JITET (Jurnal Informatika dan Teknik Elektro Terapan)

J I T E T
JURNAL INFORMATIKA DAN TEKNIK ELEKTRO TERAPAN

Vol. 13 No. 3S1, pISSN: 2303-0577 eISSN: 2830-7062 http://dx.doi.org/10.23960/jitet.v13i3S1.8148

HYBRID SENTIMENT ANALYSIS OF MAXIM APP USERS USING SUPPORT VECTOR MACHINE AND LEXICON-BASED APPROACH

Dalila Fatharani[1*], Elfitrin Syahrul[2]

[1,2]Universitas Gunadarma; Jl. Margonda Raya 100, Depok 16424; 021-78881112

Keywords:
Sentiment Analysis; Lexicon-Based; Maxim; SVM; TF-IDF.

Corespondent Email:
elfitrin@staff.gunadarma.ac.id

Abstract. Online transportation applications such as Maxim have rapidly grown alongside technological advancements. These platforms accumulate large volumes of user reviews on sources like the Google Play Store, providing valuable insights into user perceptions. However, the unstructured nature of textual data makes systematic analysis difficult. This study proposes a sentiment classification model to categorize Maxim user reviews into positive and negative sentiments, excluding neutral responses. The method integrates a lexicon-based approach using the InSet Lexicon with a Support Vector Machine (SVM) classifier. Preprocessing steps included text cleaning, case folding, normalization, tokenization, stopword removal, and stemming. Feature extraction was conducted using Term Frequency–Inverse Document Frequency (TF-IDF), followed by sentiment classification with SVM. Evaluation using a confusion matrix achieved an accuracy of 96.07%. For negative sentiment, the model obtained a precision of 79%, recall of 83%, and F1-score of 81%; for positive sentiment, precision was 89%, recall 98%, and F1-score 93%. These results indicate that integrating lexical resources with machine learning provides an effective solution for sentiment analysis of user-generated reviews.

1. INTRODUCTION

Digital platforms have significantly transformed the landscape of online transportation services, offering greater accessibility, efficiency, and convenience [1]. In Indonesia, user adoption surged by late 2024, with approximately 88.3 million individuals-or 66.7% of internet users aged 16 and above-engaging with online transportation platforms [2]. This rapid growth positioned Indonesia as the largest online transportation market in the world [2].

Maxim, one of the leading ride-hailing platforms, launched its services in 2018 and has since expanded to over 200 cities across the country [3]. By 2023, the application recorded an average of 892 thousand downloads per month—a 15.39% increase from the previous year [4]. As of 2025, Maxim surpassed 100 million total downloads, with users submitting more than 5 million reviews and assigning an average rating of 4.7 out of 5 [4].

Despite consistently high numerical ratings, these scores often fail to reflect the full range of user experiences. Textual reviews on platforms such as the Google Play Store provide richer and more nuanced insights into user satisfaction and service quality. However, the vast amount of unstructured text data makes manual interpretation difficult. Therefore, automated sentiment analysis has emerged as an effective solution for extracting opinions and emotional tone from large-scale user-generated content [5]–[7].

Several studies have explored various sentiment analysis approaches. Lexicon-based methods rely on predefined sentiment dictionaries to classify text polarity [8], while machine learning algorithms such as Support Vector Machine (SVM) offer higher adaptability in complex classification tasks [9], [10], [11]. A comprehensive review by [12] emphasizes that hybrid approaches combining lexical and statistical techniques tend to improve accuracy, particularly when supported by proper preprocessing and feature extraction. In the transportation domain, [13] highlights that sentiment analysis and topic modeling have been increasingly used to understand passenger behavior, service perception, and operational challenges, illustrating the relevance of such techniques for mobility research.

However, limited research has specifically focused on sentiment analysis for Maxim users in Indonesia, particularly leveraging local linguistic resources such as the Indonesian Sentiment Lexicon (InSet) in combination with SVM classification. Addressing this gap, the present study proposes a hybrid lexicon–machine learning approach to classify Maxim user reviews into positive and negative sentiments. The findings are expected to contribute to a deeper understanding of user perceptions and provide actionable insights for service improvement within Indonesia's growing digital mobility ecosystem.

2. LITERATURE REVIEW

2.1. Sentiment Analysis

Sentiment analysis (opinion mining) is the computational study of opinions, attitudes, and emotions expressed in text; its goal is to automatically extract and classify sentiment polarity (positive, negative, sometimes neutral) from unstructured user-generated content such as reviews and social media posts [14]–[16]. Sentiment analysis transforms large volumes of textual feedback into actionable insights for decision makers in commercial and public sectors, including transportation and mobile apps. Recent reviews emphasize that modern sentiment analysis combines careful preprocessing, robust feature engineering, and appropriate model choice to handle noise, colloquial language, and domain-specific vocabulary in real-world datasets [17].

2.2. Lexicon-Based Approach

Lexicon-based methods use sentiment dictionaries that map words or phrases to polarity labels and (optionally) intensity scores; they compute an aggregate score for text units to infer polarity. Lexicon techniques are attractive when labeled training data are scarce because they do not require supervised learning and offer interpretability - a key benefit in applied settings where explainability matters. However, lexicon methods can struggle with context, idioms, sarcasm, and domain-specific usage; hence researchers often adapt lexica (or combine lexicon outputs with machine learning models) to improve robustness. The Indonesian InSet lexicon is an established resource for Indonesian language sentiment tasks and has been used as the lexical basis for several applied studies [18].

2.3. Support Vector Machine (SVM)

Support Vector Machine is a classical supervised learning algorithm that has shown strong performance in text classification tasks when paired with appropriate feature representations such as TF-IDF. SVM constructs a maximum-margin hyperplane to separate classes in high-dimensional feature space and is often competitive with more complex models on moderate-sized review datasets. Empirical comparisons across studies show TF-IDF + SVM remains a cost-effective, performant baseline for sentiment classification, and can be further improved via kernel selection, parameter tuning, and hybridization with lexicon outputs or rule-based features [19].

2.4. Previous Studies

A few studies have explored how to combine lexicon-based methods with machine learning algorithms such as Support Vector Machine (SVM) and Term Frequency–Inverse Document Frequency (TF-IDF) for sentiment analysis [17]. Researchers have applied these techniques mainly to app and online store reviews to better understand user opinions. Several studies have developed hybrid lexicon–SVM models that can handle large and noisy text data while reducing the need for manual labeling, especially in tourism and big data applications [19]. Other works have compared TF-IDF with modern embedding methods like Word2Vec and contextual embeddings, finding that although deep learning models often perform better on massive datasets, TF-IDF with SVM still produces strong results for medium-sized corpora, offering faster processing and lower computational cost [20], [18]. Research on Google Play Store reviews also confirms that TF-IDF and SVM, combined with text preprocessing techniques such as cleaning, tokenizing, and stemming, achieve high classification accuracy [21]. In Indonesia, the InSet lexicon has been widely used for sentiment labeling in the local language and is effective in hybrid lexicon–machine learning frameworks [22]. However, few studies have examined Maxim app reviews using the InSet–SVM hybrid method [9], [23]–[26]. This study aims to fill that research gap.

3. METHODS

The methodological framework adopted in this study comprises eight sequential stages: data scraping, preprocessing, labeling, data splitting, feature extraction using Term Frequency–Inverse Document Frequency (TF-IDF), classification using Support Vector Machine (SVM), and model evaluation with visualization as seen in Figure 1.

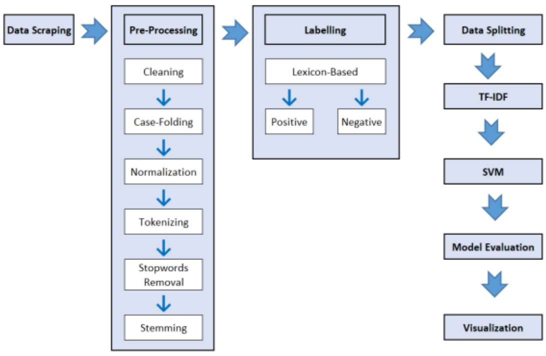

Figure 1. Methodological Framework

4. RESULTS AND DISCUSSION

4.1. Data Scraping

User reviews of the Maxim application were retrieved from the Google Play Store using automated data scraping techniques. The process focused on extracting textual review content to build a comprehensive dataset for sentiment analysis. Ethical considerations were maintained by adhering to platform usage policies and excluding any personally identifiable information.

A total of 2,000 reviews were collected on April 17, 2025. The scraping was implemented using the Python programming language and the google-play-scraper library. The keyword 'com.taxsee.taxsee' was used to locate and extract relevant reviews of the Maxim app. Four columns—userName, score, at, and content—were selected and stored in .csv format to serve as the primary dataset. The results of the scraping process and CSV data extraction are presented in Table 1.

Table 1. Sample Output From Data Scraping Process

	userName	score	at	content
0	Arvin Jody	5	2025-04-17 11:08:54	bagus dan rekomended karna pengantar yg sangat...
1	Raden alif al-khalif	5	2025-04-17 11:05:11	ok...😍
2	Kiki Alfionita 650 034	5	2025-04-17 11:03:02	baik
3	Karyono SS	5	2025-04-17 10:57:53	Ojol samarinda
4	hendry ryan	5	2025-04-17 10:55:48	Mantaaap

4.2. Pre-processing

Enhancing model performance required a series of preprocessing steps aimed at reducing noise and standardizing the input data for analysis. The procedure encompassed data cleaning, case folding, normalization, tokenization, stopword removal, and stemming. These preprocessing steps produced linguistically consistent and computationally efficient input for subsequent sentiment classification. Table 2. presents the output generated at each stage of the preprocessing process.

Table 2. Output of Each Preprocessing Stage

	userName	score	at	content	cleaning	case_folding	normalisasi	tokenize	stopwords removal	steming_data
0	Arvin Jody	5	2025-04-17 11:08:54	bagus dan rekomended karna pengantar yg sangat...	bagus dan rekomended karna pengantar yg sangat...	bagus dan rekomended karna pengantar yg sangat...	bagus dan rekomended karena pengantar yang san...	[bagus, dan, rekomended, karena, pengantar, ya...	[bagus, rekomended, pengantar, cepat]	bagus rekomended antar cepat
1	Raden alif al-khalif	5	2025-04-17 11:05:11	ok...😍	ok	ok	ok	[ok]	[ok]	ok
2	Kiki Alfionita 650 034	5	2025-04-17 11:03:02	baik	baik	baik	baik	[baik]	[]	
3	Karyono SS	5	2025-04-17 10:57:53	Ojol samarinda	Ojol samarinda	ojol samarinda	ojol samarinda	[ojol, samarinda]	[ojol, samarinda]	ojol samarinda
4	hendry ryan	5	2025-04-17 10:55:48	Mantaaap	Mantaaap	mantaaap	mantap	[mantap]	[mantap]	mantap
5	Fauzan Azhari	5	2025-04-17 10:51:38	sukses selalu	sukses selalu	sukses selalu	sukses selalu	[sukses, selalu]	[sukses]	sukses
6	Wahyu Ida	5	2025-04-17 10:51:35	pelayanan ramah	pelayanan ramah	pelayanan ramah	pelayanan ramah	[pelayanan, ramah]	[pelayanan, ramah]	layan ramah
7	Putri utari Fillah	5	2025-04-17 10:42:28	sopan baik	sopan baik	sopan baik	sopan baik	[sopan, baik]	[sopan]	sopan
8	Dedi Kasih	1	2025-04-17 10:41:57	angkosnya kecil, kasiann drivernya, OK ada s...	angkosnya kecil kasiann drivernya OK ada stak...	angkosnya kecil kasiann drivernya gk ada stak ...	angkosnya kecil kasiann drivernya tidak ada st...	[angkosnya, kecil, kasiann, drivernya, tidak, ...	[angkosnya, kasiann, drivernya, tidak, stak, a...	angkos kasiann drivernya tidak stak aplikasi m...
9	Samsul Maarif	5	2025-04-17 10:40:35	Pelayanan Driver MAXIM Makassar Sangat Baik	Pelayanan Driver MAXIM Makassar Sangat Baik	pelayanan driver maxim makassar sangat baik	pelayanan driver maxim makassar sangat baik	[pelayanan, driver, maxim, makassar, sangat, b...	[pelayanan, driver, maxim, makassar]	layan driver maxim makassar
11	Melly Yana	5	2025-04-17 10:30:30	apk yg sangat membantu	apk yg sangat membantu	apk yg sangat membantu	apk yang sangat membantu	[apk, yang, sangat, membantu]	[apk, membantu]	apk bantu
13	Kyyzzz	5	2025-04-17 10:22:15	mantap	mantap	mantap	mantap	[mantap]	[mantap]	mantap
15	Machfud Machfud	5	2025-04-17 10:20:42	bagus	bagus	bagus	bagus	[bagus]	[bagus]	bagus

The textual data underwent a structured preprocessing pipeline aimed at enhancing consistency, minimizing noise, and optimizing feature extraction for sentiment analysis. As seen in the Figure 1, the key steps of pre-processing stages included:

1. Data cleaning, which eliminated irrelevant characters such as punctuation marks, symbols, and excessive whitespace [6].
2. Case folding, applied to convert all text to lowercase, ensuring uniform representation across tokens [27].

3. Normalization, used to resolve inconsistencies in spelling and formatting, thereby standardizing textual input [28].
4. Tokenization, which segmented the text into discrete units—typically words or meaningful tokens [29].
5. Stopword removal, performed to exclude commonly used words that contribute little semantic value (e.g., "yang," "di," "dan") [10].
6. Stemming, implemented to reduce words to their base or root forms, consolidating morphological variants of the same term [28].

4.3. Labelling (Lexicon-Based Approach)

We applied a lexicon-based method to label sentiments in the dataset. The process identified 997 reviews as positive and 147 as negative from a total of 1,144 user reviews. Table 3 presents the results of the sentiment labeling.

Table 3. Example of Sentiment Labelling Output

	userName	score	at	content	cleaning	case_folding	normalisasi	tokenize	stopwords removal	steming_data	Sentiment
0	Arvin Jody	5	2025-04-17 11:08:14	bagus dan rekomended karna pengantar yg sangat...	bagus dan rekomended karna pengantar yg sangat...	bagus dan rekomended karna pengantar yg sangat...	bagus dan rekomended karena pengantar yang san...	['bagus', 'dan', 'rekomended', 'karena', 'peng...	['bagus', 'rekomended', 'pengantar', 'cepat']	bagus rekomended antar cepat	Positif
1	[illegible]	5	2025-04-17 11:05:11	ok..	ok	ok	ok	['ok']	['ok']	ok	Positif
2	hendry ryan	5	2025-04-17 10:55:48	Mantaaap	Mantaaap	mantaaap	mantap	['mantap']	['mantap']	mantap	Positif
3	Fauzan Azhari	5	2025-04-17 10:51:38	sukses selalu	sukses selalu	sukses selalu	sukses selalu	['sukses', 'selalu']	['sukses']	sukses	Positif
4	[illegible]	5	2025-04-17 10:51:35	pelayanan ramah	pelayanan ramah	pelayanan ramah	pelayanan ramah	['pelayanan', 'ramah']	['pelayanan', 'ramah']	layan ramah	Positif
5	[illegible]	1	2025-04-17 10:41:57	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	Negatif
6	[illegible]	5	2025-04-17 10:40:35	Pelayanan Driver MAXIM Makassar Sangat Baik.	Pelayanan Driver MAXIM Makassar Sangat Baik	pelayanan driver maxim makassar sangat baik	pelayanan driver maxim makassar sangat baik	['pelayanan', 'driver', 'maxim', 'makassar', '...	['pelayanan', 'driver', 'maxim', 'makassar']	layan driver maxim makassar	Positif
7	[illegible]	5	2025-04-17 10:38:38	apk yg sangat membantu	apk yg sangat membantu	apk yg sangat membantu	apk yang sangat membantu	['apk', 'yang', 'sangat', 'membantu']	['apk', 'membantu']	apk bantu	Positif
8	[illegible]	5	2025-04-17 10:22:15	mantap	mantap	mantap	mantap	['mantap']	['mantap']	mantap	Positif
9	[illegible]	5	2025-04-17 10:20:42	bagus	bagus	bagus	bagus	['bagus']	['bagus']	bagus	Positif
10	[illegible]	5	2025-04-17 10:15:52	ramah	ramah	ramah	ramah	['ramah']	['ramah']	ramah	Positif
11	[illegible]	5	2025-04-17 10:15:04	ramah cepat	ramah cepat	ramah cepat	ramah cepat	['ramah', 'cepat']	['ramah', 'cepat']	ramah cepat	Positif
12	[illegible]	5	2025-04-17 10:14:28	baik dan ramah	baik dan ramah	baik dan ramah	baik dan ramah	['baik', 'dan', 'ramah']	['ramah']	ramah	Positif
13	[illegible]	5	2025-04-17 10:11:43	good	good	good	good	['good']	['good']	good	Positif
14	[illegible]	5	2025-04-17 10:11:02	bagus	bagus	bagus	bagus	['bagus']	['bagus']	bagus	Positif
15	[illegible]	5	2025-04-17 09:59:43	layanan pengemudi ramah	layanan pengemudi ramah	layanan pengemudi ramah	layanan pengemudi ramah	['layanan', 'pengemudi', 'ramah']	['layanan', 'pengemudi', 'ramah']	layan kemudi ramah	Positif

A custom program accessed the InSet lexicon [24], through two dictionary files—positive.tsv and negative.tsv—to identify words with positive or negative polarity in the review texts. Each review was assigned a sentiment label based on the cumulative score of matched terms. Reviews exceeding the positive threshold were labeled as positive, while those falling below the negative threshold were labeled as negative.

4.4. Data Splitting

The labeled dataset was divided into two subsets—training data and testing data—using an 80:20 ratio. The training subset, comprising 915 reviews, trained the classification model to recognize patterns and relationships within the labeled data. The testing subset, containing 229 reviews, evaluated the model's performance on unseen inputs and measured its generalization ability. This partitioning strategy ensured a balanced approach between effective model learning and unbiased validation.

4.5. Term Frequency-Inverse Document Frequency (TF-IDF)

The term weighting process automatically applied the Term Frequency–Inverse Document Frequency (TF-IDF) method to transform textual data into numerical features that reflected each word's importance. Term Frequency (TF) measured how often a term appeared in a given document, while Inverse Document Frequency (IDF) quantified how unique that term was across the entire corpus. Words that appeared frequently in one review but rarely in others received higher weights, indicating stronger discriminative power for sentiment analysis.

Table 4. Sample of TF-IDF Results

	Document	Token	TF	DF	IDF	TF-IDF
0	1	layan	1	109	3.119.536	1.000000
1	2	aman	1	18	4.875.577	0.696031
2	2	cepat	1	104	3.166.056	0.451982
3	2	nyaman	1	49	3.907.993	0.557900
4	3	ya	1	152	2.789.578	1.000000
5	4	kasih	1	50	3.888.191	0.691765
6	4	terima	1	42	4.058.816	0.722122
7	5	abang	1	23	4.641.963	0.301731
8	5	bantu	1	31	4.354.280	0.283032
9	5	driver	1	127	2.967.986	0.192921
10	5	gercep	1	8	5.622.792	0.365486
11	5	jalan	1	32	4.323.509	0.2811031
12	5	maxim	1	78	3.450.569	0.224290
13	5	ramah	1	210	2.468.158	0.160432
14	5	rapi	1	1	7.126.869	0.463252
15	5	terimakasih	2	39	4.131.137	0.537054

Table 4 presents the computed TF-IDF values. For instance, the token "layan" in Document 1 achieved the highest normalized TF-IDF score of 1.0000, signifying its strong influence within that document. Conversely, common terms such as "*ramah*" (0.1604) and "*driver*" (0.1929) received lower scores due to their frequent occurrence across multiple documents. Rare yet meaningful tokens such as "*baharu*" (0.4484) and "*rapi*" (0.4632) gained moderate weights, suggesting distinctive sentiment cues. This weighting mechanism ensured that salient words contributed more to the classification model, while overly common words exerted less influence.

4.6. SVM Classification Results

A The Support Vector Machine (SVM) model utilized the TF–IDF feature matrix as input for sentiment classification. Model performance was evaluated using the confusion matrix shown in Figure 2. The matrix indicates that 201 samples were correctly classified as True Positive (TP) and 19 as True Negative (TN), while 5 samples were incorrectly predicted as False Negative (FN) and 4 as False Positive (FP).

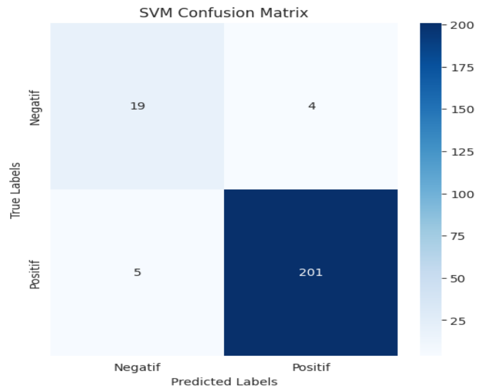

Figure 2. SVM Confusion Matrix

This distribution demonstrates the model's strong capability in detecting positive sentiment, as reflected by the high count of true positives and the minimal number of false positives. However, the relatively lower true negative count and the presence of false negatives indicate the model's limited sensitivity to negative sentiment, consistent with the dataset's class imbalance.

```
Accuracy (SVM): 0.9606986899563319
Accuracy (SVM): 96.07%

Classification Report (SVM):
              precision    recall  f1-score   support

     Negatif       0.79      0.83      0.81        23
     Positif       0.98      0.98      0.98       206

    accuracy                           0.96       229
   macro avg       0.89      0.90      0.89       229
weighted avg       0.96      0.96      0.96       229
```

Figure 3. Classification Report of SVM

Figure 3 presents the detailed classification report. The SVM achieved an overall accuracy of 96.07%, signifying that approximately 96% of reviews were classified correctly. Performance metrics for each sentiment class are as follows:

- Positive Sentiment: Precision = 0.89, Recall = 0.98, F1-score = 0.98
- Negative Sentiment: Precision = 0.79, Recall = 0.83, F1-score = 0.81

The high precision and near-perfect recall for positive sentiment indicate that the classifier accurately identified the majority of positive reviews while minimizing false positives. The resulting F1-score of 0.98 confirms the model's reliability in recognizing positive expressions. Conversely, the lower precision (0.79) for negative sentiment reveals occasional misclassification of positive reviews as negative, while the recall of 0.83 shows moderate ability in retrieving actual negative instances. The corresponding F1-score of 0.81 suggests that while the model can detect negative sentiment to a reasonable extent, its predictions are less stable compared to the positive class.

The performance discrepancy between sentiment categories largely stems from the class imbalance, where the predominance of positive reviews (997) over negative ones (147) biases the classifier toward the majority class. This imbalance reduces the decision boundary's sensitivity to minority examples, leading to skewed classification outcomes.

4.7. Data Visualization and Interpretation

Data visualization techniques help clarify the model's classification results and highlight key word frequency patterns.

Figure 4. WordCloud Visualization

Figure 4 presents a WordCloud illustrating the most frequently appearing terms in the review corpus. The font size of each word corresponds to its frequency, visually emphasizing dominant tokens. Words such as "driver," "ramah," "cepat," "bagus," "aplikasi," and "murah" appear in larger fonts, indicating their frequent use in user reviews. These terms predominantly express positive sentiment, which aligns with the model's high accuracy in identifying positive classifications.

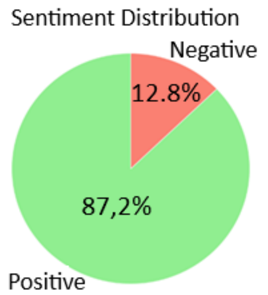

Figure 5. Sentiment Distribution Visualization

Figure 5 shows the sentiment distribution within the dataset, revealing a clear predominance of positive reviews over negative ones. This pattern confirms the class imbalance observed earlier and suggests that the model may be biased toward the majority class. The frequent occurrence of favorable expressions indicates that Maxim users often emphasize attributes such as driver friendliness, service efficiency, and affordability—factors that contribute to the high rate of positive feedback.

Overall, the visualizations offer deeper insight into the dataset's sentiment orientation and linguistic composition. They also underscore the importance of improving data balance or incorporating additional samples to enhance the SVM model's fairness and generalization performance. These insights provide a foundation for drawing meaningful conclusions regarding user sentiment and the effectiveness of the proposed hybrid sentiment analysis model.

5. CONCLUSION

The lexicon-based SVM model effectively classified 2,000 Maxim application reviews, achieving an accuracy of 96.07% and confirming its strong capability in sentiment detection. The model performed best in identifying positive sentiment, while its lower precision in detecting negative sentiment reflected the impact of class imbalance and the absence of neutral categories. WordCloud analysis revealed frequent terms such as "*driver*", "*cepat*", and "*murah*", indicating user satisfaction with service speed, affordability, and friendliness.

Future work should focus on balancing sentiment classes, expanding multilingual datasets, and integrating neutral sentiment labels to enhance model fairness and generalization. Additionally, exploring deep learning or ensemble-based techniques may further improve performance for large-scale and diverse sentiment datasets, strengthening applicability in real-world digital mobility platforms like Maxim.

ACKNOWLEDGMENT

The authors would like to express their gratitude to the academic staff of Information System Department, Universitas Gunadarma, for their valuable guidance and support during this research. Appreciation is also extended to colleagues and reviewers whose insights contributed to improving the quality of this study.

REFERENCES

[1] H. Delaere, S. Basu, C. Macharis, and I. Keseru, "Barriers and opportunities for developing, implementing and operating inclusive digital mobility services," *Eur. Transp. Res. Rev.*, vol. 16, no. 1, 2024, doi: 10.1186/s12544-024-00684-8.

[2] A. Z. Yonatan, "Indonesia Jadi Pengguna Transportasi Online Terbesar 2024," *GoodStats*. 2024, [Online]. Available: https://goodstats.id/article/indonesia-jadi-pengguna-transportasi-online-terbesar-2024-sn07c.

[3] Maxim Indonesia, "Maxim Indonesia (PT Teknologi Perdana Indonesia): Overview [LinkedIn page]," 2025. https://www.linkedin.com/company/maximid/?originalSubdomain=id (accessed Sep. 22, 2025).

[4] Maxim Indonesia, "maxim - ojek, transportasi - Apps on Google Play," 2025. https://play.google.com/store/apps/details?id=com.taxsee.taxsee (accessed Sep. 01, 2025).

[5] S. E. Safitri, W. D. Yuniarti, M. R. Handayani, and K. Umam, "User Opinion Mining on the Maxim Application Reviews Using BERT-Base Multilingual Uncased," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 14, no. 3, pp. 365–372, 2025, doi: 10.32736/sisfokom.v14i3.2391.

[6] V. H. Pranatawijaya, N. N. K. Sari, R. A. Rahman, E. Christian, and S. Geges, "Unveiling User Sentiment: Aspect-Based Analysis and Topic Modeling of Ride-Hailing and Google Play App Reviews," *J. Inf. Syst. Eng. Bus. Intell.*, vol. 10, no. 3, pp. 328–339, 2024, doi: 10.20473/jisebi.10.3.328-339.

[7] A. N. Alfarobby and H. Irawan, "Analisis

Sentimen Kepuasan Konsumen Pengguna Transportasi Online Pada Ulasan Google Playstore Menggunakan Indobert Dan Topic Modeling (Studi kasus: Gojek dan Grab)," *e-Proceeding Manag.*, vol. 11, no. 1, p. 72, 2024.

[8] H. M. U. Ali, Q. Farooq, A. Imran, and K. El Hindi, "A systematic literature review on sentiment analysis techniques, challenges, and future trends," *Knowl. Inf. Syst.*, vol. 67, no. 5, pp. 3967–4034, May 2025, doi: 10.1007/S10115-025-02365-X/METRICS.

[9] H. Firda, P. Putra, N. R. Oktadini, P. E. Sevtiyuni, and A. Meiriza, "Comparison of Rating-based and Inset Lexicon-based Labeling in Sentiment Analysis using SVM (Case Study: GoBiz Application Reviews on Google Play Store)," *Sistemasi*, vol. 14, no. 2, p. 516, 2025, doi: 10.32520/stmsi.v14i2.4795.

[10] P. K. Gautam and S. Shaw, "Sentiment Analysis Approaches: A Systematic Review," 2022, doi: 10.4108/eai.16-4-2022.2318164.

[11] S. Fransiska and A. Irham Gufroni, "Sentiment Analysis Provider by.U on Google Play Store Reviews with TF-IDF and Support Vector Machine (SVM) Method," *Sci. J. Informatics*, vol. 7, no. 2, pp. 203–212, 2020, [Online]. Available: http://journal.unnes.ac.id/nju/index.php/sji.

[12] Y. Mao, Q. Liu, and Y. Zhang, "Sentiment analysis methods, applications, and challenges: A systematic literature review," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 36, no. 4, p. 102048, 2024, doi: 10.1016/j.jksuci.2024.102048.

[13] E. C. M. Torres and L. G. de Picado-Santos, "Sentiment Analysis and Topic Modeling in Transportation: A Literature Review," *Appl. Sci.*, vol. 15, no. 12, 2025, doi: 10.3390/app15126576.

[14] H. Harnelia, "Analisis Sentimen Review Skincare Skintific Dengan Algoritma Support Vector Machine (Svm)," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 2, 2024, doi: 10.23960/jitet.v12i2.4095.

[15] A. Sitanggang, Y. Umaidah, Y. Umaidah, R. I. Adam, and R. I. Adam, "Analisis Sentimen Masyarakat Terhadap Program Makan Siang Gratis Pada Media Sosial X Menggunakan Algoritma Naïve Bayes," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3, 2024, doi: 10.23960/jitet.v12i3.4902.

[16] A. Nurian, "Analisis Sentimen Ulasan Pengguna Aplikasi Google Play Menggunakan Naïve Bayes," *J. Inform. dan Tek. Elektro Terap.*, vol. 11, no. 3s1, pp. 829–835, 2023, doi: 10.23960/jitet.v11i3s1.3348.

[17] I. Cero, J. Luo, and J. M. Falligant, "Lexicon-Based Sentiment Analysis in Behavioral Research," *Perspect. Behav. Sci.*, vol. 47, no. 1, p. 283, Mar. 2024, doi: 10.1007/S40614-023-00394-X.

[18] A. Rufaida, A. Permanasari, and N. Setiawan, "Lexicon-Based Sentiment Analysis Using Inset Dictionary: A Systematic Literature Review," 2023, doi: 10.4108/eai.5-10-2022.2327474.

[19] K. Alemerien, A. Al-Ghareeb, and M. Z. Alksasbeh, "Sentiment Analysis of Online Reviews: A Machine Learning Based Approach with TF-IDF Vectorization," *J. Mob. Multimed.*, vol. 20, no. 05, pp. 1089–1116, 2024, doi: 10.13052/jmm1550-4646.2055.

[20] F. Koto and G. Y. Rahmaningtyas, "Inset lexicon: Evaluation of a word list for Indonesian sentiment analysis in microblogs," *Proc. 2017 Int. Conf. Asian Lang. Process. IALP 2017*, vol. 2018-Janua, pp. 391–394, 2017, doi: 10.1109/IALP.2017.8300625.

[21] Z. Zhan, "Comparative Analysis of TF-IDF and Word2Vec in Sentiment Analysis: A Case of Food Reviews," *ITM Web Conf.*, vol. 70, p. 02013, 2025, doi: 10.1051/itmconf/20257002013.

[22] T. Ahmed Khan, R. Sadiq, Z. Shahid, M. M. Alam, and M. Mohd Su'ud, "Sentiment Analysis using Support Vector Machine and Random Forest," *J. Informatics Web Eng.*, vol. 3, no. 1, pp. 67–75, 2024, doi: 10.33093/jiwe.2024.3.1.5.

[23] N. W. S. Saraswati, I. K. G. D. Putra, M. Sudarma, and I. M. Sukarsa, "Enhance sentiment analysis in big data tourism using hybrid lexicon and active learning support vector machine," *Bull. Electr. Eng. Informatics*, vol. 13, no. 5, pp. 3663–3674, 2024, doi: 10.11591/eei.v13i5.7807.

[24] G. Jeffson Sagala and Y. T. Samuel, "Sentiment Analysis on ChatGPT App Reviews on Google Play Store Using Random Forest Algorithm, Support Vector Machine and Naïve Bayes," *Int. J. Eng. Bus. Soc. Sci.*, vol. 2, no. 04, pp. 1194–1204, 2024, doi: 10.58451/ijebss.v2i04.148.

[25] A. Fatihin, "Analisis Sentimen Terhadap Ulasan Aplikasi Mobile Menggunakan Metode Support Vector Machine (Svm) Dan Pendekatan Lexicon Based," 2021, [Online]. Available: https://repository.uinjkt.ac.id/dspace/handle/123456789/65009.

[26] P. Kurniawati, R. Y. Fa'rifah, and D. Witarsyah, "Sentiment Analysis of Maxim Online Transportation App Reviews using Support Vector Machine (SVM) Algorithm," *Build. Informatics, Technol. Sci.*, vol. 5, no. 2, 2023, doi: 10.47065/BITS.V5I2.4265.

[27] A. N. A. Saputra, R. E. Saputro, and D. I. S. Saputra, "Enhancing Sentiment Analysis Accuracy Using SVM and Slang Word Normalization on YouTube Comments," *Sinkron*, vol. 9, no. 2, pp. 687–699, 2025, doi: 10.33395/sinkron.v9i2.14613.

[28] A. Ligthart, C. Catal, and B. Tekinerdogan, *Systematic reviews in sentiment analysis: a tertiary study*, vol. 54, no. 7. Springer Netherlands, 2021.

[29] T. Fardiansyah, Z. Yunizar, and Maryana, "Implementation of Support Vector Machine Method with TF-IDF for Sentiment Analysis of the Al-Zaytun Islamic Boarding School Controversy," *Int. J. Eng. Sci. Inf. Technol.*, vol. 5, no. 3, pp. 58–65, 2025, doi: 10.52088/ijesty.v5i3.883.