Indonesians Perception on the South China Sea Dispute: Support Vector Machine and Naïve Bayes Approach

Adinda Aulia Hafizha[1)*], Nurfarah Nidatya[2)]
[1,2)]Universitas Pembangunan Nasional Veteran Jakarta, Indonesia
Jalan RS. Fatmawati Raya, Pondok Labu, Cilandak, South Jakarta City, Jakarta 12450
[1)]hafizhadinda@gmail.com, [2)]nurfarahnidatya@upnvj.ac.id

Abstract: In recent years, relations between Indonesia and China have become increasingly cordial. However, a potential source of tension is emerging in the form of a heightened dispute in the South China Sea. The government of Indonesia is considered an ally, however there has been a long-standing negative opinion among Indonesians regarding China, which has influenced the way both the general public and the political elite have perceived the relations between Indonesia and China. This research has two objectives. The first is to examine Indonesian perceptions regarding the South China Sea conflict. The second is to compare the performance of Support Vector Machine (SVM) and Multinomial Naïve Bayes as a method of sentiment analysis. Using 7.051 Indonesian-language posts from social media X as a dataset, the result shows that a significant portion of Indonesians view the dispute negatively, fearing potential escalation and threats to national security. Despite these concerns, there is reason to believe that Indonesia can play a proactive role in resolving the conflict through ASEAN and UNCLOS frameworks. Meanwhile, SVM has been demonstrated to be an effective method for handling sentiment analysis data, achieving an accuracy of 87.95%. This work contributes to the field of sentiment analysis by highlighting social media as a valuable platform and by demonstrating the effectiveness of SVM. Furthermore, the study offers new insights for the field of international relations by analyzing the South China Sea dispute through a machine learning lens, which may lead to the development of novel perspectives.

Keywords: Indonesia; multinomial naïve bayes; sentiment analysis; south china sea; support vector machine

INTRODUCTION

The international order, which has been dominated by the United States of America (USA) since the end of the Cold War, is currently undergoing a period of transition as the 21st century progresses towards its third decade. The principal transformation of the global order is evidenced by a decline in the economic influence of the USA and the concurrent ascension of People's Republic of China (hereby referred to as China) as a dominant economic power (Han & Paul, 2020). The changing landscape of the international order has led the USA to adopt a markedly more critical stance toward China. In response, Washington has introduced its own strategy, the "free and open Indo-Pacific," to counterbalance China's Belt and Road Initiative, also known as BRI foreign policy (Liu, 2020). In contrast to the United States' view on the matter, Brazil, the European Union, Iran, Pakistan and Russia
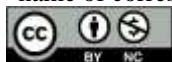
considered China to be their comprehensive strategic partners (Papageorgiou & de Melo, 2022). The divergent perspectives held by these nations demonstrate how various individuals and places have interpreted China's emergence.

The perceptions of China throughout Southeast Asia appear to be influenced by the countries' relations—particularly their cultural and economic affinities—with the Asian giant. For instance, Indonesia, which has been described as a rising power that supports East Asia's economic revival led by China, has been perceived as an ally with similar capabilities and cultural dimensions to China (Kartikasari, 2019; Pramono & Raharjo, 2024). In recent years, relations between Indonesia and China have become increasingly cordial, particularly during the presidency of Joko Widodo. The joint venture on the Jakarta-Bandung High-Speed Railway and China's vaccine diplomacy has fostered a robust relationship to the extent that President Joko Widodo has described China as "good friends and brothers" (Grossman, 2021).

However, as China and Indonesia strengthen their relationship, a potential source of tension is emerging in the form of a heightened dispute in the South China Sea. As a country with a geographical proximity to the disputed waters, Indonesia serves as a mediator, facilitating informal communication links between conflict actors with the objective of identifying issues, lowering tensions, and encouraging direct interaction in the form of negotiations (Laksmi et al., 2022). The aforementioned statement offers insight into the manner in which the Indonesian government responds to the South China Sea dispute. However, there is a notable absence of research on the Indonesian public's perception of the South China Sea dispute, particularly in the context of China's rising influence. This is an intriguing topic for study, as there has been a long-standing divergence of opinion among Indonesians regarding China, which has influenced the way both the general public and the political elite have perceived the relations between Indonesia and China (Fitriani, 2021).

The study of Indonesian sentiments towards China is a relatively new field of inquiry. Herlijanto's research indicates that there was a prevailing public skepticism about the potential impact of China's rise on Indonesia (Herlijanto, 2017). This perspective is also supported by Yeremia (2022), who posited that the dominant perceptual trend among Indonesian diplomats and scholars is to perceive China as a threat rather than an opportunity. Meanwhile, Indonesian elites regard Beijing as assertive, a perception that is shaped by the prevalence of negative perceptions of China (Herlijanto, 2022). Along the same lines, Fitriani (2018) postulates that Indonesian perceptions of China are multifaceted. The Indonesian military exhibits a consistently negative perception of China, in contrast to scholars and elites, who hold diverse views. Moreover, the involvement of China in the South China Sea dispute has prompted Indonesian public officials to engage in discourse on social media X (formerly Twitter) (Fajrina et al., 2020). In consideration of the existing research gap and the inherently intriguing topic, it is this study's first objective to examine Indonesian perceptions regarding the South China Sea conflict.

In order to enhance the novelty of our research, we will employ machine learning as a method of analysis. The machine learning techniques that will be employed in this research are Naïve Bayes and Support Vector Machine (SVM). The rationale behind selecting the aforementioned techniques is their robustness and the distinct strengths exhibited by each. Naïve Bayes is a more straightforward and expeditious training technique that demonstrates efficacy even with smaller datasets. In contrast, SVM has the potential to achieve higher accuracy, particularly when analyzing complex data with non-linear relationships. A number of studies have compared Naïve Bayes and SVM with social media X data. The results of these studies have been inconclusive, with some indicating that SVM has the highest accuracy (Andrian et al., 2022; Chong & Shah, 2022; Dey et al., 2020), and others suggesting that Naïve Bayes has the highest accuracy (Pradipta & Jayadi, 2022). The second objective of this research is to compare the performance of SVM and Naïve Bayes with social media X data of Indonesian sentiments towards the South China Sea dispute. This research will contribute to the development of machine learning literature and provide insights into the relationship between Indonesia and China from an individual level. In addition, the practical implications of this research aim to uncover the underlying concerns and hopes within the population. This can assist policymakers in anticipating potential sources of social unrest or public pressure related to the dispute.

METHOD

Data Acquisition and Preprocessing

This research uses posts from X as the dataset, ranging from 1 January 2022 to 30 April 2024. In order to ascertain the sentiment of the Indonesian population, a limitation was imposed on the data search, whereby only posts written in Indonesian were collected. Figure 1 delineates the methodology employed in this research study. It is based on the general framework for analyzing textual data and includes an additional step of comparing Naïve Bayes and SVM (Chong & Shah, 2022).
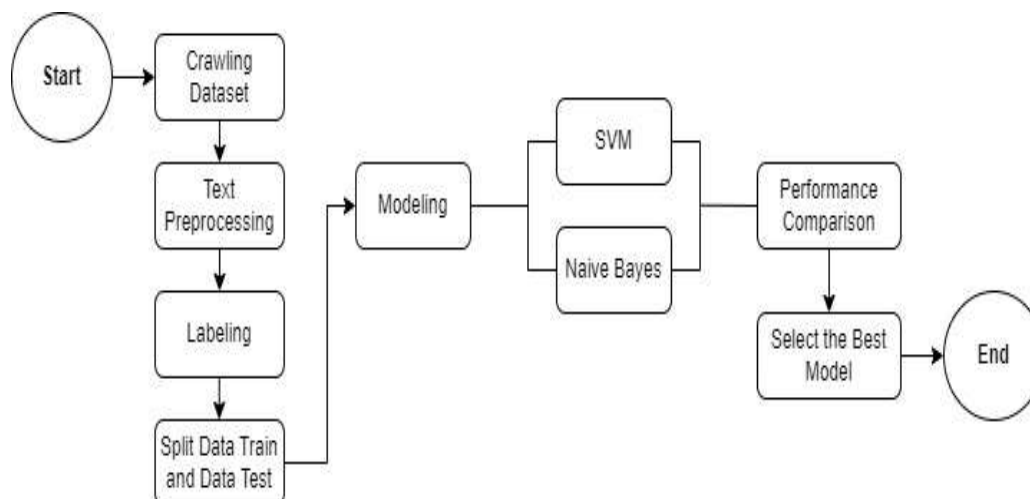

Fig. 1 Research Flowchart of Sentiment Analysis with Machine Learning Comparison.

All commands in this research were executed using the Python programming language, version 3.10.12. Next, by employing the search term "*sengketa laut china selatan*" and utilizing the Tweepy library, we successfully retrieved 8.432 posts. After the data had been extracted from social media X, it was subjected to further processing in the form of drop duplicate posts, text cleaning (cleaned of hashtags, Uniform Resource Locators (URLs), numbers, symbols, emojis, usernames, username tags, and other characters that are not relevant to the expressed opinions), case folding (converting all capital letters to lowercase to avoid case sensitivity), stopword removal (exclamation, question marks, commas, and periods will be removed from the data), and stemming (converting words with affixes into basic words). This involved the use of regular expressions (RE) to identify and eliminate any irregular characters, as well as the NLTK tokenize function to split up text into individual words. Following the removal of duplicate posts, the original dataset of 8.432 was reduced to 7.051. After the preprocessing step, the dataset was duplicated for two distinct purposes. The first was for the independent analysis of sentiments, while the second was for the experimentation with the models. For machine learning models, the preprocessed dataset was divided into training data and test data with a ratio of 80:20.

Data Labeling and Model Construction

Following the data preprocessing step, the train dataset was divided into two categories: positive and negative. This was achieved through the use of a lexicon-based method. The corpus was sourced from the Indonesian Sentiment Lexicon (InSet Lexicon), a list of words collected from Indonesian tweets that was developed to identify and categorize written opinions into positive or negative sentiment (Koto & Rahmaningtyas, 2018). Once the trained data has been annotated with the InSet Lexicon, the next step is to implement the vectorization stage with Term Frequency-Inverse Document Frequency (TF-IDF). TF-IDF is widely regarded as the most robust feature extraction technique ever employed for clustering and classifying textual data (Akuma et al., 2022; Cahyani & Patasik, 2021; Patil & Kolhe, 2022). Upon executing command 'from sklearn.feature_extraction.text import TfidfVectorizer', the TF-IDF vectorizer was employed with the following parameters: max_features (2000), min_df (5), max_df (0.7), stopwords.words ('Indonesian'), and ngram_range (1.3). Subsequent to the vectorization step, the research process will proceed with the construction of the model.
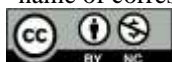
Classification models were constructed using SVM and Naive Bayes. SVM is an algorithm that employs a hyperplane to classify data, thereby creating a margin that separates each predefined dataset into two groups in a higher-dimensional space (Styawati et al., 2021). Its ability to predict polarity from text has made it a valuable tool in this domain. Unlike other algorithms, SVM is not constrained by the curse of dimensionality, a phenomenon that hinders the performance of algorithms when faced with large feature spaces (Hsu, 2020). This characteristic allows SVM to perform well on similar-sized data, regardless of the number of features. Figure 2 is an illustrative example of an SVM classifier:
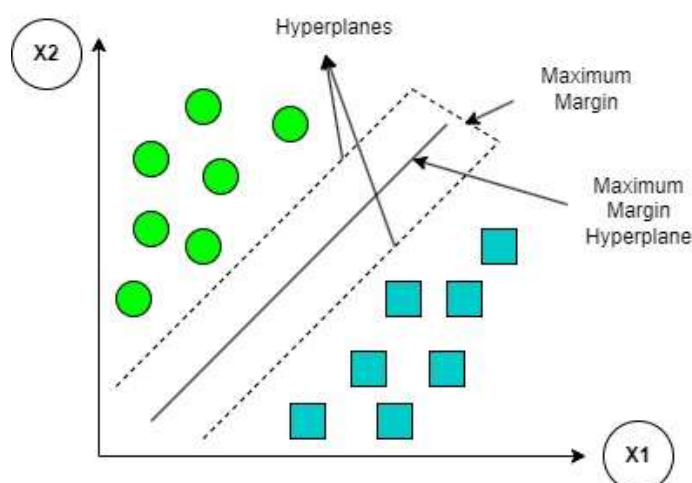

Fig. 2 Illustration of Support Vector Machine Classifier.

In comparison, Naive Bayes is based on the well-known Bayes theorem and is one of the most prominent probabilistic classification techniques utilized (Wickramasinghe & Kalutarage, 2021). The efficiency and resilience of the algorithm are significant factors in Naive Bayes' appeal, in addition to its simplicity (Arar & Ayan, 2017). The posterior probability that an item falls into a specific class is determined by the Naive Bayes algorithm using the feature values of the data. Equation 1 illustrates the manner in which this probability is calculated. In the equation, c represents the class, $P(c)$ denotes the prior probability of the class, $P(x|c)$ is the likelihood of observing the data within class c, $P(x)$ is the evidence and x is the data (Hsu, 2020):

$$\text{Posterior Probability} = P(c|x) = \frac{P(C)P(X|C)}{P(x)} \qquad (1)$$

The researcher has determined that SVM and Naive Bayes are the most suitable algorithms for identifying polarity in text, given their robustness and efficiency in textual classification, respectively. It is anticipated that the application of these algorithms will result in an accuracy rate of approximately 85%-90% in the sentiment classification cases, thereby enabling an accurate and comprehensive understanding of the perceptions of Indonesians on the South China Sea dispute. In addition, both SVM and Naive Bayes have been considered superior algorithms in sentiment analysis, and it is interesting to see which algorithm performs better when competing with the same dataset and feature extractions.Thus, this research made use of the scikit-learn libraries, executing command "from sklearn.svm import SVC" and "from sklearn.naive_bayes import MultinomialNB" to implement the SVM and Naive Bayes classifier. It is worth noting that this research employed the well-known parameters of SVM and Multinomial Naive Bayes on sentiment analysis cases, which are SVC(kernel='linear', c=1.0) and MultinomialNB(), respectively (Damanik & Setyohadi, 2021; Prastyo et al., 2020).

Once the model has been constructed, the subsequent step is to validate the models in order to ascertain their optimal performance. This is achieved through the use of evaluation metrics, which denote the classification process outcomes in four terms: accuracy, recall, precision, and F1 score. In

order to obtain scores for the evaluation metrics, each of the aforementioned terms was previously calculated using the values of True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). The calculations for the evaluation metrics are shown in the equations below:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \; x \; 100\% \tag{2}$$

$$\text{Recall} = \frac{TP}{FN + TP} \; x \; 100\% \tag{3}$$

$$\text{Precision} = \frac{TP}{FP + TP} \; x \; 100\% \tag{4}$$

$$\text{F1 Score} = 2 \; x \; \frac{Precision \; x \; Recall}{Precision + Recall} \tag{5}$$

RESULT

The initial research question sought to analyze Indonesian sentiments expressed on social media platform X in relation to the South China Sea dispute. The findings indicate that the majority of Indonesian social media users who commented on X have a negative perception of the South China Sea dispute. This is corroborated by empirical evidence in the form of emotion classification in related posts. Of the 7.051 posts classified, the majority (5.637 or 79.9%) contained negative sentiments, while the remainder (1.414 or 20.1%) were filled with positive sentiments.   To facilitate comprehension of the obtained results, we employ the R software's GGplot package for data visualization and eliminate the term "*laut china selatan*" (south china sea) to identify optimal results in related terms associated with the sentiment analysis conducted. The GGplot Wordcloud2 tool was employed to separately visualize the positive and negative sentiments, with the objective of gaining a more nuanced understanding of how Indonesians perceived the longstanding maritime dispute. Below, Figure 3 presents the most frequently occurring words in the context of positive sentiments.

Fig. 3 Wordcloud of Most Common Words in Positive Sentiments.

As illustrated in the graph above, the most prevalent words were "indonesia," "*sengketa*" (dispute), and "*sejahtera*," which can be translated as "welfare". The next most prevalent words were "*penyelesaian*" (settlement), "prabowo" (the Indonesian Minister of Defense), and "*negara*" (country). It is noteworthy that the sentiments of Indonesians extended to two international institutions, namely the Association of Southeast Asian Nations (ASEAN) and the United Nations Convention on the Law of the Sea (UNCLOS). Furthermore, the names of two countries involved in the South China Sea dispute were identified: the Philippines and Malaysia. It is notable that the results included words related to the Indonesian military, namely "tni" (Indonesian National Army), "*garda*" (garrison), and "*pertahanan*" (defense). Furthermore, these words were preceded by the terms "*asa*" (hope), "*hebat*" (great), and

"*solusi*" (solution) which suggests a correlation between the sentiment expressed in the results and the prevailing positive tone in Indonesian discourse.

We now turn to a further analysis of the sentiments of the Indonesian people in relation to the South China Sea dispute. Figure 4 presents an overview of the negative sentiments held towards this dispute. From the data displayed, it can be seen that the words used to convey negative sentiment carry bad connotations. This can be observed by noting the frequency of words such as "*konflik*" (conflict) and "*isu*" (issue). Of interest in this analysis are the terms "*natuna,*" *utara*" (north), and "*selat*" (strait), which refer to locations in Indonesia and its territorial waters. Furthermore, the data reveals that negative sentiment extends to the terms used to refer to countries and continents. In this context, the terms "*amerika*" (America), "Russia", and "Asia" appear. The relationship between these terms and other words in this context will be discussed in more detail in a later section.

Fig. 4 Wordcloud of Most Common Words in Negative Sentiments.

Moving on to the second research objective, which is to compare the performance of SVM and Naïve Bayes with social media X data of Indonesian sentiments towards the South China Sea dispute. In order to ascertain the efficacy of the constructed model, the performance metrics are sought using the following Python command: "from sklearn.metrics import classification_report, confusion_matrix, accuracy_score, recall_score, precision_score, f1_score." The output of the command is then packaged into Table 1. As illustrated in the table below, the SVM model exhibits a higher degree of accuracy than the Multinomial Naïve Bayes model, with an accuracy value of 87.95%. In contrast, the Multinomial Naïve Bayes model attained an accuracy of only 84.83%. In addition to the aforementioned accuracy value, the SVM model also demonstrated superior performance in terms of Precision, Recall, and F1 Score.

Table 1
Model Performance Metrics

Models	Accuracy	Precision	Recall	F1 Score
Support Vector Machine	**87.95%**	**85.52%**	**72.76%**	**76.77%**
Multinomial Naïve Bayes	84.83%	67.66%	41.39%	51.36%

To demonstrate the high degree of accuracy exhibited by the models, we employed both the SVM and Multinomial Naïve Bayes models to predict the sentiment of a novel sentence. As a case in point, the sentence in question is "*Indonesia mampu untuk menyelesaikan sengketa Laut China Selatan*" (Indonesia is able to resolve the South China Sea dispute) which is of a positive tone. The results predicted by both models indicate that the output is indeed positive. The results of this experiment are presented in Figure 5 below. In summary, the sentiment analysis models developed demonstrated an effective outcome. Further correlation and analysis of the results will be conducted in the Discussion section.

(a)

(b)

Fig. 5 Sentiments Prediction Model: (a) Multinomial Naïve Bayes; (b) SVM.

DISCUSSIONS

Indonesian Perceptions on The South China Sea Dispute

Previous research on Indonesian perceptions of China has yielded consistent findings, namely public skepticism regarding the rise of China, which is perceived as detrimental to local communities. In fact, some Indonesian diplomats and students view China as a threat rather than an opportunity for cooperation. This negative view is also shared by Indonesia's military, which has begun to be wary and cautious as China begins to strengthen its influence in the South China Sea region. To expand the scope of knowledge regarding Indonesian views, this research endeavors to capture the sentiments of Indonesian civil society expressed through X social media, with a specific focus on the South China Sea dispute. The results indicate that the majority of individuals hold a negative view of the maritime conflict.

The Indonesians express negative sentiment regarding the impact of the South China Sea conflict on the territorial integrity of Indonesia. This is evidenced by the co-occurrence of the terms "conflict" and "North Natuna" in the same context. The North Natuna Sea is the name given to the area of Indonesia's Exclusive Economic Zone (EEZ) directly adjacent to the South China Sea. It is located in the north of Natuna Regency, Riau Islands. This naming was undertaken as a strategic measure by Indonesia to maintain its sovereignty in the region (Purba & Burhanuddin, 2023). It is, of course, possible that if the North Natuna Sea is once again associated with negative sentiments, people may feel afraid that the escalation of conflict in the South China Sea will endanger the North Natuna and potentially threaten Indonesia's sovereignty. In light of the prevailing concerns regarding Indonesia's defense capabilities, the public has turned its attention to the military, presuming that the primary role of the Indonesian National Army (TNI) is to safeguard the territorial integrity of the country. This implies that the Indonesian military is expected to maintain a heightened state of alert and vigilance with respect to any potential security challenges in the South China Sea. Consequently, the term *"tni"* has been associated with negative sentiment.
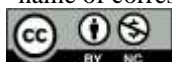
Moreover, it is noteworthy that a number of countries, including Russia and the United States, have also emerged as sources of negative sentiment regarding the South China Sea conflict. The explanation for this finding may be attributed to the intervention of the two superpowers in the South China Sea conflict. This is evidenced by the numerous military exercises conducted by Russia and the United States in the maritime conflict zone (Christiastuti, 2024). Moreover, the United States' actions extend beyond military exercises. It also provides military assistance to its allies in the Asia-Pacific region, including Taiwan and the Philippines. It is evident that this military intervention has resulted in the creation of new tensions within the South China Sea region, with the potential for these tensions to increase the conflict level. This atmosphere is indirectly reflected in the sentiments of the Indonesian people, who express concern about the preparedness of Indonesia's security and defense in the event of a conflict in the South China Sea. Overall, the negative sentiment observed is consistent with the results of previous research indicating that the Indonesian people perceive China as a threat (Fitriani, 2018; Herlijanto, 2017, 2022; Yeremia, 2022). To enrich the related literature, this research finding contributes to the discourse that Indonesians hold a negative attitude towards China due to concerns about Indonesia's sovereignty and security, particularly in the context of the South China Sea conflict.

Meanwhile, the prevailing positive sentiment towards the South China Sea conflict also encompasses the expectations of the Indonesian people regarding the country's role in the conflict. This is further reinforced by the use of words such as "Indonesia" (which emphasizes Indonesia's role as a friendly country), "sengketa" (which refers to the South China Sea conflict), and "sejahtera" (which indicates that welfare is a priority). If the narratives are considered collectively, it can be inferred that the Indonesian people desire their country to play an active role in resolving the South China Sea conflict for the benefit of Indonesia's welfare. The expectation that Indonesia will play an active role in maintaining peace is not a novel concept. Rather, it is an extension of the principle of Indonesian foreign policy, which has been in place for several decades and is known as "Bebas Aktif" (free and active foreign policy). Moreover, Indonesia has been entrusted by Joe Biden, President of the United States, with the role of mediator in the South China Sea conflict (Guinto, 2023). In fact, China, which is known for its assertive stance in disputed territories, regards Indonesia as an ally (Kartikasari, 2019). This knowledge undoubtedly influences the sentiments of the Indonesian people towards the conflict.

In addition, there are other noteworthy aspects of Indonesia's positive sentiment, including the emergence of the terms "Prabowo" and "settlement." The use of these two words may be indicative of Mr. Prabowo Subianto, who has served as the Minister of Defense of the Republic of Indonesia from 2019 until the present research was conducted. The concern for this figure also represents the Indonesian people's high expectation that the Ministry of Defense of the Republic of Indonesia will be alert and active in maintaining Indonesia's security and defense. It may be anticipated that the Indonesian Ministry of Defense will be able to contribute to the resolution of the protracted dispute. Furthermore, if we align the aforementioned expectations with the anticipated timeline of political events occurring in Indonesia at the end of 2023 to 2024, it can be expected that the Indonesian people will anticipate that Mr. Prabowo will adopt a proactive stance in the South China Sea conflict as president for the 2024-2029 period.

In addition to positive sentiment, the words "UNCLOS" and "ASEAN" can also be observed in the context of this discourse. This undoubtedly pertains to the endeavors to resolve the South China Sea dispute. Thus far, Indonesia has consistently prioritized the 1982 UNCLOS international law of the sea in asserting its EEZ territory (Sulistyani et al., 2021). It is therefore unsurprising that the Indonesian people expect the state to adhere consistently to the principles set out in UNCLOS and, where appropriate, to utilize this legal instrument in order to find a resolution to the dispute. Of equal importance is ASEAN, which the Indonesian people hope will act as a mediator in the maritime conflict. This can be inferred from the role that ASEAN has played thus far as an organization of Southeast Asian countries, namely as a mediator between the parties to the conflict by opening various discussion rooms to resolve the South China Sea dispute (Hu, 2023). Moreover, during Indonesia's tenure as chair of ASEAN, the negotiation process of the South China Sea Code of Conduct, initiated in 2002 with ASEAN countries and China, entered a new phase with the commencement of the second reading process (Kementerian Luar Negeri Republik Indonesia, 2023). This subsequently established a new basis of trust between ASEAN and UNCLOS in resolving the South China Sea conflict.
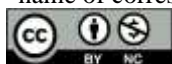
In conclusion, the negative sentiment of the Indonesian people is driven by concerns about the potentially escalating conflict in the South China Sea, which threatens Indonesia's security, sovereignty, and territorial integrity. Nevertheless, amidst the prevailing unrest, there persists a *fiducia* that Indonesia, and in particular Mr. Prabowo Subianto, will adopt a proactive stance in resolving the South China Sea conflict. This role can be fulfilled through ASEAN and in accordance with the 1982 UNCLOS.

Constructed Model: Naïve Bayes and SVM

The second objective of this research is to compare the performance of SVM and Naïve Bayes with Indonesian social media data on sentiment towards the South China Sea dispute. The study found that the best model for the collected dataset outperformed Naïve Bayes in all key evaluation metrics. This is evidenced by better Accuracy, Precision, Recall, and F1 Score values. This advantage likely stems from SVM's ability to handle complex relationships between features in text data. Sentiment regarding the South China Sea dispute can be influenced by a multitude of factors beyond individual words, including historical references, nationalistic pride, and economic concerns. SVM is capable of capturing this complexity by identifying a hyperplane that optimally separates sentiment classes with maximum margin, even in a high-dimensional feature space where these intricate relationships become more apparent (Naw & Mon, 2018). This suggests that for sentiment analysis of Indonesian social media data on this sensitive topic, SVM offers a more robust approach compared to Naïve Bayes. Consequently, this research supports the findings of previous studies that compared the performance of SVM and Naïve Bayes on the social media X posts dataset, where the results demonstrated that SVM was the most effective model for sentiment analysis (Andrian et al., 2022; Chong & Shah, 2022; Dey et al., 2020).

However, it should be noted that the dataset is not without limitations. Despite the study's significant findings, it is important to consider the limitations of the study when interpreting the results. To begin, we gathered real-time information from a limited number of users. It would be beneficial to replicate the study across multiple social media platforms in order to further examine the robustness of these findings. A further limitation of this research is the imbalanced nature of the dataset, with a greater number of negative sentiments than positive. This could result in a biased model, where the model prioritizes learning the patterns of the majority class. This may result in inaccurate classification of the minority class. Consequently, future research should employ a balanced dataset, conduct further experiments with a different case study, and explore alternative machine learning or deep learning algorithms with the objective of achieving the highest accuracy and developing a more effective sentiment analysis model.

This research, which employs machine learning to analyze Indonesian social media data, offers insights that are valuable for understanding public sentiment on the South China Sea dispute. The study underscores the value of SVM over Naive Bayes by demonstrating its superiority in modeling complex relationships within the data. This indicates that Indonesian sentiment encompasses more than a simple dichotomy of positive or negative reactions. It incorporates factors such as historical context and economic concerns. These findings can inform policymakers and researchers of the nuanced public opinion surrounding the dispute, allowing for more targeted strategies to address the issue.

CONCLUSION

The majority of Indonesians perceive the South China Sea dispute in a negative light. This negative sentiment is driven by concerns about the potential for conflict in the South China Sea to escalate, which could have a detrimental impact on Indonesia's security, sovereignty, and territorial integrity. In contrast, despite the prevailing unrest, there persists hope that Indonesia, and particularly Mr. Prabowo Subianto (as the current Minister of Defense and the elected future president of 2024-2029), will adopt a proactive stance in resolving the South China Sea conflict. This role can be fulfilled through ASEAN and in accordance with the 1982 UNCLOS. In terms of model construction, the SVM has been proven to be the most effective model for the X posts dataset, exhibiting an accuracy value of 87.95%. This exemplary performance may be attributed to the SVM hyperplane's ability to optimally separate each dataset class, which is a significant advantage over other models. In contrast, the Naïve Bayes model assumes independence between features, a premise that may not always hold true in sentiment analysis cases. In conclusion, our work contributes to the current research by advancing the theoretical
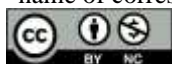
understanding of X as a channel for sentiment analysis. Furthermore, this research provides new insights to the field of International Relations (IR), with the convergence of big data and the traditional issues of IR (the South China Sea dispute), which could potentially generate new perspectives. In addition to the theoretical contributions, this work also contributes to current practice by providing a framework for machine learning comparison in order to find the best model for sentiment analysis research. However, limitations such as imbalanced and homogeneous datasets occurred, as the data were only collected from one source, which was X. In view of the above, future research should consider experimenting with other machine learning or deep learning algorithms, trying out different feature extractions, and creating a more balanced dataset by exploring additional sentiment analysis cases.

REFERENCES

Akuma, S., Lubem, T., & Adom, I. T. (2022). Comparing Bag of Words and TF-IDF with different models for hate speech detection from live tweets. *International Journal of Information Technology*, *14*(7), 3629–3635. https://doi.org/10.1007/s41870-022-01096-4

Andrian, B., Simanungkalit, T., Budi, I., & Wicaksono, A. F. (2022). Sentiment Analysis on Customer Satisfaction of Digital Banking in Indonesia. *International Journal of Advanced Computer Science and Applications*, *13*(3), 466–473. https://doi.org/10.14569/IJACSA.2022.0130356

Arar, Ö. F., & Ayan, K. (2017). A feature dependent Naive Bayes approach and its application to the software defect prediction problem. *Applied Soft Computing*, *59*, 197–209. https://doi.org/10.1016/j.asoc.2017.05.043

Cahyani, D. E., & Patasik, I. (2021). Performance comparison of tf-idf and word2vec models for emotion text classification. *Bulletin of Electrical Engineering and Informatics*, *10*(5), 2780–2788. https://doi.org/10.11591/eei.v10i5.3157

Chong, K. S., & Shah, N. (2022). Comparison of Naive Bayes and SVM Classification in Grid-Search Hyperparameter Tuned and Non-Hyperparameter Tuned Healthcare Stock Market Sentiment Analysis. *International Journal of Advanced Computer Science and Applications*, *13*(12), 90–94. https://doi.org/10.14569/IJACSA.2022.0131213

Christiastuti, N. (2024). Kapal Perang Rusia Gelar Latihan Militer di Laut China Selatan. *DetikNews*. https://news.detik.com/internasional/d-7166445/kapal-perang-rusia-gelar-latihan-militer-di-laut-china-selatan

Damanik, F. J., & Setyohadi, D. B. (2021). Analysis Of Public Sentiment About Covid-19 In Indonesia On Twitter Using Multinomial Naive Bayes And Support Vector Machine. *IOP Conference Series: Earth and Environmental Science*, *704*(1), 012027. https://doi.org/10.1088/1755-1315/704/1/012027

Dey, S., Wasif, S., Tonmoy, D. S., Sultana, S., Sarkar, J., & Dey, M. (2020). A Comparative Study of Support Vector Machine and Naive Bayes Classifier for Sentiment Analysis on Amazon Product Reviews. *2020 International Conference on Contemporary Computing and Applications (IC3A)*, 217–220. https://doi.org/10.1109/IC3A48958.2020.233300

Fajrina, A. N., Roziqin, A., & Sihidi, I. T. (2020). STUDI GEOPOLITIK LAUT CHINA SELATAN: DATA DAN ANALISIS MEDIA SOSIAL (GEOPOLITICAL STUDIES OF THE SOUTH CHINA SEA: DATA AND ANALYSIS OF SOCIAL MEDIA). *Jurnal Lembaga Ketahanan Nasional Republik Indonesia*, *8*(2), 115–130.

Fitriani, E. (2018). Indonesian perceptions of the rise of China: dare you, dare you not. *The Pacific Review*, *31*(3), 391–405. https://doi.org/10.1080/09512748.2018.1428677
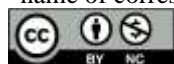
Fitriani, E. (2021). Linking the impacts of perception, domestic politics, economic engagements, and the international environment on bilateral relations between Indonesia and China in the onset of the 21 st century. *Journal of Contemporary East Asia Studies*, *10*(2), 183–202. https://doi.org/10.1080/24761028.2021.1955437

Grossman, D. (2021). Indonesia Is Quietly Warming Up to China. *Foreign Policy*. https://foreignpolicy.com/2021/06/07/indonesia-china-jokowi-natuna-sea-military-bri-cooperation-biden-united-states/

Guinto, J. (2023). South China Sea: Biden says US will defend the Philippines if China attacks. *BBC News*. https://www.bbc.com/news/world-asia-67224782

Han, Z., & Paul, T. V. (2020). China's Rise and Balance of Power Politics. *The Chinese Journal of International Politics*, *13*(1), 1–26. https://doi.org/10.1093/cjip/poz018

Herlijanto, J. (2017). Public Perceptions of China in Indonesia : The Indonesia National Survey. *ISEAS Yusof Ishak Institute*, *2017*(89), 1–11. http://www.pewglobal.org/database/indicator/1/country/101/%0Ahttps://www.iseas.edu.sg/images/pdf/ISEAS_Perspective_2017_89.pdf

Herlijanto, J. (2022). Indonesian Elite Perception of China During the Presidency of Joko (Jokowi) Widodo. In T. Tudoroiu & A. Kuteleva (Eds.), *China in the Global South* (pp. 167–190). Springer Nature Singapore. https://doi.org/10.1007/978-981-19-1344-0_9

Hsu, B. M. (2020). Comparison of supervised classification models on textual data. *Mathematics*, *8*(5). https://doi.org/10.3390/MATH8050851

Hu, L. (2023). Examining ASEAN's effectiveness in managing South China Sea disputes. *The Pacific Review*, *36*(1), 119–147. https://doi.org/10.1080/09512748.2021.1934519

Kartikasari, A. (2019). Indonesia's Image from China's Perspective on South China Sea Dispute (A Preliminary Study on China's Perception on Indonesia). *Global: Jurnal Politik Internasional*, *21*(2), 176. https://doi.org/10.7454/global.v21i2.404

Kementerian Luar Negeri Republik Indonesia. (2023, September 1). Inisiatif Indonesia Percepat Negosiasi Kode Etik di Laut Tiongkok Selatan. *Kementerian Luar Negeri Republik Indonesia*. https://kemlu.go.id/portal/id/read/5153/berita/inisiatif-indonesia-percepat-negosiasi-kode-etik-di-laut-tiongkok-selatan

Koto, F., & Rahmaningtyas, G. Y. (2018). Inset lexicon: Evaluation of a word list for Indonesian sentiment analysis in microblogs. *Proceedings of the 2017 International Conference on Asian Language Processing, IALP 2017*, *2018-Janua*, 391–394. https://doi.org/10.1109/IALP.2017.8300625

Laksmi, L. G. C. S., Mangku, D. G. S., & Yuliartini, N. P. R. (2022). PERAN INDONESIA DALAM PENYELESAIAN SENGKETA INTERNASIONAL DI LAUT CINA SELATAN. *Jurnal Komunitas Yustisia*, *5*(2), 225–242. https://doi.org/10.23887/jatayu.v5i2.51616

Liu, F. (2020). The recalibration of Chinese assertiveness: China's responses to the Indo-Pacific challenge. *International Affairs*, *96*(1), 9–27. https://doi.org/10.1093/ia/iiz226

Naw, N., & Mon, A. C. (2018). Social media data analysis in sentiment level by using support vector machine. *Journal of Pharmacognosy and Phytochemistry*, *7*(1S), 609–613.

*name of corresponding author

Papageorgiou, M., & de Melo, D. da S. N. (2022). China as a Responsible Power Amid the COVID-19 Crisis: Perceptions of Partners and Adversaries on Twitter. *Fudan Journal of the Humanities and Social Sciences*, *15*(2), 159–188. https://doi.org/10.1007/s40647-022-00344-y

Patil, R. S., & Kolhe, S. R. (2022). Supervised classifiers with TF-IDF features for sentiment analysis of Marathi tweets. *Social Network Analysis and Mining*, *12*(1), 51. https://doi.org/10.1007/s13278-022-00877-w

Pradipta, R., & Jayadi, R. (2022). The Sentiment Analysis of the Indonesian Palm Oil Industry in Social Media Using a Machine Learning Model. *Journal of Theoretical and Applied Information Technology*, *100*(12), 4532–4542.

Pramono, S., & Raharjo, W. A. (2024). Indonesia's Role in the Rise of East Asia amid China's Ascendancy. *Nation State: Journal of International Studies*, *6*(2), 139–157. https://doi.org/10.24076/nsjis.v6i2.1324

Prastyo, P. H., Ardiyanto, I., & Hidayat, R. (2020). Indonesian Sentiment Analysis: An Experimental Study of Four Kernel Functions on SVM Algorithm with TF-IDF. *2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy, ICDABI 2020*. https://doi.org/10.1109/ICDABI51230.2020.9325685

Purba, J. A. A., & Burhanuddin, A. (2023). ENCOURAGE FROM FEAR: PERUBAHAN NAMA LAUT NATUNA UTARA SEBAGAI UPAYA INDONESIA DALAM MEMPERTAHANKAN KEDAULATAN NEGARA. *Jurnal Ilmiah Multidisiplin*, *2*(6), 13–20. https://doi.org/10.56127/jukim.v2i6.987

Styawati, S., Isnain, A. R., Hendrastuty, N., & Andraini, L. (2021). Comparison of Support Vector Machine and Naïve Bayes on Twitter Data Sentiment Analysis. *Jurnal Informatika: Jurnal Pengembangan IT*, *6*(1), 56–60. https://doi.org/10.30591/jpit.v6i1.3245

Sulistyani, Y. A., Pertiwi, A. C., & Sari, M. I. (2021). Indonesia's Responses amidst the Dynamic of the South China Sea Dispute under Jokowi's Administration [Respons Indonesia di tengah Dinamika Sengketa Laut China Selatan di bawah Pemerintahan Jokowi. *Jurnal Politica Dinamika Masalah Politik Dalam Negeri Dan Hubungan Internasional*, *12*(1), 85–103. https://doi.org/10.22212/jp.v12i1.2149

Wickramasinghe, I., & Kalutarage, H. (2021). Naive Bayes: applications, variations and vulnerabilities: a review of literature with code snippets for implementation. *Soft Computing*, *25*(3), 2277–2293. https://doi.org/10.1007/s00500-020-05297-6

Yeremia, A. E. (2022). Indonesian diplomats' and foreign policy scholars' perceptions and their implications on Indonesian foreign ministry bureaucratic responses to a rising China. *The Pacific Review*, *35*(3), 529–556. https://doi.org/10.1080/09512748.2020.1851293