Comparative Analysis of Machine Learning Algorithms in Sentiment Analysis of E-Commerce Application User Reviews

Frayudi[1*], Muhammad Halmi Dar[2], Fitri Aini Nasution[3]

[1,2,3] Faculty of Science and Technology, Universitas Labuhanbatu**,** Sumatera Utara Indonesia.
[*]Corresponding Author:
Email: yudifra20@gmail.com

Abstract.

Shopee, as one of the leading e-commerce platforms, receives thousands of user reviews every day. These reviews contain valuable information that can help improve the quality of products and services. However, analyzing reviews manually is a very time-consuming task and is prone to human error. Therefore, the analysis of user review sentiment requires an automated solution. Machine learning algorithms in user review sentiment analysis can help solve this problem more efficiently. Algorithms such as SVM and Naïve Bayes can be used to classify reviews as positive or negative, providing actionable insights for management. This study aims to compare the performance of two machine learning algorithms, namely SVM and Naïve Bayes, in analyzing the sentiment of Shopee app user reviews. We collected and cleaned Shopee app user review data for the model. Next, we divided the data into a training set and a test set. Both algorithms, SVM and Naïve Bayes, were trained using the training set and evaluated using the test set. The results showed that Naïve Bayes had a precision of 81% and SVM had a precision of 80%. In terms of recall, SVM is superior with a value of 80% compared to Naïve Bayes, which has a value of 79%. Both algorithms have the same F1-score, which is 80%. In terms of accuracy, SVM is slightly superior with a value of 80% compared to Naïve Bayes, which has a value of 79%. Based on the evaluation results, SVM shows a slightly better performance than Naive Bayes in the sentiment analysis of Shopee application user reviews. Although Naive Bayes is superior in precision, SVM has higher recall and accuracy, making it the best algorithm overall in this study. E-commerce companies like Shopee can use this study's insights into the performance of two machine learning algorithms in sentiment analysis of user reviews to enhance the quality of their products and services. In addition, this study also provides guidance for researchers and practitioners in choosing the right algorithm for sentiment analysis based on their specific needs. Thus, this study not only helps in understanding the performance of SVM and Naïve Bayes algorithms but also provides a solid foundation for the practical implementation of sentiment analysis in the context of e-commerce.

Keywords: *E-commerce, Machine Learning, Naïve Bayes, Sentiment Analysis, Shopee, SVM.*

1. INTRODUCTION

The increasing development of information and communication technology has influenced the buying and selling transaction process, which was originally conventional but has now changed to modern. E-commerce is one type of modern buying and selling transaction. E-commerce is a way for consumers to conduct online transactions for goods and services [1]. E-commerce promotes products and services to

consumers. E-commerce allows consumers to make purchases of goods anywhere and anytime. Indonesia is an ASEAN country that ranks first in terms of product sales via e-commerce [2]. Indonesians frequently use Shopee, a marketplace-based e-commerce site [3]. Shopee is one of Indonesia's most visited marketplaces, with 71.5 million visits [4]. In fact, according to iPrice, Shopee is the number one largest marketplace based on the origin of international stores, with 138.7 million monthly web visitors [5].

Shopee, the largest marketplace application in Indonesia, provides users with product and service review features. The review feature is very useful for finding out whether user feedback is positive or negative [6]. The existence of user reviews of a product will be a consideration for buyers in deciding to purchase a product. The review feature will also provide input for Shopee, the developer, to enhance the quality of its services. It is not possible to manually identify user reviews in large numbers by reading them one by one [7]. This process will take a very long time and is ineffective. Therefore, we need a method that can identify user reviews more effectively and efficiently. Sentiment analysis has demonstrated its ability to identify Shopee user reviews by thoroughly examining the information, leading to a classification of positive, negative, or neutral reviews [8]. A number of studies state that sentiment analysis is the same term as opinion mining, opinion extraction, sentiment mining, or review mining [9]. Sentiment analysis is a method that can classify text into positive, neutral, or negative opinions automatically [10]. The purpose of sentiment analysis is to map a person's opinion based on a particular topic [11].

Numerous previous researchers have conducted user sentiment analysis of Shopee products and application reviews, applying various classification algorithms. The first study was conducted on Indonesian language reviews of Shopee users who purchased the Xiaomi Redmi Note 9 smartphone product using the Naïve Bayes Classifier algorithm; the resulting accuracy rate was 85%, but there was still a lot of data that was not perfectly classified [12]. Second, we applied the same algorithm to conduct sentiment analysis on Indonesian language reviews of Shopee application users on Playstore, yielding an accuracy rate of 96.667% [13]. Third, this study used the K-Nearest Neighbor (k-NN) algorithm along with Natural Language Processing to look at how Indonesian language users felt about Shopee's instant hijab products. The results were 76.92% accurate [14]. Fourth, we conducted sentiment analysis by applying the maximum entropy classification method to Indonesian language reviews of Shopee products on Google Play, yielding accuracy results of 97.32 [6]. Fifth, this study combines the K-means algorithm and the Naïve Bayes classifier to conduct sentiment analysis on Shopee product user reviews; the accuracy results obtained were 77.12%, although the classification using K-means was considered not optimal [15].

Choosing the best algorithm for conducting sentiment analysis is very important; it will affect the level of accuracy. Support Vector Machine (SVM) is one of the classification algorithms that has a high level of accuracy compared to other algorithms because it is able to define hyperplanes well [16]. While the Naïve Bayes algorithm is a machine learning classification algorithm with probability reasoning that

is no less good than other algorithms. The Naïve Bayes algorithm also has advantages, namely by using not too much training data and having a good level of accuracy. Sentiment analysis using a combination of the SVM and Naïve Bayes Classifier methods using a dataset of 2,378 has produced an accuracy level of 81.61% on SVM and an accuracy of 67.29% on Naïve Bayes [17]. Furthermore, we conducted a sentiment analysis of opinions about the Covid-19 vaccine on social media. Twitter using a dataset taken from Twitter of 1000 records showed that the SVM method was slightly better than the Naïve Bayes method, with an accuracy level of 90.47% for the SVM algorithm, while Naïve Bayes was 88.64% [18]. In other studies related to sentiment analysis, it has also been proven that the SVM algorithm is better than Naïve Bayes [19], [20].

This study aims to compare the performance of the SVM and Naïve Bayes algorithms in analyzing the sentiment of Shopee application user reviews. We conducted the study on English-language customer reviews. We used secondary data from Kaggle as the research material. This study will divide customer opinions into two, namely positive opinions and negative opinions. We measure each algorithm's performance using a performance evaluation matrix that includes metrics like accuracy, precision, recall, and f1-score. The best algorithm is the one that gets the highest evaluation matrix score. We expect this study to offer recommendations for the most effective algorithms for sentiment analysis on this topic. This study is highly significant because it can help improve understanding of how people view services from e-commerce applications.

II. METHODS

This study compares the SVM and Naïve Bayes algorithms in conducting sentiment analysis on English-language customer reviews of Shopee E-Commerce products. The study uses secondary data, namely datasets obtained from Kaggle [21]. In this study, the Python programming language is utilized on Google Collaborate online, from the process of reading the dataset to the visualization of the results. Researchers aim to simplify the algorithm implementation process by combining Google Colab with Python. Given the free nature of this application, installing the Python library is a straightforward process. Furthermore, the Microsoft Excel application facilitates straightforward data visualization. Figure 1 shows the research methodology carried out.

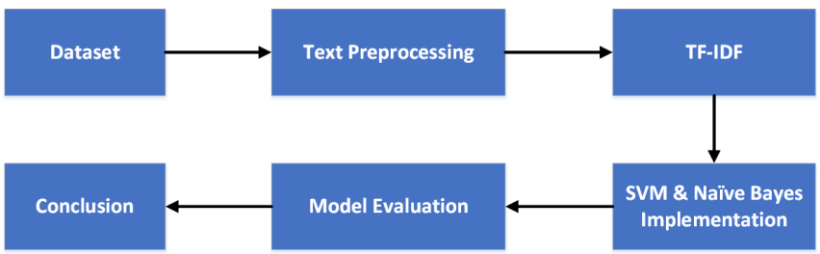

Fig. 1. Research methodology

The number of records in the dataset actually consists of 1,502,575 rows. We used only 10,000 records for this study, each containing three attributes: label, text, and sentiment. This study utilized an unstructured document as its dataset. The first thing to do is to read it. The dataset still requires text processing due to the presence of improperly structured words or sentences. During the text preprocessing phase, we remove any superfluous or non-required data from the text. Some stages for preprocessing text are case folding, tokenization, stopword removal, stemming, and lemmatization. The TF-IDF stage transforms the data into a matrix and applies the Terms Frequency-Inverse Document Frequency (TF-IDF) algorithm to weight each word's value. The results of the TF-IDF value weighting will show the similarities between documents in the dataset. In the next stage, the SVM and Naïve Bayes algorithms are implemented. During this phase, we train the machine to identify preexisting data patterns, and subsequently categorize the TF-IDF weighted text preprocessing data into two groups: positive and negative. The model evaluation stage will assess and compare the performance of the two algorithms using performance matrices such as accuracy, precision, recall, and f1-score. The conclusion stage is to determine which is the best algorithm for performing sentiment analysis.

III. RESULT AND DISCUSSION

The dataset proved to be good after data exploration because it contained no duplicates and no null values. Table 1 displays the results of the exploratory data analysis on Python.

Table 1. Summary of Dataset

Index	Label
count	10000.0
mean	4.7646
std	0.712066
min	1.0
25%,	5.0
50%	5.0
75%	5.0
max	5.0

Table 1 shows that the dataset consists of 10,000 records with an average value of 4.76 and a standard deviation of 0.71. The greater the standard deviation value, the more the sample data deviates from the average. We classify the used dataset as good because it lacks duplicate data and null values.

Table 2. The Results of Text Preprocessing

Text Preprocessing	Results
Full Text	"NOTE: DOESN'T INCLUDE THE BLUE SMALL ONE. The delivery was fast, and the quality is real good. Recommended! "
Case Folding	note: doesn't include the blue small one. the delivery was fast, and the quality is real good. recommended!
Tokenization	note doesnt include the blue small one the delivery was fast and the quality is real good recommended
Stopword Removal	note doesnt include blue small one delivery fast quality real good recommended
Stemming	note doesnt include blue small one delivery fast quality real good recommend
Lemmatization	note doesnt include blue small one delivery fast quality real good recommend

Table 2 shows the translation results from text preprocessing. The results of the case folding process show that there has been a change in the letter form of the dataset, from originally using capital letters to now using lowercase letters. The results of the tokenization process using the remove_punctuation library in Python have removed all punctuation from strings or user reviews in the dataset. The results of the stopwords removal process have reduced words that are considered unimportant in the dataset, such as the words "the," "was," "and," and "is." The results of the stemming process have removed word affixes such as prefixes, infixes, suffixes, and prefixes and confixes in the dataset. The test dataset has eliminated the suffix "ed". The stemming process's results have returned the words in the dataset to their basic form.

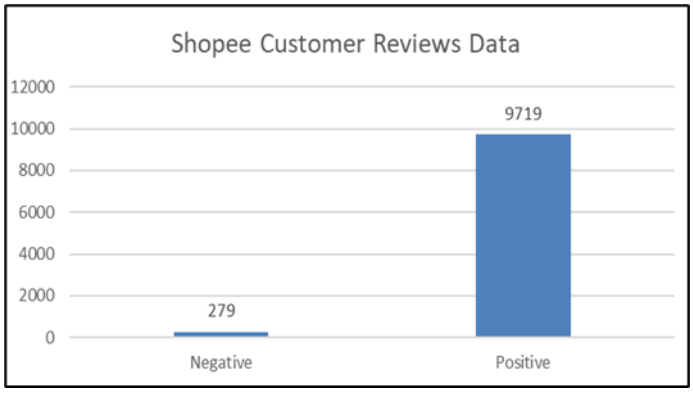

Fig. 2. Customer Opinion Classification

Figure 2 divides Shopee customer reviews into two categories: "positive" and "negative" opinions. The test results reveal that positive opinions from Shopee

customers comprise 9719 data, accounting for 97.19% of the 10,000 user data records. Shopee customers' negative opinions comprise 279 data points, representing a percentage of 27.9%. To see the emergence of words that represent each positive and negative opinion, a data visualization plot is carried out using World Cloud as shown in Figures 3 and 4.

Fig. 3. Positive Sentiment

Figure 3 shows a word cloud visualization representing the most frequently occurring words in Shopee app user reviews with positive sentiment. This wordcloud provides an overview of keywords frequently used by users when giving positive reviews of products or services on Shopee. The word "fast delivery" appears in large size, indicating that many users highlight the speed of delivery as a positive aspect of their shopping experience on Shopee. With the word "good condition," users often praise the condition of the products they receive, indicating that products sold on Shopee generally match the description and are in good condition. The word "good quality" (product quality) is also one of the most praised aspects, indicating user satisfaction with the quality of the goods they buy. The word "received" shows that many positive reviews include timely and expected product receipts. With the word "product," users give positive reviews about the products they buy on Shopee. The word "thank," or thank you from users, shows their appreciation for the service and products they receive. The frequent appearance of the word "order" suggests that users also value the ordering process. People frequently mention the word "size" to refer to the size of the product, most likely implying that it aligns with their expectations and descriptions.

Fig. 4. Negative Sentiment

Figure 4 shows a wordcloud visualization depicting the most frequently occurring words in Shopee app user reviews with negative sentiment. The word "item" appears in large size, indicating that many users mention the "item" or goods they received in negative reviews, most likely due to dissatisfaction with the item. Users often criticize the word "quality", suggesting that the product they received did not meet their expectations. The word "bad" directly indicates dissatisfaction and negative feedback from users about the product or service. Although the word "received" appears in positive reviews, it also frequently appears in negative reviews, possibly in the context that the product received did not match the description or expectations. The words "ordered" and "order," which refer to the ordering process, are also points of complaint, possibly related to errors in ordering or dissatisfaction with the ordering system. The word "thin" may refer to the thinness or quality of the material that does not match user expectations. The word "material," or criticism of the product material, indicates that many users feel that the product material received is inadequate or of low quality. In a refund request, the word "refund" indicates a fairly high level of dissatisfaction, with users feeling the need to return the product and ask for their money back.

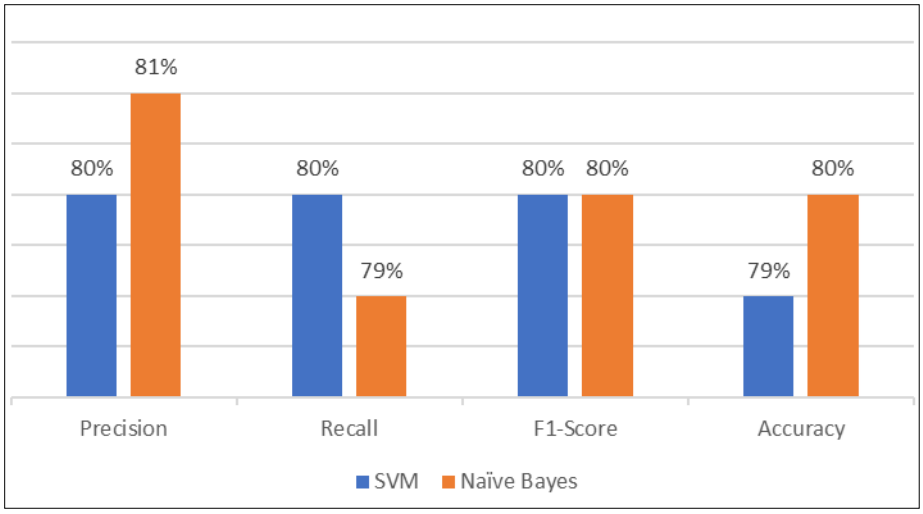

Fig. 5. Evaluation Matrix

Figure 5 shows the results of an evaluation matrix comparing the performance of two machine learning algorithms, namely Support Vector Machine (SVM) and Naïve Bayes, in the sentiment analysis of Shopee application user reviews. This evaluation matrix includes four main metrics: precision, recall, F1-score, and accuracy. Precision measures the model's ability to identify true positives from all the positive predictions generated. The results show that Naïve Bayes has a slightly higher precision compared to SVM, which is 81% vs. 80%. This means that Naïve Bayes is slightly better at generating correct positive predictions than SVM. Recall measures the model's ability to identify true positives from the total number of actual positives. SVM has a slightly higher recall than Naïve Bayes, which is 80% vs. 79%. This shows that SVM is slightly better at capturing all the actual positive data. The F1-Score is the harmonic average of precision and recall. Both algorithms have the same F1-Score, which is 80%. This shows that both algorithms have the same balance between precision and recall in this study. Accuracy is defined as the proportion of correct predictions in the total data. SVM has a slight advantage in accuracy over Naïve Bayes, at 80% vs. 79%. This shows that overall, SVM makes more correct predictions than Naïve Bayes.

According to the research results, both algorithms perform quite similarly in analyzing Shopee app user reviews sentiment. Naïve Bayes is slightly superior in precision, while SVM is superior in recall and accuracy. Both have the same F1-Score, indicating a similar balance between precision and recall. The Naïve Bayes algorithm is simple and fast in training and inference. With slightly higher precision, Naïve Bayes can be a good choice when it is important to minimize false positives. The SVM algorithm is more complex and often takes more time to train, but with higher recall and accuracy, SVM can be a better choice when it is important to minimize false negatives and increase overall correct predictions. SVM may be the preferred choice if the primary objective is to identify as many correct positive reviews as possible. If the main goal is to ensure that positive predictions are correct (for example, for promotions

or product recommendations), then Naïve Bayes may be more suitable. This evaluation matrix shows that both SVM and Naïve Bayes have their respective advantages in analyzing sentiment of Shopee app user reviews. Naïve Bayes is slightly superior in precision, while SVM is superior in recall and accuracy. The algorithm choice depends on the application's specific priorities and context of use. Both algorithms perform quite satisfactorily and yield relatively accurate results in sentiment analysis.

IV. CONCLUSION

This study aims to compare the performance of the Support Vector Machine (SVM) and Naïve Bayes algorithms in analyzing the sentiment of Shopee application user reviews, using evaluation metrics such as accuracy, precision, recall, and F1-score. Naïve Bayes has a slightly higher precision value (81%) compared to SVM (80%). This shows that Naive Bayes is more effective in identifying true positives from all positive predictions. SVM has a slightly higher recall value (80%) compared to Naïve Bayes (79%). This shows that SVM is more effective in capturing all true positive data. Both algorithms have the same F1-score value, which is 80%. This shows that both have a similar balance between precision and recall. SVM has a slightly higher accuracy value (80%) compared to Naïve Bayes (79%). This shows that SVM produces more correct predictions from all data. Based on the evaluation results, SVM has higher recall and accuracy values than Naïve Bayes, while Naïve Bayes has a slightly higher precision value. We can conclude that SVM slightly outperforms Naive Bayes in analyzing the sentiment of Shopee app user reviews, as both algorithms share the same F1-score value, which indicates a balance between precision and recall. As a result, SVM is considered the best algorithm in this study.

We recommend using SVM in the sentiment analysis of Shopee app user reviews due to its higher accuracy and recall values. This will help to capture more positive reviews and increase the overall correct prediction. We recommend performing parameter optimization on both algorithms to further improve the performance. Parameter adjustments can help improve the precision, recall, and accuracy values. In addition to SVM and Naïve Bayes, it is recommended to explore other machine learning algorithms such as Random Forest, Gradient Boosting, or Deep Learning to see if there is any performance improvement in sentiment analysis. Data quality greatly affects the performance of the algorithm. We recommend continuously improving the process of collecting and preprocessing user review data to ensure it is representative and free from noise. Implementing a continuous feedback system can help collect more user review data and iteratively improve the model based on the latest feedback. We anticipate that implementing these suggestions will enhance the Shopee app user review sentiment analysis performance, offering more precise and valuable insights for business decision-making.

REFERENCES

[1] H. Widyastuti and T. A. Prastitya, “Preferensi Konsumen Pengguna E-Commerce yang Memengaruhi Kesadaran akan Perlindungan Konsumen pada Generasi X,” *J. Sist. Inf.*

Bisnis, vol. 10, no. 1, pp. 10–19, 2020, doi: 10.21456/vol10iss1pp10-19.
[2] F. Azzahro, P. W. Handayani, S. S. Murti, and S. B. Yudhoatmojo, "The Effects of Perceived Justice and Emotions on Service Recovery Satisfaction on Indonesian B2B and C2C E-commerce Customers," *J. Sist. Inf.*, vol. 16, no. 1, pp. 38–48, 2020.
[3] Nurdin, M. Hutomi, M. Qamal, and B. Bustami, "Sistem Pengecekan Toko Online Asli atau Dropship pada Shopee Menggunakan Algoritma Breadth First Search," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 4, no. 6, pp. 1117–1123, 2020, doi: 10.29207/resti.v4i6.2514.
[4] J. J. A. Limbong, I. Sembiring, and K. D. Hartomo, "Analisis Klasifikasi Sentimen Ulasan Pada E-Commerce Shopee Berbasis Word Cloud Dengan Metode Naive Bayes Dan K-Nearest Neighbor," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 9, no. 2, pp. 347–356, 2022, doi: 10.25126/jtiik.202294960.
[5] Iprice, "Peta E-Commerce Indonesia," iprice.co.id. Accessed: Feb. 10, 2023. [Online]. Available: https://iprice.co.id/insights/mapofecommerce/
[6] U. Rhohmawati, I. Slamet, and H. Pratiwi, "Sentiment Analysis Using Maximum Entropy on Application Reviews (Study Case: Shopee on Google Play)," *J. Ilm. Tek. Elektro Komput. dan Inform.*, vol. 5, no. 1, pp. 44–49, 2019, doi: 10.26555/jiteki.v5i1.13087.
[7] G. P. Wiratama and A. Rusli, "Sentiment Analysis of Application User Feedback in Bahasa Indonesia Using Multinomial Naive Bayes," in *2019 5th International Conference on New Media Studies (CONMEDIA)*, 2019, pp. 223–227. doi: 10.1109/CONMEDIA46929.2019.8981850.
[8] I. N. Kabiru and P. K. Sari, "Analisa Konten Media Sosial E-Commerce Pada Instagram Menggunakan Metode Sentimen Analysis Dan LDA-Based Topic Modeling (Studi Kasus : Shopee Indonesia)," in *e-Proceeding of Management*, 2019, pp. 12–19.
[9] D. Rolliawati, K. Khalid, and I. S. Rozas, "Teknologi Opinion Mining untuk Mendukung Strategic Planning," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 2, p. 293, 2020, doi: 10.25126/jtiik.2020721685.
[10] F. T. Saputra, Y. Nurhadryani, S. H. Wijaya, and D. Defina, "Analisis Sentimen Bahasa Indonesia pada Twitter Menggunakan Struktur Tree Berbasis Leksikon," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 1, p. 135, 2021, doi: 10.25126/jtiik.0814133.
[11] A. B. P. Negara, H. Muhardi, and I. M. Putri, "Analisis Sentimen Maskapai Penerbangan Menggunakan Metode Naive Bayes dan Seleksi Fitur Information Gain," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 3, p. 599, 2020, doi: 10.25126/jtiik.2020711947.
[12] L. O. Sihombing, H. Hannie, and B. A. Dermawan, "Sentimen Analisis Customer Review Produk Shopee Indonesia Menggunakan Algortima Naïve Bayes Classifier," *Edumatic J. Pendidik. Inform.*, vol. 5, no. 2, pp. 233–242, 2021, doi: 10.29408/edumatic.v5i2.4089.
[13] D. Pratmanto, R. Rousyati, F. F. Wati, A. E. Widodo, S. Suleman, and R. Wijianto, "App Review Sentiment Analysis Shopee Application in Google Play Store Using Naive Bayes Algorithm," *J. Phys. Conf. Ser.*, vol. 1641, no. 1, 2020, doi: 10.1088/1742-6596/1641/1/012043.
[14] E. H. Muktafin, K. Kusrini, and E. T. Luthfi, "Analisis Sentimen pada Ulasan Pembelian Produk di Marketplace Shopee Menggunakan Pendekatan Natural Language Processing," *J. Eksplora Inform.*, vol. 10, no. 1, pp. 32–42, 2020, doi: 10.30864/eksplora.v10i1.390.
[15] T. Hariguna, W. M. Baihaqi, and A. Nurwanti, "Sentiment Analysis of Product Reviews as A Customer Recommendation Using the Naive Bayes Classifier Algorithm," *Int. J. Informatics Inf. Syst.*, vol. 2, no. 2, pp. 48–55, 2019, doi: 10.47738/ijiis.v2i2.13.
[16] N. Raisa, N. Riza, and W. I. Rahayu, "Analisis Sentimen menggunakan SVM dan KNN pada Review Drama Korea di MYDRAMALIST," *JINTEKS (Jurnal Inform. Teknol. dan Sains)*, vol. 5, no. 4, 2023, doi: 10.51401/jinteks.v5i4.3114.
[17] H. Tuhuteru and A. Iriani, "Analisis Sentimen Perusahaan Listrik Negara Cabang Ambon Menggunakan Metode Support Vector Machine dan Naive Bayes Classifier," *J. Inform. J.*

Pengemb. IT, vol. 3, pp. 394–401, Oct. 2018, doi: 10.30591/jpit.v3i3.977.
[18] F. Fitriana, E. Utami, and H. Al Fatta, "Analisis Sentimen Opini Terhadap Vaksin Covid - 19 pada Media Sosial Twitter Menggunakan Support Vector Machine dan Naive Bayes," *J. Komtika (Komputasi dan Inform.*, vol. 5, no. 1, pp. 19–25, 2021, doi: 10.31603/komtika.v5i1.5185.
[19] Siswanto, Y. P. Wibawa, W. Gata, G. Gata, and N. Kusumawardhani, "Classification Analysis of MotoGP Comments on Media Social Twitter Using Algorithm Support Vector Machine and Naive Bayes," in *2018 International Conference on Applied Information Technology and Innovation (ICAITI)*, 2018, pp. 96–101. doi: 10.1109/ICAITI.2018.8686751.
[20] A. M. Rahat, A. Kahir, and A. K. M. Masum, "Comparison of Naive Bayes and SVM Algorithm based on Sentiment Analysis Using Review Dataset," in *2019 8th International Conference System Modeling and Advancement in Research Trends (SMART)*, 2019, pp. 266–270. doi: 10.1109/SMART46866.2019.9117512.
[21] Kaggle, "Shopee Text Reviews," kaggle.com. Accessed: Jul. 10, 2024. [Online]. Available: https://www.kaggle.com/datasets/shymammoth/shopee-reviews
[22] Kaggle, "Tapera Dataset," kaggle.com. Accessed: Jun. 25, 2024. [Online]. Available: https://www.kaggle.com/datasets/unshoytable/twitter-tapera-dateset