Jurnal Informatika: Jurnal pengembangan IT

Vol. 10, No. 1, 2025
ISSN : 2477-5126, DOI:10.30591/jpit.v10i1.8180

Analysis of Electronic Wallet User Sentiment on Twitter (x) Social Media Using the Naïve Bayes Classifier Algorithm

Azhar Basir [1], Trena Yudha Prawira [1] Muhamad Aznar Abdillah[1] Ahmad Faizin[2]
[1]Faculty of Science, Technology, and Health, Informatics Engineering, Universitas Muhammadiyah Brebes, Brebes, Indonesia
[2]Faculty of Science, Technology, and Health, Information Systems, Universitas Muhammadiyah Brebes, Brebes, Indonesia

Info Artikel

Riwayat Artikel:

Received 2024-12-30
Revised 2025-01-01
Accepted 2025-01-01

Abstract *– Electronic wallets are one of the most popular payment methods in Indonesia with the number of users increasing significantly in recent years, including DANA, GoPay, and LinkAja. Along with the increasing number of users, the need to analyze user opinions and comments on social media, especially on Twitter (X) is also increasing. This study uses an experimental method with data collection from Twitter (X) using data crawling techniques. The dataset used is 1,500 data with 1 attribute. This study aims to analyze user sentiment towards electronic wallets using the Naive Bayes Classifier algorithm with the Python programming language. The results of the study showed that DANA had a negative sentiment of 16.6%, a neutral sentiment of 9.0%, and a positive sentiment of 74.4%. Followed by GoPay with a negative sentiment of 9.4%, a neutral sentiment of 11.4%, and a positive sentiment of 79.2%. Meanwhile, LinkAja had the lowest negative sentiment of 8%, with a neutral sentiment of 12.2% and a positive sentiment of 79.8%. The implementation of the Naive Bayes Classifier algorithm achieved an accuracy rate of 72% for DANA, 88% for GoPay, and 88%, for LinkAja.*

Keywords: *Electronic Wallet, Sentiment Analysis, Twitter (X), Naïve Bayes Classifier*

Corresponding Author:

Azhar Basir

Email: azharbs@stmikmpb.ac.id

Abstrak *– Dompet elektronik merupakan salah satu metode pembayaran terpopuler di Indonesia dengan jumlah pengguna yang meningkat secara signifikan dalam beberapa tahun terakhir, termasuk DANA, GoPay, dan LinkAja. Seiring dengan bertambahnya jumlah pengguna, kebutuhan untuk menganalisis opini dan komentar pengguna di media sosial khususnya di Twitter (X) juga semakin meningkat. Penelitian ini menggunakan metode eksperimen dengan pengumpulan data dari Twitter (X) menggunakan teknik data crawling. Dataset yang digunakan sebanyak 1.500 data dengan 1 atribut. Penelitian ini bertujuan untuk menganalisis sentimen pengguna terhadap dompet elektronik menggunakan algoritma Naive Bayes Classifier dengan bahasa pemrograman Python. Hasil penelitian menunjukkan DANA mempunyai sentimen negatif sebesar 16,6%, sentimen netral sebesar 9,0%, dan sentimen positif sebesar 74,4%. Disusul GoPay dengan sentimen negatif sebesar 9,4%, sentimen netral sebesar 11,4%, dan sentimen positif sebesar 79,2%. Sedangkan LinkAja memiliki sentimen negatif paling rendah sebesar 8%, sentimen netral sebesar 12,2%, dan sentimen positif sebesar 79,8%. Penerapan algoritma Naive Bayes Classifier mencapai tingkat akurasi sebesar 72% untuk DANA, 88% untuk GoPay, dan 88% untuk LinkAja.*

Kata Kunci*: Dompet Elektronik, Analisis Sentimen, Twitter (X), Naïve Bayes Classifier*

I. INTRODUCTION

Machine learning has developed quite rapidly lately, one of which is sentiment analysis [1] [2]Sentiment analysis is a method for determining whether a text written by a user is positive, negative, or neutral [3], [4] Positive sentiments communicate a positive view in a situation, negative sentiments express negative ideas in a context, and neutral sentiments state things that are neither good nor bad [4], [5]. Sentiment analysis in text mining is the study of sentiments, emotions, and attitudes in opinion texts [6]. Sentiment analysis has become very useful since the rise of social media and online review websites, and thus the requirement to analyze their sentiment in an effective and efficient way [7]. Recommender systems employ sentiment analysis and classification tasks to examine social media content such as blogs, discussion forums, Facebook postings, tweets, movie reviews, and online product reviews [8]. Sentiment analysis can be used to transform previously unstructured data into more structured data [9], [10]. Therefore, we can think of sentiment analysis as a text classification problem that uses sentiment as a category[7]. One of the objects that can be studied to classify sentiment is an electronic wallet.

Electronic wallets (e-wallets) are one of the simplest and most practical digital payment methods, allowing users to carry out various financial transactions anytime, anywhere, and from any location throughout the world[[11]. The use of e-wallets is increasingly popular in line with the progress of digitalization in the banking and financial services sector, which not only makes it easier to access transactions, but also plays an important role in implementing safety and preventive measures, especially in dealing with the Covid-19 pandemic. During the pandemic, adopting digital payments is key to reducing physical contact and minimizing the spread of the virus, as well as helping to protect public safety[12]. In addition, developments in financial technology including electronic wallets, online banking services and fintech (financial technology) provide many benefits for consumers. By using e-wallet, users no longer need a physical wallet to store money or financial cards. Instead, they can store various financial cards such as debit cards, credit cards, prepaid cards, and gift cards in virtual form, which is not only safer but also more practical to use in daily transactions [12]. The existence of this electronic wallet changes the way we interact with money, introducing convenience, efficiency and comfort in an increasingly digital financial life. The concept of an electronic wallet may be viewed as a model for performing ordinary financial transactions. Although there are many types of electronic payments, this study will only focus on three electronic wallets, namely DANA, GoPay, and LinkAja. Sentiment analysis can be used on many platforms, one of which is tweets from Twitter (X). Twitter (X) has become one of the most popular platforms for individuals to openly express their ideas on any topic from anywhere in the globe [13]. These tweets may contain text or even videos that are attached [1]. However, Twitter (X) has not been able to categorize tweets into positive, neutral, and negative sentiments, so the tweets must be processed first so that they can be used as a reference for conducting sentiment analysis, especially in text classification.

Naive Bayes classifier is a supervised machine learning method and is often used in text classification tasks [14]. This algorithm is very effective in solving classification problems, especially when dealing with large feature spaces, making it a good choice for sentiment analysis [15]. Naive Bayes functions as a classification model based on Bayes' Theorem, with the assumption that all predictor variables or features are mutually independent and have equal weight in the prediction process. [16], [17]. The main advantages of the Naive Bayes approach are its simplicity which makes it easy to implement, speed in processing large amounts of data, and its ability to produce good results, even when compared with more complex classification algorithms. That is why Naive Bayes is very popular in a variety of applications, including text analysis and natural language processing, where its ability to efficiently handle large volumes of data is highly appreciated[13]. Some machine learning approaches that have been used in sentiment analysis include Naive Bayes[1], [13], [14], [15], [18], [19], Support Vector Machines [1], [8], [13] Random Forest [5], [20], [21], Logistic Regression [1], [13], Decision Tree [18], [19] and K-Nearest Neighbor (KNN) [18], [21] [20]. Previous studies used text mining for sentiment analysis on Twitter. One of them is research from [21]proposes a comparative review analysis model visualized in the tourism domain using the Naïve Bayes Classifier algorithm with a case study of two popular tourist attractions in Bali, namely Kuta Beach and Jimbaran Beach. The stages involved in conducting sentiment analysis include data collection, initial data processing, Naïve Bayes Classifier classification, model evaluation, comparison of results, and visualization. The comparison of the two tourist locations produces an accuracy value of 80% for Kuta Beach and 64% for Jimbaran Beach.

As for research from [1] focused on analyzing the level of customer satisfaction with Traveloka services, with the aim of understanding how many people feel satisfied or dissatisfied with the various services offered by the company. In this research, the data used was obtained via the Twitter API, which includes 1200 tweets related to Traveloka. The classification method applied to analyze this data involves three main algorithms, namely Support Vector Machine (SVM), Logistic Regression, and Naïve Bayes. The research process includes several important stages, starting from data collection, data transformation, to sentiment analysis to determine how the majority of Twitter users feel about the performance of the Traveloka mobile application. Based on the analysis results, this research shows that the SVM method provides better classification performance in predicting test results, compared to the other two methods.

In addition, another study conducted by [22] focused on sentiment analysis of tweets related to the COVID-19 pandemic. The data used in this research amounted to 179,107 tweets, which were analyzed using a data pre-processing approach and the Long Short-Term Memory (LSTM) model. LSTM was chosen for its ability to handle sequential data and understand context in longer texts. The results of this research are very impressive, with an accuracy rate of 96%, which shows the model's ability to accurately identify the sentiment contained in tweets related to COVID-19. This success shows the great potential of applying sentiment analysis techniques using LSTM in the context of social media to understand public opinion regarding global issues such as the pandemic. Additionally, a pre-processing procedure is introduced to improve the accuracy of sentiment analysis. The Long Short-Term Memory (LSTM) algorithm is used in this study in order to analyze sentiments related to the COVID-19 pandemic using Twitter data.

Meanwhile, Research conducted by [23] highlights the importance of online reviews which have now become an integral part of the purchasing decision making process for many consumers. As e-commerce continues to grow, every online retailer needs to consider effective strategies to optimize the use of reviews as a marketing and feedback tool. As the number of reviews continues to increase, methods or techniques are needed that can automatically classify these reviews, whether they have positive or negative sentiment. For this reason, this research tries to apply the Naïve Bayes Classifier (NBC) method to classify e-commerce reviews. The results obtained from this research show that the Naïve Bayes Classifier can provide an accuracy of 72%, with a Recall value also reaching 72% and a Precision of 78%. However, this study will explore the use of Naive Bayes for sentiment classification in Indonesian for users of electronic wallet applications on social media Twitter (X) so that application developers can take the necessary steps to improve electronic wallet application services.

II. RESEARCH METHOD

Naïve Bayes is chosen because it is easy to implement, efficient in processing large-scale text data, and capable of delivering competitive performance even when compared to more complex algorithms. As a probability-based algorithm [24] [25]Naïve Bayes is well-suited for text classification tasks. Its speed and efficiency make it particularly effective in handling large datasets, such as reviews from Twitter[26].

As depicted in Figure 1, the stages of sentiment analysis in this study consist of six main stages, namely data collection, preprocessing, labeling using sentiment dictionary, TF-IDF weighting, classification using Naïve Bayes, and model evaluation. The output of the classification results is in three categories, namely positive, neutral, and negative.

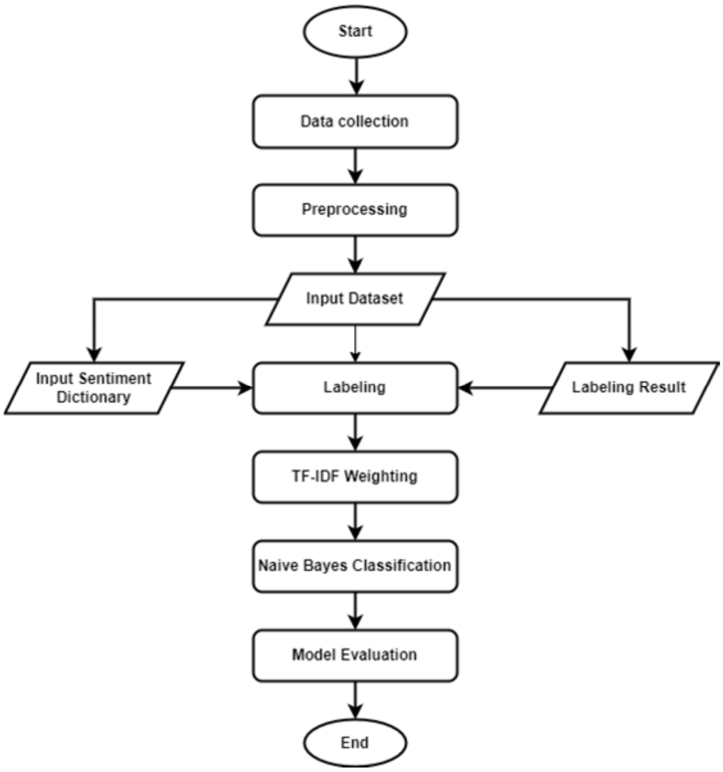

Figure **1.** Flowchart of Algorithm Experiment Stages

A. Data Collection

The dataset on the use of electronic wallets DANA, GoPay, and LinkAja is a private dataset obtained from the social media Twitter (X) by crawling data using the Python programming language. The dataset from each electronic wallet resulting from the crawled data was then downloaded and saved in .csv format consisting of 500 DANA datasets, 500 GoPay datasets, and 500 LinkAja datasets. The data containing user reviews of the electronic wallet applications DANA, GoPay, and LinkAja were taken from 2021 to 2024 with the search keywords "aplikasi dana min_faves:2", "aplikasi gopay min_faves:2", and "aplikasi linkaja min_faves:2" on the Twitter (X) platform.

B. Preprocessing

The preprocessing method involved deleting duplicate data totaling 1,500 data points and removing unnecessary features, specifically hashtags. According to [27] Sentiment analysis preprocessing requires NLTK. NLTK (Natural Language Toolkit) is one of the main libraries in Python designed specifically for natural language processing, which allows developers to build applications that can interact with human language-based data. The initial process in analyzing language data begins with data collection, which is then processed using various techniques. These steps include removing irrelevant elements such as retweets, URLs, and punctuation that could interfere with the analysis. In addition, the emojis in the text are converted into upper-lower case, then tokenization is carried out to break the text into smaller parts such as words or phrases. The next process includes removing stop words, namely words that do not make a significant contribution to the meaning of the analysis, as well as normalization which aims to equalize word forms to make them more consistent. Finally, a stemming process is carried out which changes the words into their basic form, making it easier to further analyze the text. All of these steps aim to simplify and improve the quality of the data before it is used in further analysis processes [5], [17], [28]. The preprocessing results are shown in Table 1.

TABLE 1. EXAMPLE OF TEXT PREPROCESSING RESULTS

Before Preprocessing	After Preprocessing
HELP RT !! Dear @KreditPintar @HaloBCA sodara saya ini ga pernah sekalipun ngedownload aplikasi pinjol Kredit Pintar samsek tbtb ada Dana masuk ke BCAnya dan langsung pindah ke rek lain juga. Ini maksudnya apa dan gimana ya? Mana jatuh temponya bulan depan. Tolong dibantu dong. https://t.co/MeYdrtc4HU	bantu sayang saudara tidak unduh aplikasi pinjam digital kredit pintar sama sekali tbtb dana masuk bcanya langsung pindah rek maksud bagaimana iya jatuh tempo tolong bantu meydrtchu

C. Labeling

Data classification is labeled using a sentiment dictionary divided into three classes: negative, neutral, and positive. These words are saved in a.txt file, not in a database. For example, the term "very happy" is more positive than the word "happy", however the word "less happy" is not more positive than the word "happy", implying that the word can increase or decrease the intensity of the mood of the surrounding words. The word "evil" has a negative connotation, but if the word "evil" is preceded by the negation word "not" then the word arrangement becomes "not evil" which has a positive meaning, where the word is a negation word or words in a phrase have the ability to change someone's views or opinions. The labeling results are shown in Table 2.

TABLE 2. EXAMPLE OF LABELING TEXT

aplikasi mypertamina bikin hari mudah loh iya sob yuk tulis menang total saldo linkaja rupiah sobat pertamina untung xowqjnmfg	Posistive
informasi transaksi myim bima dapat cashback aplikasi gopay dknhkiq	Neutral
artikel manusia pakai dana rupiah miliar aplikasi jelek bptsgcl	Negative

The results of text labeling on the DANA application include 372 texts containing positive sentiment, 83 texts containing negative sentiment, and 45 texts containing neutral sentiment. The results of labeling on the GoPay application include 396 texts containing positive sentiment, 47 texts containing negative sentiment, and 57 texts containing neutral sentiment. The results of text labeling on LinkAja include 399 texts containing positive sentiment, 40 texts containing negative sentiment, and 61 texts containing neutral sentiment.

D. TF-IDF Weighting

In this research, the transformation approach applied is the TF-IDF (Term Frequency-Inverse Document Frequency) method which combines two main components, namely Term Frequency (TF) and Inverse Document Frequency (IDF). This technique is used to evaluate the importance of a word in a document to the collection of documents as a whole Once the data is converted into numeric form, the first stage in constructing a machine learning-based model is to separate the datasets into training and testing sets[14] . Twitter data is divided into training and testing data with a ratio of 90%:10% training and testing data [29]. Table 3 shows the TF-IDF weighting findings for the DANA, GoPay, and LinkAja e-wallets, as well as the top 5 features (terms) and rankings (ranks) for each dataset.

TABLE 3. TF-IDF WEIGHTING RESULTS

1.	**TF-IDF Terms and Ranks for DANA**	**Term**	**Rank**
		dana	0,052372
		aplikasi	0,046557
		yang	0,027326
		iya	0,021876
		pakai	0.021869
2.	**TF-IDF Terms and Ranks for GoPay**	**Term**	**Rank**
		gopay	0,064101
		aplikasi	0,059774
		pakai	0,043840
		bayar	0,040722
		cashback	0,034653
3.	**TF-IDF Terms and Ranks for LinkAja**	**Term**	**Rank**
		linkaja	0,067321
		aplikasi	0,053159

1.	**TF-IDF Terms and Ranks for DANA**	**Term**	**Rank**
		dana	0,052372
		aplikasi	0,046557
		yang	0,027326
		iya	0,021876
		pakai	0.021869
		bayar	0,044489
		mypertamina	0,044127
		iya	0,039807

E. Naive Bayes Classification

Experiments were conducted utilizing a dataset that was divided into two parts: data training and data testing. A dataset that was split into training and test portions was used for the experiment. The goal of the categorization procedure is to use the data in the chosen category to create a model [30]. The algorithm learns from training data, which is subsequently tested on testing data. Based on the testing results, we created a confusion matrix that would be used to evaluate the algorithm[19]. Types of models in the Naive Bayes algorithm include Gaussian, Multinomial, and Bernoulli. The suggested approach employs the multinomial NB model, which predicts the badge of a text, such as an email or newspaper article. The likelihood of each badge is determined for a given sample, and the badge with the highest probability is returned as output. This technique, commonly used for natural language processing and text data analysis, proved ideal for analyzing sentiment in movie reviews [27].

F. Model Evaluation

The confusion matrix evaluates machine learning algorithms based on classification and prediction data [6]. Evaluation was performed using a 3x3 multiclass confusion matrix. The 3x3 multiclass confusion matrix includes:

TNeg : Number of negative predictions from negative actual data
FNegNet : Number of neutral predictions from negative actual data
FNegPos : Number of positive predictions from negative actual data
FNetNeg : Number of negative predictions from neutral actual data
TNet : Number of neutral predictions from neutral actual data
FNetPos : Number of positive predictions from neutral actual data
FPosNeg : Number of negative predictions from positive actual data
FPosNet : Number of neutral predictions from positive actual data
TPos : Number of positive predictions from positive actual data

The 3x3 multiclass confusion matrix of each algorithm is shown in Table 4.

TABLE 4. . 3X3 MULTICLASS CONFUSION MATRIX

		Prediction		
		Negative	Neutral	Positive
Actual	Negative	TNeg	FNegNet	FNegPos
	Neutral	FNetNeg	TNet	FNetPos
	Positive	FPosNeg	FPosNet	TPos

III. EXPERIMENTAL RESULT

Naïve Bayes is implemented using scikit-learn, a Python tool for classification. The average accuracy results of each e-wallets are shown in Table 5.

TABLE 5. AVERAGE RESULT OF NAÏVE BAYES ACCURACY ON E-WALLETS

No.	Types of Electronic Wallets	Naïve Bayes Accuracy Results
1.	DANA	72%
2.	GoPay	88%
3.	LinkAja	88%

According to Table 5, the test data prediction results for each e-wallet reveal that the model achieves an accuracy of 72% on the DANA application and 88% on GoPay and LinkAja. The Naïve Bayes classification model is evaluated using a 3x3 multiclass confusion matrix, displaying its performance on test data with known values. Table 6 shows the 3x3 multiclass confusion matrix sentiment analysis of the electronic wallets DANA, GoPay and LinkAja.

TABLE 6. MODEL EVALUATION USING A 3X3 MULTICLASS CONFUSION MATRIX

DANA		Prediction		
		Negative	Neutral	Positive
Actual	Negative	0	0	9
	Neutral	0	0	5
	Positive	0	0	36
GoPay		Prediction		
		Negative	Neutral	Positive
Actual	Negative	0	0	3
	Neutral	0	0	3
	Positive	0	0	44
LinkAja		Prediction		
		Negative	Neutral	Positive
Actual	Negative	0	0	3
	Neutral	0	0	3
	Positive	0	0	44

Table 6 shows the results of the 3x3 multiclass confusion matrix study of the DANA e-wallet, which revealed that 9 positive samples were incorrectly predicted as negative, 5 positive samples were incorrectly predicted as neutral, and 36 positive samples were correctly predicted. Meanwhile, the results of the 3x3 multiclass confusion matrix of the GoPay and LinkAja wallets both revealed that 3 positive samples were incorrectly predicted as negative, 3 positive samples were incorrectly predicted as neutral, and 44 positive samples were correctly predicted.

Table 7 shows the results of the distribution of sentiment analysis of DANA, GoPay, and LinkAja e-wallets.

TABLE 7. RESULTS OF THE DISTRIBUTION OF SENTIMENT ANALYSIS OF E-WALLETS

No.	Types of Electronic Wallets	Sentiment (%)		
		Negative	Neutral	Positive
1.	DANA	16,6%	9,0%	74,4%
2.	GoPay	9,4%	11,4%	79,2%
3.	LinkAja	8,0%	12,2%	79,8%

Table 7 shows that sentiment analysis on DANA e-wallet produces negative sentiment of 16.6%, followed by neutral sentiment of 9.0%, and positive sentiment of 74.4%, while on GoPay e-wallet it produces negative sentiment of 9.4%, neutral sentiment of 11.4%, and positive sentiment of 79.2%, then LinkAja e-wallet produces negative sentiment of 8.0%, neutral sentiment of 12.2%, and positive sentiment of 79.8%. So, it can be concluded that LinkAja e-wallet produces the highest percentage of positive sentiment, which is 79.8%, only slightly different from GoPay which has a positive sentiment of 79.2%, while DANA has the lowest positive sentiment, which is 74.4%.

IV. CONCLUSION

User perceptions of electronic wallet applications are influenced by factors such as features, promotions, and service integration. The research findings show that the LinkAja e-wallet app ranked first due to a higher percentage of positive sentiment compared to the others, followed closely by GoPay, which only slightly differed in positive sentiment from LinkAja, and lastly, DANA. The Naïve Bayes Classifier algorithm achieved an accuracy of 72% in sentiment analysis for the DANA e-wallet app, while GoPay and LinkAja achieved accuracies of 88%. This contributes to addressing the research problem and objectives. It is hoped that future studies will expand the dataset from Twitter (X) and the Google Play Store review section to enhance the results, as well as use other classification methods to compare the model testing results to identify the best data classification method.

REFERENCES

[1] Z. A. Diekson, M. R. B. Prakoso, M. S. Q. Putra, M. S. A. F. Syaputra, S. Achmad, and R. Sutoyo, "Sentiment analysis for customer review: Case study of Traveloka," in *Procedia Computer Science*, Elsevier B.V., 2022, pp. 682–690. doi: 10.1016/j.procs.2022.12.184.

[2] A. Basir, koidah koidah, and A. Jamil, "Exploration of the Dempster-Shafer Theory for Diagnosing Hardware Damage: Comparative Study with the Certainty Factor Method," *Journal of Informatics Information System Software Engineering and Applications (INISTA)*, vol. 6, no. 2, pp. 98–107, Mar. 2024, doi: 10.20895/inista.v6i2.1401.

[3] Hertina *et al.*, "Data mining applied about polygamy using sentiment analysis on twitters in indonesian perception," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 4, pp. 2231–2236, Aug. 2021, doi: 10.11591/EEI.V10I4.2325.

[4] F. A. 2 Pablo A Henríquez 1, "Analyzing digital societal interactions and sentiment classification in Twitter (X) during critical events in Chile," *national library of medicine*, 2024.

[5] V. A. Fitri, R. Andreswari, and M. A. Hasibuan, "Sentiment analysis of social media Twitter with case of Anti-LGBT campaign in Indonesia using Naïve Bayes, decision tree, and random forest algorithm," in *Procedia Computer Science*, Elsevier B.V., 2019, pp. 765–772. doi: 10.1016/j.procs.2019.11.181.

[6] P. H. Prastyo, A. S. Sumi, A. W. Dian, and A. E. Permanasari, "Tweets Responding to the Indonesian Government's Handling of COVID-19: Sentiment Analysis Using SVM with Normalized Poly Kernel," *Journal of Information Systems Engineering and Business Intelligence*, vol. 6, no. 2, p. 112, Oct. 2020, doi: 10.20473/jisebi.6.2.112-122.

[7] M. A. Fauzi, "Random forest approach fo sentiment analysis in Indonesian language," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 12, no. 1, pp. 46–50, Oct. 2018, doi: 10.11591/ijeecs.v12.i1.pp46-50.

[8] L. K. Ramasamy, S. Kadry, and S. Lim, "Selection of optimal hyper-parameter values of support vector machine for sentiment analysis tasks using nature-inspired optimization methods," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 1, pp. 290–298, Feb. 2021, doi: 10.11591/eei.v10i1.2098.

[9] M. Isnan, G. N. Elwirehardja, and B. Pardamean, "Sentiment Analysis for TikTok Review Using VADER Sentiment and SVM Model," in *Procedia Computer Science*, Elsevier B.V., 2023, pp. 168–175. doi: 10.1016/j.procs.2023.10.514.

[10] Y. Mao, Q. Liu, and Y. Zhang, "Sentiment analysis methods, applications, and challenges: A systematic literature review," Apr. 01, 2024, *King Saud bin Abdulaziz University*. doi: 10.1016/j.jksuci.2024.102048.

[11] N. I. Abas and D. Puspawati, "E-Wallet Adoption in Continuance Intention As A e-Payment System for Live Streaming Shopping," in *Procedia Computer Science*, Elsevier B.V., 2024, pp. 1137–1144. doi: 10.1016/j.procs.2024.03.109.

[12] A. Daragmeh, J. Sági, and Z. Zéman, "Continuous intention to use e-wallet in the context of the covid-19 pandemic: Integrating the health belief model (hbm) and technology continuous theory (tct)," *Journal of Open Innovation: Technology, Market, and Complexity*, vol. 7, no. 2, Jun. 2021, doi: 10.3390/joitmc7020132.

[13] D. Locatelli, G. Damo, and D. Nozza, "A Cross-Lingual Study of Homotransphobia on Twitter," 2023. [Online]. Available: https://www.ustranssurvey.org/reports

[14] Y. Shang, "Prevention and detection of DDOS attack in virtual cloud computing environment using Naive Bayes algorithm of machine learning," *Measurement: Sensors*, vol. 31, Feb. 2024, doi: 10.1016/j.measen.2023.100991.

[15] R. A. Sagum, P. A. C. Clacio, R. E. R. Cayetano, and A. D. F. Lobrio, "Philippine Court Case Summarizer using Latent Semantic Analysis," in *Procedia Computer Science*, Elsevier B.V., 2023, pp. 474–481. doi: 10.1016/j.procs.2023.10.548.

[16] H. Cam, A. V. Cam, U. Demirel, and S. Ahmed, "Sentiment analysis of financial Twitter posts on Twitter with the machine learning classifiers," *Heliyon*, vol. 10, no. 1, Jan. 2024, doi: 10.1016/j.heliyon.2023.e23784.

[17] M. Qorib, T. Oladunni, M. Denis, E. Ososanya, and P. Cotae, "Covid-19 vaccine hesitancy: Text mining, sentiment analysis and machine learning on COVID-19 vaccination Twitter dataset," *Expert Syst Appl*, vol. 212, Feb. 2023, doi: 10.1016/j.eswa.2022.118715.

[18] M. Bilal, H. Israr, M. Shahid, and A. Khan, "Sentiment classification of Roman-Urdu opinions using Naïve Bayesian, Decision Tree and KNN classification techniques," *Journal of King Saud University - Computer and Information Sciences*, vol. 28, no. 3, pp. 330–344, Jul. 2016, doi: 10.1016/j.jksuci.2015.11.003.

[19] B. M. Susanto, "Naïve Bayes Decision Tree Hybrid Approach for Intrusion Detection System," *Buletin Teknik Elektro dan Informatika (Bulletin of Electrical Engineering and Informatics*, vol. 2, no. 3, pp. 225–232, 2013.

[20] A. A. S. E. D. M. H. N. A. P. and C. P. M. S. A. V. D. Sano, "Proposing a Visualized Comparative Review Analysis Model on Tourism Domain Using Naïve Bayes Classifier," pp. 482–489, 2023.

[21] N. M. N. Mathivanan, N. A. Nor, and R. M. Janor, "Improving classification accuracy using clustering technique," *Bulletin of Electrical Engineering and Informatics*, vol. 7, no. 3, pp. 465–470, Sep. 2018, doi: 10.11591/eei.v7i3.1272.

[22] B. Valarmathi, N. S. Gupta, V. Karthick, T. Chellatamilan, K. Santhi, and D. Chalicheemala, "Sentiment Analysis of Covid-19 Twitter Data using Deep Learning Algorithm," in *Procedia Computer Science*, Elsevier B.V., 2024, pp. 3397–3407. doi: 10.1016/j.procs.2024.04.320.

[23] H. Huang, A. A. Zavareh, and M. B. Mustafa, "Sentiment Analysis in E-Commerce Platforms: A Review of Current Techniques and Future Directions," 2023, *Institute of Electrical and Electronics Engineers Inc.* doi: 10.1109/ACCESS.2023.3307308.

[24] C. Naulak, "A comparative study of Naive Bayes Classifiers with improved technique on Text Classification," May 31, 2022. doi: 10.36227/techrxiv.19918360.v1.

[25] A. Basir *et al.*, "Sistemasi: Jurnal Sistem Informasi Penerapan Rule Base pada Sistem Pakar Identifikasi Jenis Kulit Wajah menggunakan Forward Chaining Application Rule Base on Facial Skin Type Identification Expert System using Forward Chaining." [Online]. Available: http://sistemasi.ftik.unisi.ac.id

[26] Q. Zeng *et al.*, "Improved Naive Bayes with Mislabeled Data," Apr. 2023, [Online]. Available: http://arxiv.org/abs/2304.06292

[27] N. Pavitha *et al.*, "Movie recommendation and sentiment analysis using machine learning," *Global Transitions Proceedings*, vol. 3, no. 1, pp. 279–284, Jun. 2022, doi: 10.1016/j.gltp.2022.03.012.

[28] J. Junifer Pangaribuan and O. Putra Barus, "IMPLEMENTASI ALGORITMA TF-IDF DAN SUPPORT VECTOR MACHINE TERHADAP ANALISIS PENDETEKSI KOMENTAR CYBERBULLYING DI MEDIA SOSIAL TIKTOK," *JURNAL DEVICE*, vol. 13, no. 1, pp. 124–134.

[29] S. R. Anggraeni, N. A. Ranggianto, I. Ghozali, C. Fatichah, and D. Purwitasari, "Deep Learning Approaches for Multi-Label Incidents Classification from Twitter Textual Information," *Journal of Information Systems Engineering and Business Intelligence*, vol. 8, no. 1, pp. 31–41, Apr. 2022, doi: 10.20473/jisebi.8.1.31-41.

[30] Z. Parveen and M. P. Quadri, "Classification and Prediction of Student Academic Performance using Machine Learning A Review," *International Journal of Computer Sciences and Engineering*, vol. 7, no. 3, pp. 607–614, Mar. 2019, doi: 10.26438/ijcse/v7i3.607614.