electrical Power & Energy Systems. 2022 Mar 1;136:107744.
[62] Lee I, Shin YJ. Machine learning for enterprises: Applications, algorithm selection, and challenges. Business Horizons. 2020 Mar 1;63(2):157-70.
[63] Martua, E (2020) Twitter-COVID19-Indonesia-Sentiment-Analysis---Lexicon-Based [Source code]. https://github.com/evanmartua34/Twitter-COVID19-Indonesia-Sentiment-Analysis---Lexicon-Based
[64] Ting SL, Ip WH, Tsang AH. Is Naive Bayes a good classifier for document classification. International Journal of Software Engineering and Its Applications. 2011 Jul 3;5(3):37-46.
[65] Schapire RE, Singer Y, Singhal A. Boosting and Rocchio applied to text filtering. InProceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval 1998 Aug 1 (pp. 215-223).
[66] Gunavathi C, Premalatha K. Performance analysis of genetic algorithm with kNN and SVM for feature selection in tumor classification. International Journal of Computer and Information Engineering. 2014 Sep 3;8(8):1490-7.

[67] Kabiru OA, Owolabi TO, Ssennoga T, Olatunji SO. Performance comparison of SVM and ANN in predicting compressive strength of concrete.
[68] Primartha R, Tama BA. Anomaly detection using random forest: A performance revisited. In2017 International conference on data and software engineering (ICoDSE) 2017 Nov 1 (pp. 1-6). IEEE.
[69] Rennie JD, Shih L, Teevan J, Karger DR. Tackling the poor assumptions of naive bayes text classifiers. InProceedings of the 20th international conference on machine learning (ICML-03) 2003 (pp. 616-623).
[70] Mogotsi IC. Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze: Introduction to information retrieval: Cambridge University Press, Cambridge, England, 2008, 482 pp, ISBN: 978-0-521-86571-5.
[71] Fix E, Hodges JL. Discriminatory analysis. Nonparametric discrimination: Consistency properties. International Statistical Review/Revue Internationale de Statistique. 1989 Dec 1;57(3):238-47.
[72] Cover T, Hart P. Nearest neighbor pattern classification. IEEE transactions on information theory. 1967 Jan;13(1):21-7.
[73] Wu X, Kumar V, Ross Quinlan J, Ghosh J, Yang Q, Motoda H, McLachlan GJ, Ng A, Liu B, Yu PS, Zhou ZH. Top 10 algorithms in data mining. Knowledge and information systems. 2008 Jan;14:1-37.
[74] Bhatia, N., & Vandana. (2010). Survey of Nearest Neighbor Techniques. International Journal of Computer Science and Information Security, 8 (2), 302–305.
[75] Xu S, Wu Y. An algorithm for remote sensing image classification based on artificial immune B-cell network. The international archives of the photogrammetry, remote sensing and spatial information sciences. 2008;37:107-12.
[76] Manne S, Kotha SK, Sameen Fatima S. Text categorization with K-nearest neighbor approach. In Proceedings of the International Conference on Information Systems Design and Intelligent Applications 2012 (INDIA 2012) held in Visakhapatnam, India, January 2012 2012 Jan (pp. 413-420). Berlin, Heidelberg: Springer Berlin Heidelberg.
[77] Geng X, Liu TY, Qin T, Arnold A, Li H, Shum HY. Query dependent ranking using k-nearest neighbor. InProceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval 2008 Jul 20 (pp. 115-122).
[78] Bajramovic, F., Mattern, F., Butko, N., & Denzler, J. (2006). A comparison of nearest neighbor search algorithms for generic object recognition. In Advanced Concepts for Intelligent Vision Systems (pp. 1186–1197). Springer.
[79] Yang Y, Ault T, Pierce T, Lattimer CW. Improving text categorization methods for event tracking. In Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval 2000 Jul 1 (pp. 65-72).
[80] Takahashi K, Yamamoto K, Kuchiba A, Koyama T. Confidence interval for micro-averaged F 1 and macro-averaged F 1 scores. Applied Intelligence. 2022 Mar;52(5):4961-72.
[81] Hand DJ, Christen P, Kirielle N. F*: an interpretable transformation of the F-measure. Machine Learning. 2021 Mar;110(3):451-6.
[82] Grandini M, Bagli E, Visani G. Metrics for multi-class classification: an overview. arXiv preprint arXiv:2008.05756. 2020 Aug 13.
[83] Hossin, M., & Sulaiman, and M. N. (2015). A Review on Evaluation Metrics for Data Classification Evaluations. International Journal of Data Mining & Knowledge Management Process (IJDKP), 5(2), 1–11.
[84] Vujović, Ž. (2021). Classification Model Evaluation Metrics. International Journal of Advanced Computer Science and Applications, 12(6), 599–606. https://doi.org/10.14569/IJACSA.2021.0120670