SMATIKA : STIKI Informatika Jurnal
Accredited Sinta 4 Ministry of Higher Education, Science and Technology Republic of Indonesia SK No. 79/E/KPT/2023
E-ISSN: 2580-6939 || P-ISSN: 2087-0256

smatika
STIKI Informatika Jurnal

Sentiment Analysis of User Reviews on the Bus Simulator Indonesia Application Using Machine Learning Approaches

Iswatun Maimanah [3], Rakhmad Maulidi [2], Dany Candra Febrianto [1*]
[1,2,3] Telkom University, Faculty of Informatics, Informatics, DI Panjaitan Street No. 128, Karangreja, Purwokerto Kidul, Purwokerto Sel. District, Banyumas Regency, Central Java 53147, Indonesia

Keyword

Bus Simulator Indonesia; Logistic Regression; Naïve Bayes; Sentiment Analysis; Support Vector Machine

****Corresponding Author:***
danycandra@telkomuniversity.ac.id

Abstract

User reviews on the Google Play Store serve as a primary source of information for assessing the quality and satisfaction levels of an application. As expectations for game quality and realism continue to rise, Bus Simulator Indonesia still faces a range of technical issues — including bugs, glitches, visual disturbances, and suboptimal controls — that directly impact user comfort and overall experience. This study aims to classify the sentiments expressed by Bus Simulator Indonesia users into three categories: positive, neutral, and negative. In addition, it compares the performance of three machine learning algorithms: Naive Bayes, Support Vector Machine (SVM), and Logistic Regression. The research pipeline consists of data collection, text preprocessing, word weighting, modeling, and model evaluation. The dataset comprises 19,482 Indonesian-language reviews gathered from the Google Play Store between 2021 and 2025, evaluated across three data-splitting scenarios: 90:10, 80:20, and 70:30. The results show that Support Vector Machine consistently achieved the highest accuracy across all scenarios, recording 84.3%, 82.0%, and 81.0%, with weighted F1-score values of 0.83, 0.82, and 0.81, respectively, indicating a more stable classification performance. Logistic Regression followed with accuracies of 81.1%, 79.9%, and 79.0%, with weighted F1-score values of 0.79, 0.78, and 0.77, respectively, while Naive Bayes produced the lowest scores overall. These findings confirm that Support Vector Machine is the most effective algorithm for sentiment classification of Bus Simulator Indonesia user reviews.

1. Introduction

User reviews on the Google Play Store platform serve as an important source of information for assessing application quality, user satisfaction levels, and perceptions of the features and services offered. Reviews directly represent users' opinions, experiences, and responses to their use of an application, making them strategically valuable for application evaluation and development. The analysis of digital reviews has also been widely utilized in sentiment analysis research to systematically identify trends in user opinion [1].

Bus Simulator Indonesia (BUSSID) is a simulation game developed by a local Indonesian developer that demonstrates internationally competitive quality. The game is designed to deliver a bus-driving experience that closely resembles real conditions in Indonesia spanning road environments, vehicles, and driving culture and is often referred to as a locally made game with an international feel [2]. BUSSID's strengths lie in its distinctive features, including livery design, local street atmospheres, authentic Indonesian bus routes and operators, the iconic *telolet* horn, vehicle mods, multiplayer convoy mode, and automatic save functionality.

Beyond entertainment, BUSSID also incorporates elements of local wisdom that reflect the social and transportation landscape of Indonesia [3]. Nevertheless, numerous user reviews indicate persistent technical issues such as bugs and glitches, visual disturbances, and vehicle controls deemed suboptimal. Furthermore, excessive advertisement frequency has emerged as a primary complaint, contributing to a decline in user comfort and satisfaction [4].

The Google Play Store acts as the main platform where users provide ratings, share their experiences, and express complaints regarding the applications they use. The sheer volume and variety of available reviews pose significant challenges to manual analysis [5]. Consequently, an automated approach is needed to process and extract meaningful insights from user reviews. One such approach is sentiment analysis a technique for classifying user opinions into positive, negative, and neutral categories based on review text enabling a more structured understanding of user perceptions toward a given application [1].

A number of prior studies have demonstrated that machine learning algorithms such as Naïve Bayes and SVM perform well in sentiment classification tasks. Research on Bus Simulator Indonesia review sentiment analysis using TF-IDF feature representation and the Naïve Bayes algorithm yielded a considerably high accuracy rate [4]. Another study comparing Naïve Bayes, SVM, and k-NN on gadget reviews concluded that SVM delivered the best overall performance [6]. Similar findings were reported in sentiment analyses of M-Paspor application reviews [7] and Twitter data [8], where SVM consistently achieved higher accuracy than Naïve Bayes. Additionally, research on Gojek application reviews from the Google Play Store found that Logistic Regression outperformed Multinomial NB, SVM, and k-NN [9]. In the e-commerce domain, sentiment analysis of Lazada reviews similarly demonstrated that SVM outperforms Naïve Bayes [10],[11],[12].

This research applies three classification algorithms to determine which is best suited for sentiment classification: SVM, Logistic Regression, and Naïve Bayes. SVM is a machine learning method that aims to identify the optimal hyperplane for separating data between classes [13]. Logistic Regression offers a notable advantage in terms of model interpretability, allowing the relationship between features and class probabilities to be understood more clearly [14]. Meanwhile, Naïve Bayes is a probabilistic classification algorithm that leverages Bayes' Theorem to determine the most likely class based on a given set of attributes [15].

Research on sentiment analysis has been widely applied to understand public opinion on various issues and digital services, spanning both social media and application reviews. A number of studies demonstrate that machine learning approaches — particularly Naïve Bayes, Support Vector Machine, and Logistic Regression — are effective in classifying user sentiment. A study examining public sentiment toward the development of Indonesia's new capital city, Nusantara (IKN), using data from social media platform X, found that SVM achieved an accuracy of 80%, slightly outperforming both NB and Logistic Regression, which each reached 79% [16]. These findings affirm SVM's advantage in modeling text-based public opinion on policy issues.

In the context of application reviews, research on the Gojek application on the Google Play Store found that Logistic Regression delivered the best performance with an accuracy of 82.45%, surpassing Multinomial NB, SVM, and k-Nearest Neighbor (k-NN) [9]. Another study on aspect-based application reviews using TF-IDF feature representation found that SVM with a Radial Basis Function (RBF) kernel achieved the highest accuracy of 92.86%, marginally outperforming NB [17]. Sentiment analysis of PeduliLindungi application user reviews based on Twitter data showed that SVM attained an accuracy of 86%, while NB reached 85%, with NB holding an advantage in computational efficiency [18]. Further, research on TikTok application user reviews in Indonesia demonstrated that SVM consistently outperformed NB across accuracy, precision, recall, F1-Score, and AUC metrics [19].

Despite its limitations, several studies indicate that NB maintains notable strengths under particular circumstances. In a sentiment analysis conducted on TikTok Shop application reviews, SVM demonstrated superior consistency in classification, while NB showed greater effectiveness in identifying negative sentiment. [20]. Similar findings were obtained in a Twitter-based sentiment analysis of the PeduliLindungi application, where both algorithms exhibited relatively comparable performance, despite SVM achieving a marginally higher accuracy [21]. In the public service domain, sentiment analysis of Information and Communication Technology (ICT) service satisfaction indicated that SVM demonstrated superior classification performance over NB [22]. Conversely, in highly polarized social issues such as the boycott of Israeli products on Twitter, NB achieved higher accuracy than SVM [23].

Table 1. Summary of Previous Sentiment Analysis Studies

Ref.	Dataset / Research Object	Algorithms Used	Language	Reported Accuracy / Findings
[4]	BUSSID user reviews from Google Play Store	TF-IDF, Multinomial Naïve Bayes	Indonesian	Highest accuracy of 85% using 90:10 train-test split
[6]	Samsung Galaxy Z Flip 3 YouTube comments	Naïve Bayes, SVM, k-NN	Indonesian	SVM achieved the best average accuracy of 96.43%
[7]	M-Paspor application reviews	Naïve Bayes, SVM	Indonesian	SVM achieved 80.76%, outperforming NB with 78.12%
[9]	Gojek application reviews	Logistic Regression, Multinomial NB, SVM, k-NN	Indonesian	Logistic Regression achieved the highest accuracy of 82.45%
[16]	Public sentiment on IKN development from social media X	SVM, Logistic Regression, Naïve Bayes	Indonesian	SVM achieved the highest accuracy of 80%
[17]	Application user reviews	SVM, Naïve Bayes Classifier	Indonesian	SVM achieved 92.86%, slightly higher than NB with 92.54%
[20]	TikTok Shop user reviews	SVM, Naïve Bayes, TF-IDF	Indonesian	SVM outperformed NB with 68.86% accuracy, while NB excelled in negative sentiment detection

Based on Table 1, previous studies indicate that Support Vector Machine frequently achieves superior performance in sentiment classification tasks, particularly in application review datasets. However, Naïve Bayes and Logistic Regression also demonstrate competitive performance under certain dataset characteristics and sentiment distributions. Most previous studies focused on e-commerce platforms, public services, and social media discussions, while sentiment analysis studies on Indonesian local game applications such as Bus Simulator Indonesia are still limited. Therefore, this study compares the performance of Naïve Bayes, Support Vector Machine, and Logistic Regression using TF-IDF feature representation on BUSSID user reviews from the Google Play Store.

This study aims to analyze the sentiment of user reviews for the Bus Simulator Indonesia application on the Google Play Store using Naïve Bayes, SVM, and Logistic Regression algorithms with TF-IDF feature representation, as well as to compare the performance of the three algorithms based on evaluation metrics. The findings are expected to contribute academically to the advancement of Indonesian-language sentiment analysis research particularly in the domain of local gaming applications while also providing practical benefits for BUSSID developers in systematically understanding user perceptions as a basis for informed decision-making in application development and quality improvement.

2. Research Method

The research stages are structured as shown in Figure 1, encompassing the processes of data collection, preprocessing, feature extraction, sentiment labeling, classification modeling, model evaluation, and result visualization.

The data analyzed consists of mobile application user reviews obtained from the Google Play Store. User reviews were selected as the data source because they are capable of directly representing users' satisfaction levels, complaints, and expectations regarding application quality and features [17]. The research object is focused on the BUSSID application a simulation game developed by a local Indonesian developer that delivers a realistic bus-driving experience with a distinctly Indonesian feel [4].

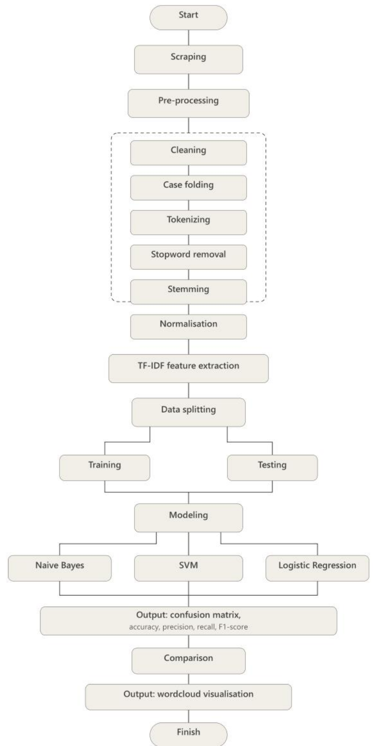


Figure 1. Research flowchart.

2.1 Data Collection

Data collection was carried out using a web scraping technique, a method of automatically retrieving data from web pages using a specific program [24]. The data collected consisted of user review texts from the BUSSID application on the Google Play Store over a five-year period (2021–2025). The reviews retrieved were limited to those written in Indonesian and relevant to the application's main features, such as livery, graphics, the telolet horn, convoy mode, vehicle mods, the save system, and advertisements. Sample size determination was conducted using a percentage-based approach from the available data population. Several studies recommend retrieving approximately 10–20% of the reachable population to ensure the sample remains representative [25], [26].

2.2 Data Processing

To ensure the quality of the text data before analysis, the collected user reviews were processed through a sequential preprocessing pipeline. This pipeline was constructed in accordance with approaches used in preceding studies. [19], [27], [28], The pipeline comprised several stages, including cleaning, case folding, tokenization, stopword removal, stemming, and normalization. The cleaning stage was carried out to eliminate irrelevant characters such as symbols, numbers, and punctuation marks, whereas case folding was employed to convert all text into lowercase for standardization purposes. Tokenization was used to split text into individual words, stopword removal was conducted using the Indonesian stopword list provided by the NLTK library, which was further customized manually by adding informal and commonly used words that frequently appeared in the review dataset but did not contribute significantly to sentiment analysis, stemming was performed using the PySastrawi library to reduce affixed words to their base forms. The normalization process was carried out using a pre-existing Indonesian slang dictionary that was further adjusted manually based on non-standard words found in the dataset to convert non-standard words into their standard equivalents and maintain data consistency. The preprocessed data was then represented in numerical form using the Term Frequency–Inverse Document Frequency (TF-IDF) method to measure the importance of each word within a document [29], and to highlight opinion-bearing words in the sentiment analysis [19].

The dataset was subsequently divided into training and testing sets using simple random sampling with split ratios of 90:10, 80:20, and 70:30 to examine the effect of data proportion on classification model performance [30],[31]. Sentiment labeling was performed using a lexicon-based approach with the VADER Lexicon to classify reviews into three sentiment categories: positive, neutral, and negative [32]. The classification modeling process was carried out using the Naïve Bayes, Support Vector Machine (SVM), and Logistic Regression algorithms, which were selected for their ability to learn patterns from textual data and perform predictions effectively [33],[34],[35]. In this study, the models were implemented using the sklearn library, with most hyperparameters set to their default values. The Multinomial Naive Bayes model used an alpha parameter of 1.0, the LinearSVC model used C = 1.0, while the Logistic Regression model used C = 1.0 and max_iter = 1000. Model performance evaluation was conducted using accuracy, precision, recall, and F1-score metrics, along with a confusion matrix to analyze classification errors [36]. The analysis results were then visualized using a wordcloud to display the dominant words frequently appearing in the reviews, thereby providing a general overview of users' opinion focus toward the BUSSID application.

3. Result and Discussions

This section presents the results of the sentiment analysis research on user reviews of the Bus Simulator Indonesia (BUSSID) application, obtained through the stages of the research methodology. Each stage is presented along with visualization outputs in the form of figures and tables, and is discussed concisely and systematically in accordance with the research pipeline.

3.1 Data Collection

Data collection was conducted through a web scraping technique applied to user reviews of the BUSSID application on the Google Play Store covering the period 2021–2025. The results of the data collection process are shown in Figure 2, which displays the complete set of review data obtained through scraping.

	reviewId	userName	userImage	content	score	thumbsUpCount	reviewCreatedVersion	at	replyContent	repliedAt	appVersion
0	1b109463-be6b-44f5-965d-58eeb6e2dd77	avil ramadani	https://play-lh.googleusercontent.com/a/ACg8oc...	tolong di perbaiki bug nya pasang mod tidak bisa	5	0	4.4.1	2025-11-28 00:22:06	NaN	NaN	4.4.1
1	5a4bafa0-ff3c-4149-b0a8-d31711969e45	Daffa Irmansyah	https://play-lh.googleusercontent.com/a/ACg8oc...	bang tolong tambahin bus sddnya sama tambain k...	5	0	2.6	2025-11-27 17:16:32	NaN	NaN	2.6
2	e63ec1ff-b277-4fcf-b6c1-18096554f13f	Fikri Eka Wicaksono	https://play-lh.googleusercontent.com/a/ACg8oc...	tolong maleo kasih fitur untuk mabar pake mod ...	5	0	4.4.1	2025-11-27 16:34:15	NaN	NaN	4.4.1
3	f4b5f5ce-844a-494f-940f-e62a1168939d	Alfan Dm	https://play-lh.googleusercontent.com/a/ACg8oc...	memang bagus dan grafik nya juga keren cuman y...	1	0	NaN	2025-11-27 16:09:18	NaN	NaN	NaN
4	c579f605-685f-48b7-ab7b-2d2c1322b259	Aku uwong Saya orang	https://play-lh.googleusercontent.com/a/ACg8oc...	tolong lah hapus tuh iklan judol	1	0	4.4.1	2025-11-27 15:34:19	NaN	NaN	4.4.1
...	...	...	...	...	...	...	...	...	...	...	...
97289	e5fb3857-dd92-4d7e-9e9e-2253cf90ffdb	Ahmad Dani	https://play-lh.googleusercontent.com/a-/ALV-U...	game nya bisa mod apa aja	5	0	NaN	2021-01-01 01:18:25	NaN	NaN	NaN
97290	24bd5ea6-b6c2-4623-81f6-281d9761bdd1	Imam Jakfar	https://play-lh.googleusercontent.com/a/ACg8oc...	game nya udah bagu ,min tolong turunin dong mb...	5	1	3.4.3	2021-01-01 00:37:58	NaN	NaN	3.4.3
97291	952e1ec9-11b2-46b1-964e-c6d54c7834db	rahman official	https://play-lh.googleusercontent.com/a-/ALV-U...	mohon bussid tolong kalo narik kita dari garas...	5	0	3.3.2	2021-01-01 00:19:07	NaN	NaN	3.3.2
97292	7b0a5c8f-2e2d-4249-b6a0-1cfcbe174151	Syafina & Kenzie	https://play-lh.googleusercontent.com/a-/ALV-U...	gem ya, bisapake mod truck new bekaje biar tam...	5	0	3.4.3	2021-01-01 00:18:10	NaN	NaN	3.4.3
97293	fadf0fec-0826-4fb4-8655-32e694296b47	Iko Prastiyo	https://play-lh.googleusercontent.com/a-/ALV-U...	masih kalah jauh dengan euro truck simulator 2...	1	0	3.3.3	2021-01-01 00:16:22	NaN	NaN	3.3.3

97294 rows × 11 columns

Figure 2 Sample review data

Once the population was obtained, a sample of 20% of the total population was drawn using the simple random sampling method to reduce bias and maintain data representativeness [36], [42].

3.2 Pre-Processing

The preprocessing stage was conducted to ensure that the text data was in optimal condition prior to analysis. This process encompasses cleaning to remove unnecessary symbols and characters, case folding to standardize all text to lowercase, tokenizing to split text into individual words, stopword removal to eliminate common words, stemming to reduce words to their base forms, and normalization to convert non-standard words into their standard equivalents.

	at	normalisasiBaru
0	2021-01-01 00:18:10	gim pakai mod truk baru bekas biar keren
1	2021-01-01 00:19:07	mohon bussid tarik garasi terminal tambah kota...
2	2021-01-01 00:37:58	gim bagus admin turun buat mod knalpot serigal...
3	2021-01-01 03:22:15	abang baru tumpang terminal rute tol abang
4	2021-01-01 03:32:00	semoga baru baru masuk kota medan

Figure 3. Pre-processing result.

3.3 Data Labeling

Sentiment labeling was performed using a lexicon-based approach with the VADER Lexicon to classify reviews into positive, negative, and neutral sentiments [37].As shown in Figure 4, negative sentiment is dominant at 37.10%, followed by positive sentiment at 32.84%, and neutral sentiment at 30.06%. This distribution indicates a reasonably balanced variation in user opinions. Feature extraction was then carried out using the TF-IDF method with unigram and bigram representations.

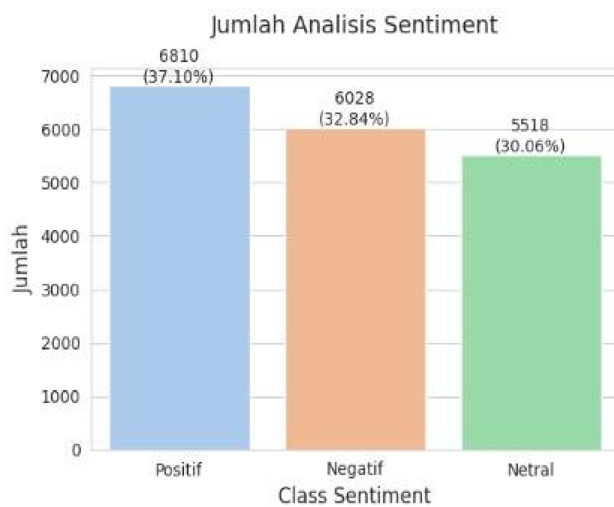


Figure 4. Label distribution

3.4 Data Splitting

The feature-extracted data was split using three ratio schemes: 90:10, 80:20, and 70:30. The data splitting results are presented in Table 2. The use of multiple ratios aims to evaluate the effect of training data proportion on model performance [47], [48].

Table 2. Data splitting

Ratio	Train Data	Test Data
90:10	16,520	1,836
80:20	14,684	3,672
70:30	12,849	5,507

3.5 Model Evaluation Results

Model performance evaluation was conducted to measure the ability of the Naïve Bayes, SVM, and Logistic Regression algorithms in classifying the sentiment of BUSSID user reviews. The evaluation was performed using three data-splitting schemes — 90:10, 80:20, and 70:30 — with accuracy and confusion matrix as the primary metrics.

Table 3. Model Accuracy and Confusion matrix

Algoritma	Split Data	Accuracy	Precision Macro	Recall Macro	F1-Score Macro	Precision Weighted	Recall Weighted	F1-Score Weighted
Naïve Bayes	90:10	65.95%	0.71	0.63	0.61	0.70	0.66	0.62
Naïve Bayes	80:20	64.46%	0.70	0.62	0.59	0.69	0.64	0.60
Naïve Bayes	70:30	64.45%	0.70	0.62	0.59	0.69	0.64	0.60
SVM	90:10	84.31%	0.84	0.84	0.83	0.84	0.84	0.83
SVM	80:20	82.03%	0.81	0.81	0.81	0.82	0.82	0.82
SVM	70:30	80.95%	0.80	0.80	0.80	0.81	0.81	0.81
Logistic Regression	90:10	81.10%	0.81	0.81	0.81	0.81	0.81	0.81
Logistic Regression	80:20	79.92%	0.79	0.79	0.79	0.79	0.79	0.79
Logistic Regression	70:30	79.01%	0.78	0.78	0.78	0.78	0.78	0.78

To provide an overview of model performance per sentiment class, a confusion matrix summary showing the number of correct predictions (TP) per class is presented in Table 4.

Table 4. Confusion Matrix

Split Data	Model	TP Total	FP Total	FN Total	TN Total	True Neutral	False Neutral
90:10	Naïve Bayes	1211	625	625	3047	124	428
80:20	SVM	1548	288	288	3384	383	169
70:30	Logistic Regression	1489	347	347	3325	381	171
90:10	Naïve Bayes	2368	1304	1304	6040	226	878
80:20	SVM	3012	660	660	6684	737	367
70:30	Logistic Regression	2935	737	737	6607	752	352
90:10	Naïve Bayes	3549	1958	1958	9056	325	1330
80:20	SVM	4458	1049	1049	9965	1071	584
70:30	Logistic Regression	4351	1156	1156	9858	1084	571

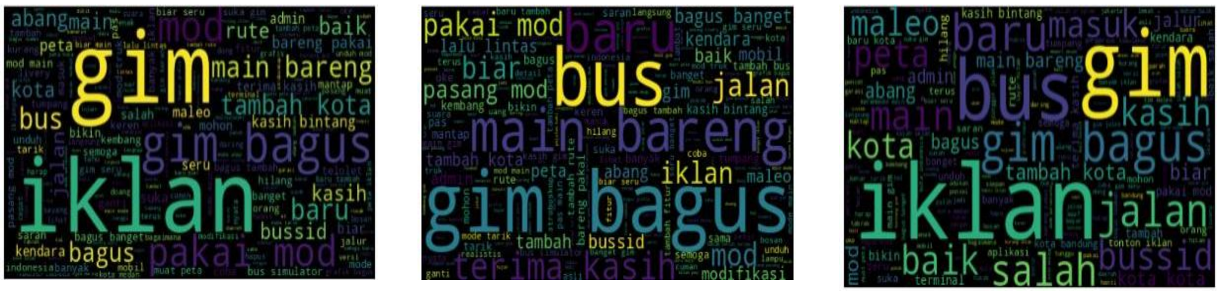
Based on Figure 5 the comparison results show that SVM consistently achieved the highest accuracy across all scenarios.

Figure 5. Model Accuracy Comparison

Based on Figures 6, the word cloud visualizations show differences in the most frequently occurring words across each sentiment category. In the neutral sentiment, the dominant words include "*gim*", "*bagus*", "*iklan*", "*tambah kota*", and "*baru*", which indicates that users provided comments in the form of suggestions, requests for new features, and feedback related to game development. The positive sentiment is dominated by words such as "*bagus*", "*gim*", "*bus*", "*main*", and "*bareng*", which indicates that users are satisfied with the gameplay quality, multiplayer features, and the engaging as well as entertaining gaming experience. Meanwhile, the negative sentiment features words like "*iklan*", "*gim*", "*bus*", "*jalan*", "*salah*", and "*kota*", which indicates that user complaints are largely related to excessive advertisement frequency, bugs on maps or travel routes, and the incompatibility of certain in-game features. The occurrence of the word "advertisement" in both neutral and negative sentiments suggests that the advertising system has become a dominant issue directly affecting user experience. These findings provide insights that the BUSSID developers need to optimize advertisement placement, improve the stability of travel-related features, and fix bugs as well as game performance issues in order to enhance user satisfaction more effectively.

The results of the study indicate that the preprocessing stage and TF-IDF feature extraction play a crucial role in improving the quality of textual data, and the SVM algorithm demonstrates the most optimal and balanced performance across all sentiment classes. This finding differs from research on Gojek application reviews, where Logistic Regression outperformed SVM, Multinomial NB, and k-NN with an accuracy of 82.45% [9]. The divergence is plausible given differences in dataset scale and sentiment distribution: the present study draws on a substantially larger BUSSID review corpus spanning five years, which may allow SVM's margin-based decision boundary to separate sentiment classes more effectively than a linear probability model once the feature space grows denser. The result instead aligns more closely with studies reporting SVM as the top performer among the same three algorithms, such as sentiment analysis of public opinion on the IKN capital relocation, where SVM reached 80% accuracy against 79% for both Logistic Regression and Naïve Bayes under comparable conditions [16], and an evaluation of TikTok application reviews, where SVM consistently outperformed Naïve Bayes across accuracy, precision, recall, and F1-score [19].

However, SVM's advantage is not unconditional across the sentiment analysis literature, and the exceptions are informative rather than contradictory. In a sentiment analysis of TikTok Shop reviews, SVM achieved higher overall accuracy, yet Naïve Bayes proved more effective specifically at identifying negative sentiment [20], a pattern that resonates with the present study's own finding that the neutral class consistently yielded lower recall and F1-score across all three algorithms, as shown in Table 3. This suggests that algorithm advantage is not uniform across sentiment classes, and that SVM's overall superiority can still coexist with localized weaknesses on harder-to-separate categories. A more striking exception comes from a study on the boycott of Israeli products on Twitter, where Naïve Bayes achieved higher accuracy than SVM [23]. That dataset, however, centers on a sharply polarized and emotionally charged social and political issue, conditions known to produce highly distinctive, repetitive lexical markers that align well with Naïve Bayes's independence assumption. BUSSID reviews, by contrast, concern a consumer application rather than a polarized public issue, and the vocabulary used by reviewers is comparatively varied and context-dependent, spanning gameplay mechanics, technical bugs, and advertisement complaints. This difference in domain character likely explains why SVM, which is better suited to capturing more nuanced and overlapping feature boundaries, retained its advantage in the present study despite Naïve Bayes performing competitively or better in more polarized or topically narrow datasets elsewhere. The uniqueness of this research lies in the use of application review data from a simulation game based on the local Indonesian context, as well as the combination of unigram and bigram TF-IDF, which is able to represent variations in user opinions more comprehensively.



(a) (b) (c)

Figure 6. (a) Word Cloud Neutral Sentiment, (b) Word Cloud Positive Sentiment, (c)Word Cloud Negative Sentiment

4. Conclusions

This study aims to analyse the sentiment of user reviews for the Bus Simulator Indonesia (BUSSID) application on the Google Play Store, as well as to compare the performance of the Naïve Bayes, SVM, and Logistic Regression algorithms in classifying review sentiment into three categories: positive, negative, and neutral. Based on the analysis of 19,482 user reviews sampled from a total of 97,294 records, the sentiment distribution was found to be 37.10% negative, 32.84% positive, and 30.06% neutral, indicating that negative sentiment still dominates user reviews, although the proportions of positive and neutral sentiment are relatively balanced.

The model evaluation results indicate that the SVM algorithm achieved the best performance across all data split scenarios. In the 90:10 data split, SVM obtained an accuracy of 84.31% with a weighted F1-score of 0.83. In the 80:20 and 70:30 data splits, SVM also achieved accuracies of 82.03% and 80.95%, respectively. Logistic Regression ranked second with stable performance, while Naïve Bayes showed lower performance, particularly in the neutral class, which had lower recall and F1-score values compared to the other algorithms. These findings confirm that SVM is better capable of capturing complex sentiment patterns in BUSSID review data compared to the other two algorithms. Accordingly, the research objectives of conducting sentiment analysis and identifying the best-performing classification algorithm have been achieved, and the results of this study can serve as a data-driven evaluation basis for BUSSID application developers in improving application quality and user experience.

This study still has several limitations, particularly regarding the use of the VADER Lexicon for Indonesian-language texts, especially in understanding informal language contexts, abbreviations, slang, and local expression variations used by application users. In addition, this study only compared traditional machine learning algorithms, namely Naïve Bayes, SVM, and Logistic Regression, without involving deep learning methods such as BERT or IndoBERT, which have stronger capabilities in understanding complex language contexts. Therefore, future studies are recommended to employ deep learning–based models such as IndoBERT or LSTM to improve sentiment classification performance, as well as to develop aspect-based sentiment analysis to analyze user opinions on specific BUSSID application features, such as graphics, gameplay, performance, and application bugs, in a more detailed manner. The use of larger datasets and more optimized Indonesian-language preprocessing techniques may also support improvements in the quality of sentiment analysis results in future research.

5. References

[1] A. Bahar, T. Astuti, and P. Arsi, "Performance comparison of SVM, Naive Bayes, and support vector machine classification algorithms in analyzing Noice app user reviews," *J. Tek. Inform.*, vol. 5, no. 4, pp. 469–477, 2024, doi: 10.52436/1.jutif.2024.5.4.2061.

[2] S. Z. Gunawan, "Bus Simulator Indonesia (BUSSID), game buatan lokal rasa internasional," 2021.

[3] M. F. Ravizaldi, N. Z. Saputra, S. Syabani, and N. Suciana, "Analisis kearifan lokal pada Bus Simulator Indonesia," in *Prosiding Seminar Nasional Desain dan Media*, Nov. 2023, pp. 750–755.

[4] T. I. Alfawas, A. Rahim, and R. Rudiman, "Application of TF-IDF feature extraction for sentiment analysis of Bus Simulator Indonesia game reviews using the Naive Bayes algorithm," *J. Soc. Sci. Res.*, vol. 4, no. 5, pp. 3177–3193, 2024.

[5] S. Imron, E. I. Setiawan, J. Santoso, and M. H. Purnomo, "Aspect based sentiment analysis marketplace product reviews using BERT, LSTM, and CNN," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 7, no. 3, pp. 586–591, 2023, doi: 10.29207/resti.v7i3.4751.

[6] J. W. Iskandar and Y. Nataliani, "Comparison of Naïve Bayes, SVM, and k-NN for aspect-based gadget sentiment analysis," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 6, pp. 1120–1126, Dec. 2021, doi: 10.29207/resti.v5i6.3588.

[7] R. Maheri, F. N. Salisah, F. Muttakin, and M. Megawati, "Sentiment analysis of M-Paspor application reviews using Naive Bayes and Support Vector Machine," *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 10, no. 1, pp. 448–458, 2025, doi: 10.29100/jipi.v10i1.5826.

[8] S. Styawati, A. R. Isnain, N. Hendrastuty, and L. Andraini, "Comparison of Support Vector Machine and Naïve Bayes on Twitter data sentiment analysis," *J. Inform. J. Pengemb. IT*, vol. 6, no. 1, pp. 56–60, 2021, doi: 10.30591/jpit.v6i1.3245.

[9] A. Maulana, I. K. Afifah, A. Mubarrak, K. R. Fauzan, A. Dwintara, and B. P. Zen, "Comparison of Logistic Regression, Multinomial NB, SVM, and k-NN methods on sentiment analysis of Gojek app reviews on the Google Play Store," *J. Tek. Inform.*, vol. 4, no. 6, pp. 1487–1494, Dec. 2023, doi: 10.52436/1.jutif.2023.4.6.863.

[10] S. Setyabudi and E. Aryanny, "Sentiment analysis of Lazada marketplace user ratings with Naïve Bayes and Support Vector Machine methods," *J. Inovtek Polbeng - Seri Inform.*, vol. 10, no. 1, pp. 422–433, Mar. 2025, doi: 10.35314/sww8cg21.

[11] T. S. Ningsih, T. I. Hermanto, and I. M. Nugroho, "Sentiment analysis of mobile provider application reviews using Naive Bayes algorithm and Support Vector Machine," *SinkrOn*, vol. 8, no. 2, pp. 824–835, Mar. 2024, doi: 10.33395/sinkron.v8i2.13469.

[12] Y. Christian, T. Wibowo, and M. Lyawati, "Sentiment analysis by using Naïve Bayes classification and Support Vector Machine, study case Sea Bank," *SinkrOn*, vol. 8, no. 1, pp. 258–275, Jan. 2024, doi: 10.33395/sinkron.v9i1.13141.

[13] E. Suryati, Styawati, and A. A. Aldino, "Online transportation sentiment analysis using Word2Vec text embedding feature extraction and Support Vector Machine (SVM) algorithm," *J. Teknol. dan Sist. Inf.*, vol. 4, no. 1, pp. 96–106, 2023, doi: 10.1109/ISMODE53584.2022.9742906.

[14] M. Prasetya, M. Wulandari, and S. A. Nikmah, "Implementation of basic NLP (natural language processing) in Spotify review sentiment analysis," in *Seminar Nasional Teknologi \& Sains*, 2024, pp.

145–153.

[15] D. Novira and A. Lattu, "Public sentiment analysis on Instagram regarding the Jakarta-Bandung high-speed train using the Naïve Bayes classifier algorithm," in *Prosiding Seminar Nasional Sistem Informasi dan Manajemen Informatika Universitas Nusa Putra*, 2023, pp. 142–149.

[16] N. Hadi and D. Sugiarto, "Sentiment analysis of IKN development on social media X using SVM, Logistic Regression, and Naïve Bayes algorithms," *J. Inform. J. Pengemb. IT*, vol. 10, no. 1, pp. 37–49, 2025, doi: 10.30591/jpit.v10i1.7106.

[17] M. F. Madjid, D. E. Ratnawati, and B. Rahayudi, "Sentiment analysis on app reviews using Support Vector Machine and Naïve Bayes classification," *SinkrOn*, vol. 7, no. 1, pp. 556–562, Feb. 2023, doi: 10.33395/sinkron.v8i1.12161.

[18] R. Syahputra, G. J. Yanris, and D. Irmayani, "SVM and Naïve Bayes algorithm comparison for user sentiment analysis on Twitter," *SinkrOn*, vol. 6, no. 2, pp. 671–678, May 2022, doi: 10.33395/sinkron.v7i2.11430.

[19] J. O. Leandro and M. I. Fianty, "Evaluation of sentiment analysis methods for social media applications: A comparison of Support Vector Machines and Naïve Bayes," *JOIV Int. J. Informatics Vis.*, vol. 9, no. 2, 2025, doi: 10.62527/joiv.9.2.2905.

[20] A. M. Rizki, B. Bustami, and S. F. Anshari, "Comparison of Support Vector Machine and Naïve Bayes algorithms in sentiment analysis of TikTok Shop application user reviews," *J. Renew. Energy, Electr. Comput. Eng.*, vol. 5, no. 1, pp. 18–29, 2025, doi: 10.29103/jreece.v5i1.21342.

[21] I. Yunanto and S. Yulianto, "Twitter sentiment analysis PeduliLindungi application using Naïve Bayes and Support Vector Machine," *J. Tek. Inform.*, vol. 3, no. 4, pp. 807–814, Aug. 2022, doi: 10.20884/1.jutif.2022.3.4.292.

[22] W. Trisnawati and A. Wibowo, "Sentiment analysis of ICT service user using Naive Bayes classifier and SVM methods with TF-IDF text weighting," *J. Tek. Inform.*, vol. 5, no. 3, pp. 709–719, May 2024, doi: 10.52436/1.jutif.2024.5.3.1784.

[23] L. A. Hayurian and N. Hendrastuty, "Comparison of Naïve Bayes algorithm and Support Vector Machine in sentiment analysis of boycott Israeli products on Twitter," *J. Tek. Inform.*, vol. 5, no. 3, pp. 731–738, May 2024, doi: 10.52436/1.jutif.2024.5.3.1813.

[24] B. Zhao, "Web scraping," in *Encyclopedia of Big Data*, Cham: Springer International Publishing, 2022, pp. 951–953. doi: 10.1007/978-3-319-32010-6_483.

[25] A. B. Muyembe and I. P. Anselemo, "Scientific research sample size determination," *Int. J. Sci. Technol.*, vol. 11, no. 7, 2023.

[26] Y. K. Singh, *Fundamental of Research Methodology and Statistics*. New Age International, 2006.

[27] H. T. Duong and T. A. Nguyen-Thi, "A review: preprocessing techniques and data augmentation for sentiment analysis," *Comput. Soc. Networks*, vol. 8, no. 1, Dec. 2021, doi: 10.1186/s40649-020-00080-x.

[28] D. A. Tarigan, Z. Situmorang, and R. Rosnelly, "Sentiment analysis of the Play Store Sirekap 2024 application post-presidential election using a comparison of Support Vector Machine (SVM), Naïve Bayes classifier, and random forest methods," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 12, no. 3, pp. 661–670, 2025, doi: 10.25126/jtiik.2025129608.

[29] Z. Tang, W. Li, and Y. Li, "An improved supervised term weighting scheme for text representation and classification," *Expert Syst. Appl.*, vol. 189, p. 115985, 2022, doi: 10.1016/j.eswa.2021.115985.

[30] V. R. Joseph, "Optimal ratio for data splitting," Feb. 2022, doi: 10.1002/sam.11583.

[31] A. M. Asror and I. K. D. Nuryana, "Prediction and analysis of customer churn at Telkomsel using

machine learning approach," *J. Emerg. Inf. Syst. Bus. Intell.*, vol. 6, no. 2, pp. 230–240, Jul. 2025, doi: 10.26740/jeisbi.v6i2.65825.

[32] T. Gori, A. Sunyoto, and H. Al Fatta, "Data preprocessing and classification for predicting student academic performance," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 11, no. 1, pp. 215–224, 2024, doi: 10.25126/jtiik.20241118074.

[33] O. Peretz, M. Koren, and O. Koren, "Naive Bayes classifier - An ensemble procedure for recall and precision enrichment," *Eng. Appl. Artif. Intell.*, vol. 136, p. 108972, 2024, doi: 10.1016/j.engappai.2024.108972.

[34] T. Ke and others, "Maximal margin hyper-sphere SVM for binary pattern classification," *Eng. Appl. Artif. Intell.*, vol. 117, p. 105615, 2023, doi: 10.1016/j.engappai.2022.105615.

[35] K. Hasanah, "Comparison of sentiment analysis model for Shopee comments on Google Play Store," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 13, no. 1, pp. 21–30, 2024, doi: 10.32736/sisfokom.v13i1.1916.

[36] S. Szabó, I. J. Holb, V. É. Abriha-Molnár, G. Szatmári, S. K. Singh, and D. Abriha, "Classification Assessment Tool: A program to measure the uncertainty of classification models in terms of class-level metrics," *Appl. Soft Comput.*, vol. 155, p. 111468, 2024, doi: 10.1016/j.asoc.2024.111468.

[37] M. Gultom, J. Marikros, and W. Rusli, "Application of VADER sentiment for detecting English-based sentiment via website," in *Seminar Nasional Corisindo*, 2024, pp. 13–18.