Bulletin of Community Engagement
Vol. 6, No. 3, December 2026
https://attractivejournal.com/index.php/bce//index

A Comparative Sentiment Analysis of Learning Management Systems Reviews Using Machine Learning within the SEMMA Framework

Rahmat Rizki Hidayah [1*], Vianny Utami [1]
[1] *Bina Nusantara University, Indonesian*

rahmatrizkihidayah@gmail.com*

ARTICLE INFO
Article history:
Received
August 18, 2026
Revised
October 04, 2026
Accepted
November 08, 2026

Abstract

Mobile Learning Management Systems (LMS) have become essential in supporting digital learning by providing flexible access to academic resources and communication between students and educational institutions. User reviews published on the Google Play Store provide valuable feedback for evaluating service quality; however, their unstructured nature makes systematic analysis challenging. This study aims to analyze user sentiment toward the LMS Mobile application of Universitas XYZ and compare the performance of the Naïve Bayes and Support Vector Machine (SVM) algorithms in sentiment classification. A quantitative experimental approach was employed using 129 user reviews collected through web scraping from the Google Play Store between November 2020 and December 2025. The analysis followed the SEMMA (Sample, Explore, Modify, Model, Assess) framework and applied Natural Language Processing (NLP) techniques, including text cleaning, tokenization, case folding, stopword removal, and stemming. Textual data were transformed into numerical features using the Term Frequency–Inverse Document Frequency (TF-IDF) method before being classified with Naïve Bayes and SVM. Model performance was evaluated using confusion matrix metrics, including accuracy, precision, recall, and F1-score. The results indicate that negative sentiment dominates user reviews, primarily concerning login problems, application errors, and system performance. The Naïve Bayes model achieved a higher accuracy (0.69) than SVM (0.65), whereas SVM demonstrated more balanced classification performance across sentiment categories. These findings provide a data-driven basis for evaluating user feedback and supporting continuous improvement of digital learning services within higher education.

Keywords: Sentiment Analysis, Machine Learning, Learning Management System, Support Vector Machine, SEMMA.

Published by CV. Creative Tugu Pena
ISSN 2774-7077
Website https://attractivejournal.com/index.php/bce/
This is an open access article under the CC BY SA license
https://creativecommons.org/licenses/by-sa/4.0/

@ 2026 by the authors

INTRODUCTION

The The rapid development of digital technology has significantly transformed higher education, particularly through the adoption of online learning systems that provide flexible access to educational resources and academic activities. Mobile technology enables students to participate in learning anytime and anywhere, making digital learning an essential component of modern education. As the demand for technology-supported learning continues to grow, educational institutions increasingly rely on Learning Management Systems (LMS) to facilitate teaching, communication, assignment management, and learning assessment (Mechelen et al., 2023).

Learning Management Systems have become one of the primary platforms supporting distance and blended learning by integrating various academic services into a single digital environment. The effectiveness of an LMS, however, depends not only on its technical functionality but also on users' experiences when interacting with the system. Previous studies have reported that usability, system quality, feature availability, and user readiness strongly influence the

acceptance of digital learning technologies (Ikhsan., 2023; Rini et al., 2024). Therefore, continuous evaluation of user experiences is essential to improve the quality and sustainability of digital learning services (Aljouni., 2024).

Universitas XYZ has implemented the LMS Mobile application to support students in accessing learning materials, submitting assignments, viewing academic schedules, and conducting various academic activities through mobile devices. User experiences with the application are publicly reflected through reviews and ratings on the Google Play Store (Sarker., 2021). User-generated ratings and reviews represent an important source of feedback that can support software quality improvement and help identify application features required by users. As unstructured data, user reviews can also provide valuable knowledge for supporting organizational decision-making and application development (Mustaqim & Suryono, 2025). However, because user reviews are presented as unstructured textual data, manually identifying users' opinions and recurring issues becomes increasingly difficult as the volume of reviews grows. In this study, 129 user reviews collected from the Google Play Store between November 2020 and December 2025 indicate that one-star ratings dominate the dataset, with frequent complaints related to login failures, application errors, delayed notifications, system updates, and limited access to learning features (Kurniawan et al., 2024). These findings suggest that systematic analysis of user feedback is necessary to support evidence-based improvements in digital learning services.

Sentiment analysis has been widely applied to automatically classify user opinions from textual data using Natural Language Processing (NLP) and machine learning techniques. Algorithms such as Naïve Bayes and Support Vector Machine (SVM) have demonstrated promising performance in sentiment classification across various domains, including social media, e-commerce, and public service applications (Setiawan et al., 2024). Nevertheless, studies focusing on user reviews of university mobile Learning Management Systems remain limited, particularly those employing a structured data mining framework to guide the entire analytical process. Most previous studies emphasize classification accuracy while providing limited discussion of systematic stages from data collection and preprocessing to model development and evaluation.

To address this gap, this study applies the Sample, Explore, Modify, Model, and Assess (SEMMA) framework to conduct sentiment analysis of user reviews for the LMS Mobile application of Universitas XYZ. The framework is integrated with Natural Language Processing techniques, TF-IDF feature extraction, and two machine learning algorithms, namely Naïve Bayes and Support Vector Machine, to classify user sentiment and evaluate their performance. The findings are expected to provide data-driven insights into user experiences and support continuous improvement of digital learning services in higher education while demonstrating the applicability of the SEMMA framework for systematic sentiment analysis.

METHOD

The This study employed a quantitative experimental approach to analyze user sentiment toward the LMS Mobile application of Universitas XYZ using user reviews collected from the Google Play Store. A total of 129 reviews posted between November 2020 and December 2025 were collected through web scraping and used as the research dataset. User reviews were selected because they reflect users' experiences and perceptions regarding application usability, system performance, and service quality. Prior to analysis, the collected reviews were manually labeled into three sentiment categories, namely positive, negative, and neutral, to support supervised machine learning classification and comparative model evaluation (Pambudi et al., 2025).

This study adopted the Sample, Explore, Modify, Model, and Assess (SEMMA) framework to ensure a systematic sentiment analysis process. During the Sample stage, user reviews were collected and prepared as the research dataset (Andrade-Arenas et al., 2024). The Explore stage examined the characteristics of the reviews to identify sentiment tendencies and data distribution. The Modify stage applied Natural Language Processing (NLP) techniques, including cleaning, case folding, tokenization, stopword removal, and stemming, to transform unstructured textual data into a format suitable for machine learning (Siddharth et al., 2022). The processed text was subsequently converted into numerical features using the Term Frequency–Inverse Document Frequency (TF-IDF) method (Cahyani & Patasik., 2021). Finally, the Model stage implemented Naïve Bayes and Support Vector Machine (SVM) algorithms for sentiment classification (Kaisiang & Shah., 2022; Jannah & Kusnawi., 2024), while the Assess stage evaluated model performance using a confusion matrix (Rahmad et al., 2020).

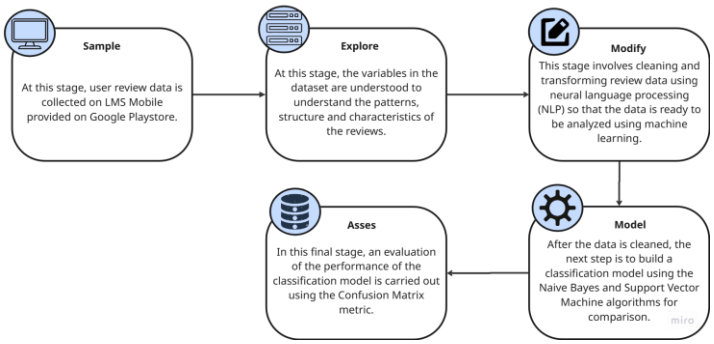


Figure 1. Research Framework Based on the SEMMA Methodology

Model performance was evaluated using a confusion matrix to measure the effectiveness of the sentiment classification models. The confusion matrix identifies correctly and incorrectly classified instances, allowing the calculation of several evaluation metrics, including accuracy, precision, recall, and F1-score (Sathyanarayanan & Tantri., 2024). These metrics were used to compare the classification performance of the Naïve Bayes and Support Vector Machine models.

Table 1. Confusion Matrix for Classification Evaluation

Predicted / Actual	Positive (Actual)	Negative (Actual)
Positive (Predicted)	True Positive (TP): Positive reviews correctly classified as positive	False Positive (FP): Negative reviews incorrectly classified as positive
Negative (Predicted)	False Negative (FN): Positive reviews incorrectly classified as negative	True Negative (TN): Negative reviews correctly classified as negative

RESULT AND DISCUSSION

A total of 129 user reviews of the LMS Mobile application of Universitas XYZ were collected from the Google Play Store between November 2020 and December 2025. The reviews represent users' experiences regarding application functionality, usability, system performance, and the quality of digital learning services (Yustanti et al., 2023). Before the classification process, each review was manually labeled into three sentiment categories: positive, negative, and neutral to support supervised machine learning classification.

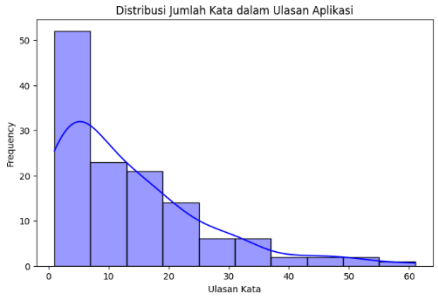


Figure 2. Distribution of Sentiment Categories

The sentiment distribution shows that negative sentiment dominated the dataset, indicating that many users encountered technical problems while using the application. Most negative reviews were associated with login failures, application errors, delayed system responses, and limited accessibility to learning features. These findings are consistent with previous studies reporting that system quality, application stability, and usability significantly influence user satisfaction with mobile learning applications (Aljouni et al., 2024; Ikhsan et al., 2022). Therefore, user-generated reviews can serve as valuable feedback for evaluating and improving the quality of digital learning services. To provide a broader understanding of user perceptions, the distribution of user ratings was also analyzed. The rating distribution demonstrates different levels of user satisfaction, with one-star ratings accounting for the largest proportion of the dataset.

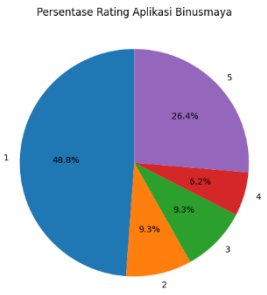


Figure 3. Rating Distribution

The predominance of low ratings supports the sentiment analysis results, indicating that many users were dissatisfied with the LMS Mobile application. Nevertheless, positive ratings were also observed, suggesting that some users considered the application beneficial for accessing learning materials and supporting academic activities. These findings demonstrate that user reviews provide important information regarding both the strengths and weaknesses of the application, enabling institutions to prioritize improvements based on actual user experiences. Before the sentiment classification process, several Natural Language Processing (NLP) techniques were applied during the Modify stage of the SEMMA framework to transform raw textual data into a structured format suitable for machine learning analysis. The preprocessing stage included text cleaning, tokenization, case folding, stopword removal, and stemming. These techniques reduce irrelevant information, eliminate textual inconsistencies, and improve the overall quality of textual data prior to feature extraction (Siddharth et al., 2022). The first preprocessing step involved text cleaning by removing punctuation marks, numbers, URLs, symbols, and other unnecessary characters while preserving meaningful textual information.

Table 2. Text Cleaning Results

At	Username	Score	Content	Clenning
2025-12-03 12:04:59	Rahmat Rizki Hidayah	4	Cukup bagus aplikasi nya	Cukup bagus aplikasi nya
2025-11-23 12:56:57	Sri Mini	3	Setelah pembaruan aplikasicukup membantu	Setelah pembaruan aplikasicukup membantu
2025-11-30 11:07:01	Aziz	3	masih ada beberapa fitur yg masih belum bisa	masih ada beberapa fitur yg masih belum bisa
2025-11-27 15:10:58	All	2	kenapa lemot bngt skrng ni woiiii, jadwal jg	kenapa lemot bngt skrng ni woiiii jadwal jg
2025-10-30 03:59:15	Vincent Marvellino	2	sering error	sering error

After text cleaning, tokenization divided each review into individual words to facilitate computational analysis. Case folding was subsequently performed to convert all characters into lowercase, ensuring consistency between identical words written with different letter cases. Stopword removal eliminated frequently occurring words with limited semantic value, while stemming transformed words into their root forms to reduce linguistic variation and improve feature consistency.

```
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory

factory = StemmerFactory()
stemmer = factory.create_stemmer()
df['stemming'] = df['stopword_removal'].apply(
    lambda tokens: [stemmer.stem(word) for word in tokens]
    if isinstance(tokens, list) else tokens)
df[['at', 'userName', 'score', 'content', 'cleaning', 'tokenized', 'case_folding', 'stopword_removal', 'stemming']].head(10)
```

	at	userName	score	content	cleaning	tokenized	case_folding	stopword_removal	stemming
0	2025-12-03 12:04:59	Rahmat Rizki Hidayah	4	Cukup bagus aplikasi nya	Cukup bagus aplikasi nya	[Cukup, bagus, aplikasi, nya]	[cukup, bagus, aplikasi, nya]	[bagus, aplikasi, nya]	[bagus, aplikasi, nya]
1	2025-11-30 11:07:01	Aziz	3	masih ada beberapa fitur yg masih belum bisa d...	masih ada beberapa fitur yg masih belum bisa d...	[masih, ada, beberapa, fitur, yg, masih, belum...	[masih, ada, beberapa, fitur, yg, masih, belum...	[fitur, yg, dibuka]	[fitur, yg, buka]
2	2025-11-27 15:10:58	Ali	2	kenapa lemot bngt skrng ni woiiii, jadwal jg k...	kenapa lemot bngt skrng ni woiiii jadwal jg ka...	[kenapa, lemot, bngt, skrng, ni, woiiii, jadwa...	[kenapa, lemot, bngt, skrng, ni, woiiii, jadwa...	[lemot, bngt, skrng, ni, woiiii, jadwal, jg, k...	[lot, bngt, skrng, ni, woiiii, jadwal, jg, kad...
3	2025-11-23 12:56:57	Sri Mini	3	setelah pembaruan aplikasi cukup membantu	setelah pembaruan aplikasi cukup membantu	[setelah, pembaruan, aplikasi, cukup, membantu]	[setelah, pembaruan, aplikasi, cukup, membantu]	[pembaruan, aplikasi, membantu]	[baru, aplikasi, bantu]

Figure 4. Final Preprocessing Results

The preprocessing results indicate that the textual data became more standardized and suitable for machine learning analysis. By removing redundant terms and normalizing word representations, the preprocessing stage reduced data complexity while preserving important semantic information required for sentiment classification. Following preprocessing, the cleaned textual data were transformed into numerical representations using the Term Frequency–Inverse Document Frequency (TF-IDF) method. TF-IDF assigns weights to each term based on its frequency within individual documents and across the entire dataset, allowing informative words to receive greater importance while reducing the influence of commonly occurring but less discriminative terms (Cahyani & Patasik., 2021).

The TF-IDF transformation produced a feature matrix consisting of 129 documents and 490 unique features, indicating that each user review had been successfully converted into numerical vectors suitable for machine learning classification.

```
from sklearn.feature_extraction.text import TfidfVectorizer

# Join the stemmed tokens into a single string for each review
df['final_text'] = df['stemming'].apply(lambda x: ' '.join(x) if isinstance(x, list) else '')

tfidf = TfidfVectorizer(
    max_features=3000   # aman untuk dataset kecil (129 data)
)

X = tfidf.fit_transform(df['final_text'])

print(X.shape)
```

```
(129, 490)
```

Figure 5. TF-IDF Feature Extraction Results

To evaluate the classification models objectively, the dataset was divided into training and testing subsets using an 80:20 ratio. A total of 103 reviews were allocated for model training, while the remaining 26 reviews were reserved for testing.

```
Jumlah data latih : (103, 490)
Jumlah data uji   : (26, 490)
```

Figure 6. Training and Testing Data Distribution

Separating the dataset into training and testing subsets enables the classification models to learn sentiment patterns from historical data while evaluating their predictive performance using previously unseen reviews. This strategy helps reduce overfitting and provides a more reliable assessment of the models' ability to generalize to new data (Verbraaken et al., 2020).

Following the feature extraction and data partitioning stages, sentiment classification was performed using two supervised machine learning algorithms, namely Naïve Bayes and Support Vector Machine (SVM). Both models were trained using the TF-IDF feature representation and classified user reviews into three sentiment categories: negative, neutral, and positive. The

classification results were first analyzed using confusion matrices to evaluate how accurately each model predicted the predefined sentiment classes.

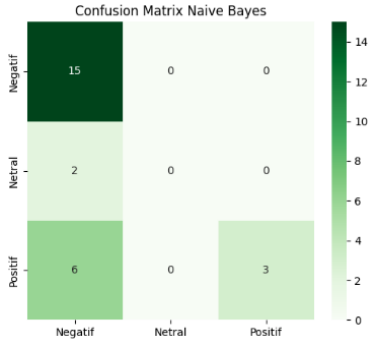


Figure 7A. Confusion matrix of Naïve Bayes

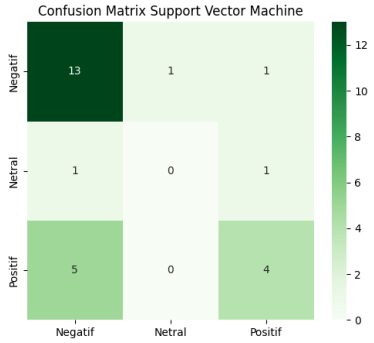


Figure 7B. Confusion matrix of Support Vector Machine

The confusion matrix indicates that the Naïve Bayes model correctly classified 15 negative reviews and 3 positive reviews. However, the model was unable to correctly identify reviews with neutral sentiment, as all neutral instances were classified into other sentiment categories. Furthermore, six positive reviews were incorrectly predicted as negative. These findings suggest that the Naïve Bayes classifier exhibited a strong tendency toward the dominant negative class, indicating that class imbalance affected the classification performance (Ressan & Hassan., 2022).

The Support Vector Machine model demonstrated a relatively more balanced prediction pattern. The model correctly classified 13 negative reviews, while two negative reviews were incorrectly assigned to the neutral and positive classes. In addition, four positive reviews were correctly identified, whereas five positive reviews were predicted as negative. Similar to the Naïve Bayes model, the SVM algorithm also failed to accurately recognize neutral sentiment (Sinulingga & Sitorus., 2024). Nevertheless, its prediction distribution across the sentiment classes was more balanced, suggesting better robustness despite the imbalanced dataset. These findings are consistent with previous comparative studies reporting that SVM generally provides more stable decision boundaries, while Naïve Bayes performs efficiently but may be more sensitive to class distribution (Jannah & Kusnawi., 2024; Kaisiang & Shah., 2022).

To further evaluate classification performance, both models were assessed using confusion matrix-based evaluation metrics, including precision, recall, F1-score, and overall accuracy (Sathyanarayanan & Tantri., 2024).

```
print(classification_report(y_test, y_pred_nb))
```

	precision	recall	f1-score	support
Negatif	0.65	1.00	0.79	15
Netral	0.00	0.00	0.00	2
Positif	1.00	0.33	0.50	9
accuracy			0.69	26
macro avg	0.55	0.44	0.43	26
weighted avg	0.72	0.69	0.63	26

Figure 8A. Classification report of Naïve Bayes

```
print(classification_report(y_test, y_pred_svm))
```

	precision	recall	f1-score	support
Negatif	0.68	0.87	0.76	15
Netral	0.00	0.00	0.00	2
Positif	0.67	0.44	0.53	9
accuracy			0.65	26
macro avg	0.45	0.44	0.43	26
weighted avg	0.63	0.65	0.63	26

Figure 8B. Classification report of Support Vector Machine

The evaluation results indicate that the Naïve Bayes model achieved an overall accuracy of 0.69 with a weighted F1-score of 0.63. The model obtained a recall value of 1.00 for the negative sentiment class, indicating that all negative reviews in the testing dataset were successfully identified. However, the precision for this class reached 0.65, suggesting that some reviews belonging to other sentiment categories were incorrectly predicted as negative. For the positive class, the model achieved perfect precision (1.00) but a relatively low recall (0.33), indicating that although positive predictions were highly accurate, many actual positive reviews remained undetected. Meanwhile, both precision and recall for the neutral class were 0.00, demonstrating the model's inability to distinguish neutral sentiment effectively.

The Support Vector Machine model achieved an overall accuracy of 0.65 with a weighted F1-score of 0.63. For the negative sentiment class, the model obtained a precision of 0.68 and a recall of 0.87, indicating relatively balanced performance in recognizing negative reviews. In the

positive sentiment category, the precision and recall values were 0.67 and 0.44, respectively, reflecting moderate capability in identifying positive opinions. Similar to Naïve Bayes, the SVM model was unable to correctly classify neutral reviews, resulting in zero precision and recall for that category. The limited number of neutral reviews and the linguistic similarity between neutral and other sentiment categories likely contributed to this performance.

Overall, both machine learning algorithms demonstrated satisfactory performance in identifying user sentiment toward the LMS Mobile application. Based on the evaluation metrics, Naïve Bayes achieved the highest overall accuracy (0.69), whereas Support Vector Machine exhibited a more balanced classification performance across the sentiment classes. The superior accuracy of Naïve Bayes may be attributed to its effectiveness in learning dominant patterns from relatively small textual datasets. In contrast, the SVM classifier generated more evenly distributed predictions despite slightly lower accuracy. Similar findings have been reported in previous studies comparing Naïve Bayes and SVM for sentiment classification, where Naïve Bayes often achieves competitive accuracy on limited datasets, while SVM demonstrates better generalization under more balanced data conditions (Jannah & Kusnawi., 2024; Pambudi et al., 2025).

These findings also reveal that the principal challenge in this study was the classification of neutral sentiment. The absence of correctly classified neutral reviews indicates that future research should consider larger and more balanced datasets or alternative feature extraction and classification techniques to improve model performance. Despite this limitation, the proposed SEMMA-based sentiment analysis framework successfully identified dominant user concerns, particularly regarding login issues, application errors, and system performance. Consequently, the findings provide practical insights for developers and higher education institutions in prioritizing improvements to mobile learning services based on user-generated feedback.

The novelty of this study lies in the integration of the SEMMA (Sample, Explore, Modify, Model, and Assess) framework (Gómez Palacios et al., 2017) with Natural Language Processing (NLP) techniques (Patil et al., 2023) and a comparative evaluation of Naïve Bayes and Support Vector Machine (SVM) algorithms for sentiment analysis of user reviews on a university mobile Learning Management System (LMS) (Rini et al., 2023). Previous sentiment analysis studies have been conducted across various domains, including social media, online reviews, and tourism-related digital content (Singgalen, 2024), while relatively limited attention has been given to user-generated feedback on mobile LMS applications in higher education. Furthermore, this study not only compares the classification performance of two widely used machine learning algorithms but also demonstrates the application of a structured data mining framework to systematically transform unstructured user reviews into meaningful insights for evaluating digital learning services. This approach provides a comprehensive perspective that combines data preprocessing, feature extraction, model development, and performance evaluation within a single analytical workflow.

The findings of this study reveal that negative sentiment dominated user reviews of the LMS Mobile application, indicating that technical issues remain the primary concern among users. Most complaints were associated with login failures, application errors, delayed system responses, and limited accessibility to learning features. The comparative evaluation further demonstrated that the Naïve Bayes algorithm achieved a higher classification accuracy (0.69) than the Support Vector Machine model (0.65), whereas the SVM algorithm produced a more balanced classification performance across sentiment categories. However, both models experienced difficulties in correctly identifying neutral sentiment, suggesting that class imbalance and the limited number of neutral reviews influenced the classification results (Ganganwar & Rajalakshmi, 2024). Overall, these findings highlight the importance of systematically analyzing user-generated feedback to identify system weaknesses and support evidence-based improvements in mobile learning applications (Hokijuliandy et al., 2023).

The results of this study are generally consistent with previous research investigating sentiment analysis using machine learning techniques. Previous research has demonstrated that Naïve Bayes can achieve strong classification performance in sentiment analysis when combined with TF-IDF feature weighting (Wati et al., 2023 ; Zulfikar et al., 2023). Other studies have also shown that SVM provides strong classification performance and is capable of handling high-dimensional data (Abdullah & Abdulazeez, 2021), particularly when textual data are represented using TF-IDF features (Zhou et al., 2024). In addition, previous research on Learning Management Systems has demonstrated that user satisfaction is influenced by several factors, including

information quality, service quality, and perceived ease of use (Ohliati & Abbas, 2019). The predominance of negative sentiment identified in this study therefore supports the broader understanding that technology and system quality issues can negatively affect users' experiences and their confidence in digital learning platforms (Ngulube & Ncube, 2025). These similarities indicate that sentiment analysis can serve as an effective approach for evaluating user satisfaction and identifying priority areas for improving educational technology services.

Although the overall findings are largely consistent with previous studies, several differences should also be considered. The higher classification accuracy achieved by the Naïve Bayes model in this study may be influenced by the relatively small dataset and the imbalanced distribution of sentiment classes, where negative reviews constituted the largest proportion of the collected data (Johnson & Khoshgoftaar, 2019). Under such conditions, probabilistic classifiers often learn dominant patterns more effectively than margin-based classifiers. On the other hand, the inability of both algorithms to accurately classify neutral sentiment may have resulted from the limited number of neutral samples and the semantic similarity between neutral expressions and other sentiment categories, as previous research has shown that neutral and positive sentiment samples can be difficult to separate due to their similarity in meaning (Salur & Aydin, 2020). Differences between this study and previous research may also be attributed to variations in research contexts, preprocessing techniques, feature extraction methods, dataset characteristics, and model configurations, as preprocessing strategies can substantially influence text classification performance depending on the dataset and classification model employed (Siino et al., 2024). Consequently, model performance should be interpreted within the specific characteristics of each dataset rather than being generalized across different application domains (Zhang et al., 2024).

This study contributes to the growing body of knowledge on sentiment analysis in educational technology by demonstrating the applicability of the SEMMA framework for systematically analyzing user reviews of mobile Learning Management Systems. Previous research has similarly demonstrated the potential of NLP-based approaches for systematically analyzing student feedback and generating actionable insights for educational institutions (Indriani et al., 2024). From a practical perspective, the findings provide valuable information for universities and application developers to prioritize improvements related to system stability, login accessibility, application performance, and feature usability. Previous research has similarly emphasized the importance of addressing technical issues, IT assistance, and user-related challenges to improve the effectiveness of LMS implementation in higher education (Ali et al., 2023 Academically, this study expands the application of machine learning-based sentiment analysis within higher education and provides a foundation for future comparative research involving larger datasets, additional machine learning algorithms, and deep learning approaches, which offer automated feature learning capabilities and have demonstrated strong applicability to high-dimensional textual data and NLP tasks (Janiesch et al., 2021). Future research could extend this study by incorporating transformer-based approaches such as BERT, which leverage bidirectional contextual representations and self-attention mechanisms to capture contextual relationships in textual data and have been widely applied to sentiment analysis tasks (Gardazi et al., 2025). Such developments are expected to improve sentiment classification performance and generate more comprehensive insights into user experiences across various educational technology platforms.

CONCLUSION

This study analyzed user sentiment toward the LMS Mobile application of Universitas XYZ using a machine learning-based sentiment analysis approach within the SEMMA (Sample, Explore, Modify, Model, Assess) framework. User reviews collected from the Google Play Store were processed using Natural Language Processing (NLP) techniques and TF-IDF feature extraction before being classified using the Naïve Bayes and Support Vector Machine (SVM) algorithms. The findings revealed that user sentiment was predominantly negative, reflecting recurring issues related to login accessibility, application errors, system performance, and feature usability. These results indicate that technical reliability remains a key factor influencing user satisfaction with mobile learning applications.

The comparative evaluation showed that the Naïve Bayes model achieved higher classification accuracy (0.69) than the SVM model (0.65), suggesting that Naïve Bayes was more effective for the available dataset. Nevertheless, the SVM model demonstrated more balanced performance across sentiment classes, indicating better stability in handling sentiment

classification. Both models experienced difficulties in identifying neutral sentiment, primarily due to class imbalance, ambiguous expressions, and the limited number of neutral reviews.

This study contributes to the growing body of research on sentiment analysis in educational technology by demonstrating the applicability of the SEMMA framework for systematically analyzing user feedback from mobile learning platforms. From a practical perspective, the findings provide evidence-based insights that can assist Universitas XYZ in improving the quality, reliability, and usability of its LMS Mobile application based on actual user experiences.

Despite these contributions, this study is limited by the relatively small dataset of 129 user reviews and the imbalance among sentiment classes, which may have affected classification performance. Future research is recommended to utilize larger and more balanced datasets, investigate advanced machine learning or deep learning approaches, and evaluate user sentiment across broader educational technology platforms to obtain more comprehensive and generalizable findings.

REFERENCES

Abdullah, D. M., & Abdulazeez, A. M. (2021). *Machine Learning Applications based on SVM Classification: A Review.* Qubahan Academic Journal, 1(2), 81–90. DOI: 10.48161/qaj.v1n2a50.

Ali, S., Fatima, F., Hussain, J., Qureshi, M. I., Fatima, S., & Zahoor, A. (2023). International Journal of Interactive Mobile Technologies (iJIM), 17(13), 4–21. DOI: 10.3991/ijim.v17i13.39785.

Aljouni, A., Alkhawaldeh, R. S., Al-Okaily, M. M., Al-Debei, M. M., & Alshurideh, M. (2024). Factors influencing student satisfaction with LMS mobile application: A TAM-based SEM analysis. *International Journal of Interactive Mobile Technologies, 18*(19). https://doi.org/10.3991/ijim.v18i19.50237

Andrade-Arenas, L., Rubio-Paucar, I., & Yactayo-Arias, C. (2024). Data mining for predictive analysis in gynecology: A focus on cervical health. *International Journal of Electrical and Computer Engineering, 14*(3), 2822–2833. https://doi.org/10.11591/ijece.v14i3.pp2822-2833

Cahyani, D. E., & Patasik, I. (2021). Performance comparison of TF-IDF and Word2Vec models for emotion text classification. *Bulletin of Electrical Engineering and Informatics, 10*(5), 2780–2788. https://doi.org/10.11591/eei.v10i5.3157

Ganganwar, V., & Rajalakshmi, R. (2024). *Employing synthetic data for addressing the class imbalance in aspect-based sentiment classification.* Journal of Information and Telecommunication, 8(2), 167–188. DOI: 10.1080/24751839.2023.2270824.

Gardazi, N. M., Daud, A., Malik, M. K., Bukhari, A., Alsahfi, T., & Alshemaimri, B. (2025). BERT applications in natural language processing: A review. *Artificial Intelligence Review, 58*, 166. DOI: 10.1007/s10462-025-11162-5.

Gómez Palacios et al. (2017). *A Comparative between CRISP-DM and SEMMA through the Construction of a MODIS Repository for Studies of Land Use and Cover Change*. DOI: 10.25046/aj020376

Hokijuliandy, E., Napitupulu, H., & Firdaniza. (2023). *Application of SVM and Chi-Square Feature Selection for Sentiment Analysis of Indonesia's National Health Insurance Mobile Application.* Mathematics, 11(17), 3765. DOI: 10.3390/math11173765.

Ikhsan, R. B., Prabowo, H., Yuniarty, Simamora, B., Ruan, X., & Kumar, V. (2022). Predicting students' use of mobile-learning management systems in Indonesia. *Journal of Educators Online, 19*(3). https://doi.org/10.9743/JEO.2022.19.3.5

Indriani, F., Nugroho, R. A., Faisal, M. R., & Kartini, D. (2024). Comparative evaluation of IndoBERT, IndoBERTweet, and mBERT for multilabel student feedback classification. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi), 8*(6), 748–757. DOI: 10.29207/resti.v8i6.6100.

Janiesch, C., Zschech, P., & Heinrich, K. (2021). *Machine learning and deep learning.* Electronic Markets, 31, 685–695. DOI: 10.1007/s12525-021-00475-2.

Jannah, N. Z. B., & Kusnawi. (2024). Comparison of Naïve Bayes and SVM in sentiment analysis of product reviews on marketplaces. *Sinkron: Jurnal dan Penelitian Teknik Informatika, 8*(2), 727–733. https://doi.org/10.33395/sinkron.v8i2.13559.

Johnson, J. M., & Khoshgoftaar, T. M. (2019). *Survey on deep learning with class imbalance.* Journal of Big Data, 6, 27. DOI: 10.1186/s40537-019-0192-5.

Kaisiang, C., & Shah, N. (2022). Comparison of Naive Bayes and SVM classification in grid-search hyperparameter tuned and non-hyperparameter tuned healthcare stock market sentiment analysis. *IAENG International Journal of Computer Science, 49*(4), 536–544.

Kurniawan, R., Wijaya, H. O. L., & Aprisusanti, R. P. (2024). Sentiment analysis of Google Play Store user reviews on digital population identity app using K-nearest neighbors. *Jurnal SISFOKOM (Sistem Informasi dan Komputer), 13*(2), 170–178. https://doi.org/10.1145/3569897

Mustaqim, I. Z., & Suryono, R. R. (2025). *A systematic literature review of topic modeling techniques in user reviews.* Journal of Information Systems Engineering and Business Intelligence, 11(2), 238–253. DOI: 10.20473/jisebi.11.2.238-253.

Ngulube, P., & Ncube, M. M. (2025). Leveraging learning analytics to improve the user experience of learning management systems in higher education institutions. *Information, 16*(5), 419. https://doi.org/10.3390/info16050419

Ohliati, J., & Abbas, B. S. (2019). *Measuring Students Satisfaction in Using Learning Management System.* International Journal of Emerging Technologies in Learning (iJET), 14(4), 180–189. DOI: 10.3991/ijet.v14.i04.9427.

Pambudi, S., Setiaji, P., & Triyanto, W. A. (2025). Sentiment analysis of Fizzo Novel application using Support Vector Machine and Naïve Bayes algorithm with SEMMA framework. *Jurnal Teknik Informatika (JUTIF), 6*(4), 1861–1880. https://doi.org/10.52436/1.jutif.2025.6.4.4875

Patil, R., Boit, S., Gudivada, V., & Nandigam, J. (2023). *A Survey of Text Representation and Embedding Techniques in NLP.* IEEE Access, 11, 36120–36146. DOI: 10.1109/ACCESS.2023.3266377.

Rahmad, F., Suryanto, Y., & Ramli, K. (2020). Performance comparison of anti-spam technology using confusion matrix classification. *IOP Conference Series: Materials Science and Engineering, 879*(1). https://doi.org/10.188/1757-899X/879/1/012076

Ressan, M. B., & Hassan, R. F. (2022). Naïve-Bayes family for sentiment analysis during COVID-19 pandemic and classification tweets. *Indonesian Journal of Electrical Engineering and Computer Science, 28*(1), 375–383. https://doi.org/10.11591/ijeecs.v28.i1.pp375-383

Rini, F., Yelfiza, & Pernanda, A. Y. (2023). *New LMS Mobile Framework Based on Multiplatform: A Literature Review of Mobile LMS Theory, Design and Implementation.* DOI: 10.47191/ijmra/v6-i6-54.

Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN Computer Science, 2*(3), 160. http://doi.org/10.1007/s42979-021-00592-x

Sathyanarayanan, S., & Tantri, B. R. (2024). Confusion matrix-based performance evaluation metrics. *African Journal of Biomedical Research, 27*(4S), 4023–4031. https://doi.org/10.53555/AJBR.v27i4S.4345

Salur, M. U., & Aydin, I. (2020). *A novel hybrid deep learning model for sentiment classification.* IEEE Access, 8, 58080–58093. DOI: 10.1109/ACCESS.2020.2982538.

Setiawan, T., Liem, S., & Pribadi, M. R. (2024). *Perbandingan algoritma SVM dan Naïve Bayes dalam analisis sentimen komentar TikTok pada produk skincare.* Applied Information Technology and Computer Science (AICOMS), 3(2), 28–32. http://doi.org/10.58466/aicoms.v3i1.1523

Siddharth, L., Blessing, L., & Luo, J. (2022). Natural language processing in-and-for design research. *Design Science, 8*, e21. http://doi.org/10.1017/dsj.2022.16.

Siino, M., Tinnirello, I., & La Cascia, M. (2024). *Is text preprocessing still worth the time? A comparative survey on the influence of popular preprocessing methods on Transformers and traditional classifiers.* Information Systems, 121, 102342. DOI: 10.1016/j.is.2023.102342.

Singgalen, Y. A. (2024). *Distribution of Polarity Value between VADER and TextBlob in Sentiment Classification of Tourist Vlog Content Reviews.* Journal of Information System Research (JOSH), 5(4), 1445–1454. DOI: 10.47065/josh.v5i4.5501.

Sinulingga, J. E. B., & Sitorus, H. C. K. (2024). Analisis sentimen masyarakat terhadap film horor Indonesia menggunakan metode SVM dan TF-IDF. *Jurnal Manajemen Informatika (JAMIKA), 14*(1), 42–53. http://doi.org/10.34010/jamika.v14i1.11946

Van Mechelen, M., et al. (2023). Emerging technologies in K–12 education: A future HCI research agenda. *ACM Transactions on Computer-Human Interaction, 30*(3), 1–40. http://doi.org/0.1145/3569897

Verbraaken, J., Wolting, M., Katzy, J., Kloppenburg, J., Verbelen, T., & Rellermeyer, J. S. (2020). A survey on distributed machine learning. *ACM Computing Surveys, 53*(2), 1–33. https://doi.org/10.1145/3377454

Wati, R., Ernawati, S., & Rachmi, H. (2023). *Pembobotan TF-IDF Menggunakan Naïve Bayes Pada Sentimen Masyarakat Mengenai Isu Kenaikan BIPIH.* DOI: 10.34010/jamika.v13i1.9424.

Yelfiza, Rini, F., & Pernanda, A. Y. (2024). Measuring user satisfaction: Open source LMS mobile application as a learning technology. *Journal of Advanced Research in Applied Sciences and Engineering Technology, 38*(2), 176–185. http://doi.org/10.37934/araset.38.2.176185.

Yustanti, W., Iriawan, N., & Irhamah. (2023). Categorical encoder based performance comparison in pre-processing imbalanced multiclass classification. *Jurnal Teknik Industri, 25*(1), 39–54. https://doi.org/10.9744/jti.25.1.39-54

Zhang, Y., Li, Q., & Xin, Y. (2024). *Research on eight machine learning algorithms applicability on different characteristics data sets in medical classification tasks.* DOI: 10.3389/fncom.2024.1345575.

Zhou, J., Ye, Z., Zhang, S., Geng, Z., Han, N., & Yang, T. (2024). *Investigating response behavior through TF-IDF and Word2vec text analysis: A case study of PISA 2012 problem-solving process data.* DOI: 10.1016/j.heliyon.2024.e35945.

Zulfikar, W. B., Atmadja, A. R., & Pratama, S. F. (2023). Sentiment Analysis on Social Media Against Public Policy Using Multinomial Naive Bayes, 10(1), 25–34. DOI: 10.15294/sji.v10i1.39952.