SENTIMENT ANALYSIS OF PUBLIC OPINION ON TRANSPORTATION SERVICES IN INDONESIA USING MACHINE LEARNING

Fina Sifaul Nufus[1*]; Windu Gata[2]

Computer Science[1,2]
Universitas Nusa Mandiri, Jakarta, Indonesia[1,2]
https://www.nusamandiri.ac.id/[1,2]
14230025@nusamandiri.ac.id[1], windu@nusamandiri.ac.id[2]
(*) Corresponding Author

The creation is distributed under the Creative Commons Attribution-NonCommercial 4.0 International License.

Abstract— *This study analyzes public sentiment towards transportation services in Indonesia through social media using Naïve Bayes and Support Vector Machine (SVM) algorithms. Data was collected from Twitter using an API with transportation-related keywords over a three-month period. The analysis results indicate that 93.5% of the opinions are neutral, 3.5% are positive, and 3% are negative. The dominance of neutral sentiment suggests potential dataset imbalance or user hesitation in expressing strong opinions. SVM achieved a higher accuracy (100%) compared to Naïve Bayes (92%), which may be influenced by dataset limitations or the model's validation method. Data preprocessing involved several steps, including tokenization, stopword removal, stemming, lemmatization, and handling of missing data to ensure cleaner and more structured text input. These findings highlight the potential of sentiment analysis for transportation policy improvements, providing insights for policymakers and transport service providers. Future research should address data balancing and broader dataset usage to enhance the robustness of findings and support better decision-making in the transportation sector.*

Keywords: *Naïve Bayes, Sentiment Analysis, Support Vector Machine, Transportation, Twitter.*

Intisari— Transportasi memiliki peran penting dalam mobilitas masyarakat Indonesia, terutama di perkotaan yang menghadapi tantangan seperti kemacetan, polusi, dan keterbatasan infrastruktur. Media sosial menjadi platform utama bagi masyarakat untuk menyampaikan opini terkait layanan transportasi. Penelitian ini bertujuan untuk menganalisis sentimen publik terhadap transportasi di Indonesia menggunakan algoritma Naïve Bayes dan Support Vector Machine (SVM). Data dikumpulkan dari Twitter menggunakan API dengan kata kunci terkait transportasi. Hasil analisis menunjukkan bahwa mayoritas opini bersifat netral (93,5%), dengan sentimen positif sebesar 3,5% dan negatif 3%. SVM menunjukkan akurasi lebih tinggi (100%) dibandingkan dengan Naïve Bayes (92%). Temuan ini menunjukkan bahwa analisis sentimen berbasis media sosial dapat menjadi alat yang efektif bagi pembuat kebijakan dan penyedia layanan dalam meningkatkan kualitas transportasi di Indonesia.

Kata Kunci: *Naïve Bayes, Analisis Sentimen, Support Vector Machine, Transportasi, Twitter.*

INTRODUCTION

Transportation plays a crucial role in the mobility of Indonesian society, especially in major cities facing challenges such as congestion, pollution, and limited infrastructure**.** The low quality of public transportation drives an increase in private vehicle usage, highlighting the need for improvements in transportation management, accessibility, and service convenience (Said & Syafey, 2022) (Nalle et al., 2023). In the digital era, social media platforms like Twitter have become key channels for the public to voice their opinions on transportation**.** Additionally, public participation in transportation decision-making plays a crucial role in creating a more inclusive and sustainable system. Research indicates that integrating community participation at every stage of transportation planning can enhance policy effectiveness and increase user satisfaction (Hidayati, 2023) (Pratama et al., 2023).

(Hidayati, 2023). Active public engagement is expected to result in a more efficient, comfortable, and user-oriented transportation system (Nalle et al., 2023) (Pratama et al., 2023). Therefore, efforts to enhance public comfort in transportation and reduce congestion are highly necessary. Twitter has become one of the primary platforms for sentiment analysis research. This platform allows users to share opinions and information in short text formats, making it an effective data source for analyzing public perceptions. For instance, using Twitter to analyze user responses to Gojek services revealed that the Support Vector Machine (SVM) algorithm achieved a high accuracy rate of up to 99% in sentiment classification (Khoiruddin et al., 2023). Similarly, Google Machine Learning has been utilized to analyze public sentiment toward online transportation services across Twitter, YouTube, and Google Search. The analysis results showed an accuracy of 82.6%, with the majority of user opinions categorized as negative sentiment for Gojek (Savita et al., 2021).

Furthermore, researchers developed a sentiment analysis system based on Ensemble Stacking using Twitter data from Gojek and Grab, achieving the highest accuracy of 88% with an F1-score of 87% (Setiawan et al., 2022). As public reliance on transportation continues to grow, it is essential to understand the factors that influence user satisfaction and perception. One increasingly popular approach is social media-based sentiment analysis, which allows for broad and real-time exploration of public opinion (R. Sari & B. Nugroho, 2021).

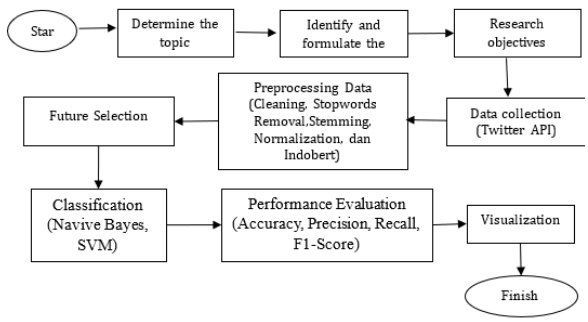
Social media platforms like Twitter not only serve as a space for people to express their experiences but also provide valuable data that can be used to identify trends and patterns in public perception of transportation services (A. Pratama & T. W. Hadi, 2022). Overall, the challenges faced by Indonesia's transportation system require serious attention from all stakeholders. By leveraging sentiment analysis from social media and enhancing public participation in decision-making, it is expected that the quality of transportation services can be improved, ultimately supporting public mobility and sustainable economic growth.

This study aims to analyze public sentiment toward transportation services in Indonesia using Twitter data. By applying sentiment classification techniques, this research seeks to identify dominant sentiment trends and explore their implications for policy and service development. The findings are expected to provide a deeper understanding of public perception and evidence-based recommendations to improve transportation services in Indonesia.

MATERIALS AND METHODS

The utilization of artificial intelligence (AI) and machine learning in sentiment analysis has proven effective in processing large volumes of data and generating insights that can aid strategic decision-making (D. Kusuma et al, 2021). Various studies have shown that methods such as Support Vector Machine (SVM), Naïve Bayes, and deep learning-based approaches can achieve high accuracy in classifying user sentiment toward online transportation services (M. Yusuf & L. Kurniawan, 2023). By leveraging these data analysis techniques, policymakers and service providers can gain a deeper understanding of the aspects that need improvement in Indonesia's transportation system (N. Dewi & P. Santoso, 2022).

Similar research has been conducted by Enos Dwiyanto (Dwianto & Sadikin, 2021). In his study, Enos Dwiyanto utilized the Naïve Bayes and Support Vector Machine algorithms. During the research phase, the collected data was categorized into two classes: positive and negative. The results of this sentiment analysis are expected to provide valuable insights for stakeholders, including users, the government, and transportation officials, to support more informed decision-making regarding an integrated public transportation system. Thus, sentiment analysis from social media can contribute to improving transportation services, enhancing user satisfaction, and supporting the development of a more sustainable and inclusive transportation system (T. Ramadhani & R. Widodo, 2021). This research is a sentiment analysis study using data from Twitter to understand public perception of transportation in Indonesia. The data was collected using a Python script that leveraged the Twitter API with transportation-related keywords, utilizing 200 user reviews.



Source : (Research Result, 2025)
Figure 1. Research Method

Data Collection

Data were collected from Twitter using the Twitter API over a three-month period (January - March 2024). Tweets containing transportation-related keywords such as "transportasi umum," "kemacetan," and "KRL" were retrieved. To maintain relevancy, tweets were filtered by Indonesian language and geographic location within Indonesia. The collected dataset consisted of 200 tweets, which were used for training and evaluation.

Preprocessing

Data preprocessing is the initial stage in data analysis aimed at cleaning and preparing data before further analysis. This preprocessing step is crucial in text analysis, including sentiment analysis on social media, as it helps improve the accuracy of analytical models by eliminating noise and structuring data in a more optimal format (Khoiruddin et al., 2023). Before analysis, the raw text data underwent preprocessing to improve accuracy and reduce noise. The steps included:

1. Tokenization: Splitting text into individual words or tokens.
2. Stopword Removal: Eliminating common words (e.g., "dan," "atau," "yang") that do not contribute to sentiment analysis.
3. Stemming & Lemmatization: Converting words to their root form (e.g., "berjalan" to "jalan").
4. Handling Missing Data: Removing incomplete or irrelevant tweets to maintain dataset quality.
5. TF-IDF Transformation: Converting text into numerical representations by assigning weight to words based on their importance in the document.

Feature Selection

Feature selection is the process of choosing the most relevant attributes or features for classification. In online transportation sentiment analysis, commonly used features include frequently occurring keywords in customer comments. By analyzing specific words in tweets, it is possible to determine whether a review carries a positive or negative sentiment (Novaneliza et al., 2023).

Justification for Algorithm Selection Naïve Bayes and Support Vector Machine (SVM) were selected based on their effectiveness in text classification. Naïve Bayes is computationally efficient and works well with small datasets, while SVM provides robust classification in high-dimensional spaces, making it suitable for sentiment analysis (Singgalen, 2021). Validation Technique to assess model performance and prevent overfitting, 5-fold cross-validation was employed. This method ensures that the dataset is split into different training and testing subsets, improving generalizability.

Classification

Classification is the process of dividing data into specific categories based on patterns found in the dataset. In sentiment analysis, classification is used to determine whether a review has a positive, negative, or neutral sentiment. Some commonly used classification methods include:

1. Naïve Bayes

A probability-based algorithm frequently used in text analysis. It works by calculating the likelihood of a word appearing in a specific category based on prior training data. Naïve Bayes is effective for sentiment analysis due to its simplicity while still providing fairly accurate results (Novaneliza et al., 2023).

2. Support Vector Machine (SVM)

A hyperplane-based algorithm that separates data into two or more categories. SVM is widely used in research due to its ability to handle high-dimensional data and provide more accurate classifications compared to other methods. Studies comparing classification techniques often show that SVM outperforms Naïve Bayes in sentiment analysis (Nalle et al., 2023).

Performance Evaluation

Accuracy is the most intuitive performance metric, representing the ratio of correctly predicted observations to the total observations. Precision measures the ratio of correctly predicted positive observations to the total predicted positive observations. Recall, on the other hand, evaluates how well the model identifies all actual positive observations in the dataset. Lastly, F1-Score is the weighted average of precision and recall, considering both false positives and false negatives, making it a useful metric when dealing with imbalanced datasets (Fadlisyah & Muhathir, 2023).

To ensure robustness and prevent overfitting, K-Fold Cross-Validation (k=5) was applied. This method divides the dataset into five equal parts, where four parts are used for training and one part for testing in each iteration. The final performance metrics are obtained by averaging the results across all five folds, ensuring a more generalized evaluation of the model (Wijiyanto et al., 2024).

Visualization

Visualization in data analysis is the process of presenting data in graphical or visual form to facilitate understanding and interpretation. In the context of sentiment analysis, visualization is often used to display the distribution of positive and negative sentiments through various methods such as histograms, word clouds, and other graphical representations (Iwandini et al., 2023).

RESULTS AND DISCUSSION

The sentiment analysis results show that 93.5% of the tweets are neutral, indicating that most discussions on transportation do not express strong positive or negative opinions. This may be due to the nature of transportation-related conversations on social media, where users provide factual statements rather than subjective opinions. Additionally, dataset imbalance could contribute to this distribution, as the majority of collected tweets were informational rather than opinion-based. To further validate model performance, K-Fold Cross-Validation (k=5) was applied, ensuring a more generalized evaluation. The results showed an average accuracy of 98.4% across all folds, confirming the robustness of the classification model while mitigating potential overfitting.

However, despite high accuracy, the imbalanced dataset may cause biased predictions, and future studies should explore techniques like SMOTE (Synthetic Minority Over-sampling Technique) to improve class distribution. A deeper analysis of sentiment keywords provides insights into public perception. Negative sentiment tweets frequently mention "kemacetan" (traffic congestion), "penundaan transportasi" (transportation delays), and "harga" (cost), indicating dissatisfaction with urban mobility, public transport reliability, and affordability.

On the other hand, positive tweets highlight aspects such as "nyaman" (comfortable) and "bersih" (clean), suggesting that some transportation services meet public expectations. These findings have significant implications for transportation policy. Policymakers can leverage sentiment analysis to identify pressing issues and develop targeted solutions. For example:

1. Frequent mentions of "kemacetan" suggest an urgent need for improved traffic management and public transport incentives.
2. Concerns about "harga" indicate that affordability remains a critical barrier, highlighting the importance of subsidies or fare adjustments.
3. The dominance of neutral sentiments suggests limited public engagement, emphasizing the need for better communication strategies between transport authorities and users.

Further research should incorporate more advanced NLP techniques, such as Topic Modeling (LDA) and Sentiment Score Analysis, to extract deeper insights. Expanding the dataset across multiple social media platforms may also help capture a more diverse range of opinions on transportation services in Indonesia.

Data Collection

The data collection process in this study consists of two key steps: data extraction and cleaning. The first step, data extraction, involves retrieving tweets related to transportation in Indonesia using predefined keywords via the Twitter API. The retrieved data is then stored in the appropriate format for further processing. The second step, data cleaning, focuses on refining the dataset by eliminating irrelevant elements to enhance its quality and usability.

The final dataset is categorized based on sentiment analysis, comprising 7 positive reviews, 6 negative reviews, and 187 neutral reviews. This distribution ensures a well-represented dataset for both training and evaluation purposes.

Preprocessing Data

After data collection, the next stage is preprocessing, which aims to clean and prepare the data before analysis. Some important steps in preprocessing include:

1. Cleaning

Cleaning the text from irrelevant elements is the first step. This involves removing special characters, emojis, numbers, and unnecessary punctuation. Then, the text is broken down into smaller components through a process called tokenization (Khairunnisa et al., 2021).

```
0  Hai kak. Rute 4B: Universitas Indonesia - Stas...
0  hai kak rute universitas indonesia stasiun man.
```

Source : (Research Result, 2025)

Figure 2. Cleaning

2. Stopword Removal

Words that do not have significant meaning in the analysis, such as conjunctions ("and," "or," "from"), are removed using the stopword removal technique. To simplify words into their root form, stemming or lemmatization is applied, depending on the research needs. The result of this stage is cleaner text, ready for analysis (Khairunnisa et al., 2021).

```
3  Menurut laporan e-Conomy SEA 2024, sektor tran...
3  lapor economy sea sektor transportasi makan on...
```

Source : (Research Result, 2025)

Figure 3. Stopword Removal

3. Stemming

Converting words into their root form. For example, "berjalan" becomes "jalan," "membantu" becomes "bantu," and "dilayani" becomes "layan." Stemming is highly useful for reducing variations of words that have the same meaning, thereby improving the accuracy of text analysis (Novaneliza et al., 2023).

```
4  Menko AHY Ajak Para Perwira Transportasi Siap ...
4  menko ahy ajak perwira transportasi emban tuga...
```

Source : (Research Result, 2025)

Figure 4. Stemming

4. Normalization

Normalization is a stage in data preprocessing that aims to standardize data formats, making them more uniform and easier to process in further analysis (Khairunnisa et al., 2021).

```
keren banget nih mobil listrik Hyundai IONIQ 5 ...

keren mobil listrik hyundai ioniq ...
```

Source : (Research Result, 2025)

Figure 5. Normalization

5. IndoBert

IndoBERT is a BERT-based (Bidirectional Encoder Representations from Transformers) language model specifically trained to understand the Indonesian language. This model is a version of BERT adapted to the characteristics of the Indonesian language by using a large Indonesian language corpus in its pretraining process (Geni et al., 2023).

```
                                             Content Sentiment
0  Hai kak. Rute 4B: Universitas Indonesia - Stas...   Neutral
1  Keren banget nih mobil listrik Hyundai IONIQ 5...  Positive
2  Hai EPIzen\n\nPT PLN (Persero) melalui Subhold...   Neutral
3  Menurut laporan e-Conomy SEA 2024, sektor tran...   Neutral
4  Menko AHY Ajak Para Perwira Transportasi Siap ...   Neutral
```

Source : (Research Result, 2025)

Figure 6. IndoBert

Classification Method

This study employs two classification methods: Naïve Bayes (NB) and Support Vector Machine (SVM) to analyze sentiment in transportation-related tweets. The classification process was carried out after preprocessing and feature extraction using IndoBERT. To address data imbalance. The results show that SVM achieved an accuracy of 100%, with both precision and recall values reaching 1.00 for both sentiment classes. This indicates that SVM effectively differentiates between positive and neutral sentiments with perfect classification. On the other hand, Naïve Bayes attained an accuracy of 92%, with slightly lower recall for positive sentiment (0.86). While NB performed well, its probabilistic assumptions might have limited its ability to fully capture the contextual dependencies in the text data. These findings suggest that SVM is the superior classifier for this dataset, benefiting from its ability to handle high-dimensional text features and its robust separation of sentiment classes.

Table 1. Result Using SVM Method

	Precision	Recall	F1-Score	Support
Neutral	1.00	1.00	1.00	39
Positive	1.00	1.00	1.00	38
Accuracy			1.00	77
Macro Avg	1.00	1.00	1.00	77
Weighted Avg	1.00	1.00	1.00	77

Source : (Research Result, 2025)

The evaluation results of the Support Vector Machine (SVM) model demonstrate outstanding performance in sentiment classification. The precision for both Neutral and Positive classes reaches 1.00, meaning every prediction made by the model in these classes is completely accurate. Additionally, the recall value is also 1.00, indicating that the model successfully identifies all tweets in each class without misclassification. The F1-score, which balances precision and recall, also achieves a perfect score of 1.00 for both classes.

Regarding the test data count (support), there are 39 data points for the Neutral class and 38 for the Positive class, totaling 77 test samples. The model's overall accuracy reaches 100%, meaning there are no misclassifications in the test data. Both the macro average and weighted average scores also achieve 1.00, indicating that the model performs exceptionally well without bias toward any specific class.

Table 2. Result Using Naive Bayes Method

	Precision	Recall	F1-Score	Support
Neutral	1.00	0.85	0.92	39
Positive	0.86	1.00	0.93	38
Accuracy			0.92	77
Macro Avg	0.93	0.92	0.92	77
Weighted Avg	0.93	0.92	0.92	77

Source : (Research result, 2025)

The evaluation results of the Naïve Bayes model show a fairly good performance in sentiment classification, although it is not as accurate as the SVM model. The precision for the Neutral class reaches 1.00, meaning all predictions made by the model for this class are correct. However, the recall for the same class is only 0.85, indicating that 15% of Neutral data was misclassified as Positive. Conversely, for the Positive class, the precision is lower at 0.86, meaning some Positive data was misclassified as Neutral. However, the recall for the Positive class is 1.00, indicating that all positive sentiment data was correctly identified by the model without any being overlooked.

The F1-score for both classes is 0.92 (Neutral) and 0.93 (Positive), demonstrating a good balance between precision and recall. The overall accuracy of the model is 92%, meaning that out of 77 test data points, around 8% were misclassified.
Both the macro average and weighted average scores range between 0.92 and 0.93, indicating stable performance across both classes. While these

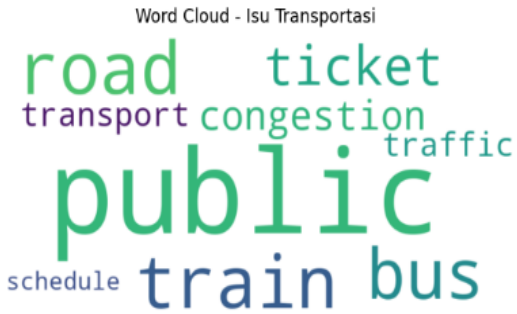
Vol. 22, No. 2 September 2025 | DOI https://doi.org/10.33480/techno.v22i2.6577

results are quite good, the performance of Naïve Bayes is slightly lower compared to SVM, primarily due to its assumption of feature independence, which is often less suitable for text-based data.

This model can still be optimized through better feature processing techniques or by increasing the amount of training data to improve its generalization capability.

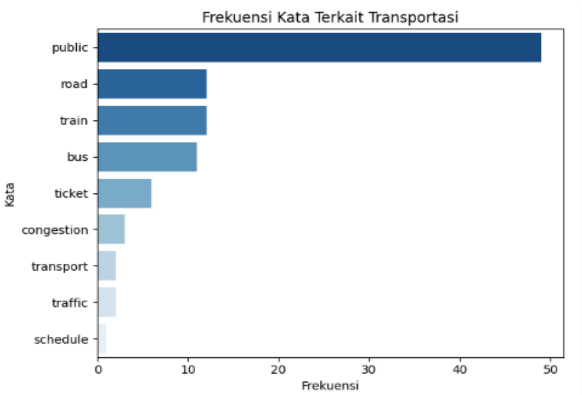
Visualization

Once the data was preprocessed and categorized, the sentiment analysis results were represented through a word cloud, emphasizing the most commonly appearing words based on their frequency. Words that appear larger indicate higher occurrences, allowing for quick identification of dominant keywords within the text data. This visualization was created using a Python script.



Source : (Research Result, 2025)
Figure 7. Wordcloud Isu Transportasi

The word cloud displays the most frequently occurring words in discussions about transportation. The size of the words in the word cloud reflects their frequency in the dataset the larger the word, the more frequently it appears. Based on the results, the words "public," "road," and "train" are the most dominant, indicating that discussions in the dataset are largely related to public transportation, roads, and trains.



Source : (Research Result, 2025)
Figure 8. Bar Chart

This bar chart illustrates the distribution of word occurrences related to transportation. The word "public" has the highest frequency (49 times), followed by "road" and "train" (each appearing 12 times). Words like "bus," "ticket," and "congestion" also appear, though with lower frequencies. This indicates that discussions about public transportation and road infrastructure are quite dominant in the dataset.

Table 3. Transportation Word Frequency Table

	Precision	Recall
1	Public	49
0	Road	12
3	Train	12
2	Bus	11
8	Ticket	6
4	Congestion	3
6	Transport	2
7	Traffic	2
5	Schedule	1

Source : (Research result, 2025)

This table provides a more detailed breakdown of the occurrence of each word featured in the word cloud and bar chart. With a more systematic format, we can observe that:

1. "public" appears 49 times, reinforcing that public transportation is the main topic.
2. "road" and "train" each appear 12 times, highlighting a focus on road and rail infrastructure.
3. Other words like "bus" (11 times), "ticket" (6 times), "congestion" (3 times), and "transport" (2 times) add context to the ongoing transportation-related discussions.

This dataset primarily discusses public transportation, road conditions, as well as tickets and congestion as relevant issues. These findings can be linked to transportation policies, such as improving public transit services, traffic management, and ticketing regulations.

CONCLUSION

This study highlights that public sentiment toward transportation services in Indonesia is predominantly neutral (93.5%), which may be attributed to dataset imbalance, the nature of discussions on social media, or limited emotional engagement in transportation-related topics. Support Vector Machine (SVM) achieved an accuracy of 100%, while Naïve Bayes reached 92%, reinforcing the importance of algorithm selection in sentiment classification. To enhance future research, it is recommended to expand the dataset, integrate advanced NLP techniques such as topic modeling (LDA) for deeper analysis, and conduct cross-platform studies for broader insights.

Furthermore, linking sentiment analysis with transportation policy-making processes can facilitate data-driven decision-making, improving service quality and addressing key public concerns effectively.

REFERENCE

A. Pratama & T. W. Hadi. (2022). Utilizing Social Media as a Data Source for Analyzing Public Perception of Public Transportation. Journal of Information Systems, 10, n.

D. Kusuma et al. (2021). Implementation of Machine Learning in Public Sentiment Analysis of App-Based Transportation Services. Journal of Information Technology, 18, n.

Dwianto, E., & Sadikin, M. (2021). Sentiment Analysis of Online Transportation on Twitter Using Naïve Bayes and Support Vector Machine Classification Methods. Format: Scientific Journal of Informatics Engineering, 10(1), 94. https://doi.org/10.22441/format.2021.v10.i1.009

Fadlisyah, F., & Muhathir, M. (2023). Performance Evaluation Of Variations Boosting Algorithms For Classifying Formalin Fish From Photos. *Journal of Informatics and Telecommunication Engineering*, *6*(2), 621–629. https://doi.org/10.31289/jite.v6i2.6614

Geni, L., Yulianti, E., & Sensuse, D. I. (2023). Sentiment Analysis of Tweets Before the 2024 Elections in Indonesia Using IndoBERT Language Models. *Jurnal Ilmiah Teknik Elektro Komputer Dan Informatika (JITEKI)*, *9*(3), 746–757. https://doi.org/10.26555/jiteki.v9i3.26490

Hidayati, I. (2023). Breaking the commute barrier: How women in Jabodetabek overcome daily challenges on commuting for work. *ETNOSIA : Jurnal Etnografi Indonesia*, *8*(1), 44–62. https://doi.org/10.31947/etnosia.v8i1.26085

Iwandini, I., Triayudi, A., & Soepriyono, G. (2023). Sentiment Analysis of Jakarta Transportation Users Towards TransJakarta Using Naïve Bayes and K-Nearest Neighbor Methods. Journal of Information System Research *(JOSH)*, *4*(2), 543–550. https://doi.org/10.47065/josh.v4i2.2937

Khairunnisa, S., Adiwijaya, A., & Faraby, S. Al. (2021). Effect of Text Preprocessing on Sentiment Analysis of Public Comments on Twitter (Case Study: COVID-19 Pandemic). Journal of Media Informatics Budidarma, *5*(2), 406. https://doi.org/10.30865/mib.v5i2.2835

Khoiruddin, Y., Fauzi, A., & Siregar, A. M. (2023). Sentiment Analysis of Gojek Indonesia on Twitter Using Naïve Bayes and Support Vector Machine Algorithms. Scientific Journal of Computer Science, *19*, 391–400.

M. Yusuf & L. Kurniawan. (2023). Comparison of Support Vector Machine and Deep Learning Algorithms in Online Transportation Sentiment Analysis. Journal of Artificial Intelligence, 5, no.

N. Dewi & P. Santoso. (2022). Analysis of Public Opinion Trends on Public Transportation Through Social Media. Journal of Computer Science, 12, n.

Nalle, V. I. W., Syaputri, M. D., Krisnanto, W., & Tjandra, O. C. P. (2023). Public Participation in Bus Transit Policymaking: The Case of Semarang, Indonesia. *International Journal of Transport Development and Integration*, *7*(3), 235–245. https://doi.org/10.18280/ijtdi.070307

Novaneliza, R., Handayani, F., Suhandar, R. J., Surono, H., Azzahra, N. S., & Nadilla, D. (2023). Comparison of Algorithms for Sentiment Analysis on Twitter for Public Transportation Commuter Line. Journal of Computer Science & Informatics *(J-SAKTI*, *7*(1), 13–21.

Pratama, D., Rifai, A. I., & Handayani, S. (2023). The Passenger Satisfaction Analysis of Commuter Line in the New-Normal Period. *Indonesian Journal of Multidisciplinary Science*, *1*(1), 409–418. https://doi.org/10.55324/ijoms.v1i1.398

R. Sari & B. Nugroho. (2021). **S**entiment Analysis of Online Transportation Users on Twitter Using Naïve Bayes Method. Journal of Informatics, 14, n.

Said, L. B., & Syafey, I. (2022). User acceptance oF public transport systems based on a perception model. *International Journal of Transport Development and Integration*, *6*(4), 399–414. https://doi.org/10.2495/TDI-V6-N4-399-414

Savita, D. A., Putra, I. K. G. D., & Rusjayanthi, N. K. D. (2021). Public Sentiment Analysis of Online Transportation in Indonesia through Social Media Using Google Machine Learning. *Jurnal Ilmiah Merpati (Menara Penelitian Akademika Teknologi Informasi)*, *9*(2), 153. https://doi.org/10.24843/jim.2021.v09.i02.p06

Setiawan, Y., Jondri, J., & Astuti, W. (2022). Twitter Sentiment Analysis on Online Transportation in Indonesia Using Ensemble Stacking. *Jurnal Media Informatika Budidarma*, *6*(3), 1452. https://doi.org/10.30865/mib.v6i3.4359

Singgalen, Y. A. (2021). Selection of Methods and Algorithms in Sentiment Analysis on Social Media: Systematic Literature Review. Journal of Information Systems and Informatics, *3*(2), 278–302. https://doi.org/10.33557/journalisi.v3i2.125

T. Ramadhani & R. Widodo. (2021). Twitter-Based Sentiment Analysis for Public Transportation Service Evaluation. Journal of Data Science, 6, no.

Wijiyanto, W., Pradana, A. I., Sopingi, S., & Atina, V. (2024). K-Fold Cross Validation Technique for Evaluating Student Performance. Journal of Algorithms, *21*(1), 239–248. https://doi.org/10.33364/algoritma/v.21-1.1618