International Journal Software Engineering and Computer Science (IJSECS)
6 (1), 2026, 183-193
Published Online April 2026 in IJSECS (http://www.journal.lembagakita.org/index.php/ijsecs)
P-ISSN: 2776-4869, E-ISSN: 2776-3242. DOI: https://doi.org/10.35870/ijsecs.v6i1.6862.

RESEARCH ARTICLE Open Access

A Comparative Performance Analysis of Naïve Bayes, LSTM, and BiLSTM with Data Balancing Techniques for Sentiment Analysis of EasyCash Application Reviews

Fitri Abelia [1*], **Fitriyani** [2]

[1*,2] Institut Sains dan Bisnis Atma Luhur, Pangkal Pinang City, Bangka Belitung Islands Province, Indonesia.

**Corresponding author*: 2222500102@mahasiswa.atmaluhur.ac.id.

Received: March 27, 2026; Accepted: April 1, 2026; Published: April 10, 2026.

Abstract: This study compares the performance of Naïve Bayes, Long Short-Term Memory (LSTM), and Bidirectional LSTM (BiLSTM) models in sentiment analysis of EasyCash application reviews, with data balancing techniques applied throughout the process. The dataset was collected from the Google Play Store and processed through cleaning, tokenization, stemming, and normalization. Sentiment labeling classified reviews into positive, neutral, and negative categories. To address class imbalance, the Synthetic Minority Oversampling Technique (SMOTE) was applied prior to model training. Feature extraction was conducted using TF-IDF, and models were evaluated on accuracy, precision, recall, and F1-score. Naïve Bayes outperformed both LSTM and BiLSTM, producing higher accuracy and more stable results across evaluation metrics. The findings suggest that simpler machine learning models can be more effective than deep learning approaches when working with limited and imbalanced datasets. Careful data preprocessing, appropriate balancing techniques, and deliberate model selection remain central to achieving reliable sentiment classification performance in fintech applications.

Keywords: Sentiment Analysis; Naïve Bayes; Fintech Reviews.

1. Introduction

The rapid advancement of information technology has compelled various industrial sectors to continuously adapt to evolving societal and economic demands, including the financial technology (fintech) sector. Fintech has emerged as a transformative force that reshapes conventional financial services into more flexible, efficient, and technology-driven systems. One of the most prominent innovations in this domain is online lending, which provides faster and more accessible financial services through digital platforms. Applications such as EasyCash exemplify this transformation by enabling users to access loan services conveniently via smartphones — a shift that aligns with the broader objective of fintech in promoting financial inclusion and expanding access to financial services for diverse populations (Feriyanto *et al.*, 2024; Riska *et al.*, 2025). In Indonesia, fintech has experienced significant growth over the past decade, driven by increasing digital adoption, supportive regulatory frameworks, and the demand for inclusive financial services. Regulatory bodies such as the Financial Services Authority (OJK) and Bank Indonesia (BI) have implemented various policies, including regulatory sandboxes, to ensure innovation while maintaining financial stability and consumer protection. The presence of licensed fintech lending companies and the increasing volume of online loan

disbursements indicate strong adoption of digital financial services among the population (Adji *et al.*, 2023; Gandasari *et al.*, 2025; Kwon *et al.*, 2023; Tritto *et al.*, 2020; Wulandari *et al.*, 2025).

Despite its benefits, the rapid expansion of fintech also introduces critical challenges related to consumer protection, data privacy, and cybersecurity. The increasing reliance on digital platforms necessitates robust governance frameworks to manage risks such as data misuse, cyberattacks, and fraud. Studies emphasize that the integration of data protection regulations and cybersecurity practices is essential to sustain user trust and ensure the long-term viability of fintech ecosystems. The harmonization of cross-sectoral regulations and the adoption of global best practices in data governance are equally pressing concerns (Algamar & Ismail, 2023; Hesniati & Limgestu, 2023; Laínez & Gardner, 2023; Mehrban *et al.*, 2020). The widespread use of fintech applications has, in turn, generated a large volume of user-generated content in the form of online reviews on platforms such as the Google Play Store. These reviews serve as valuable sources of information for evaluating service quality, user satisfaction, and potential issues within applications. Prior studies highlight that the quality, credibility, and content of reviews significantly affect perceived trust, risk, and behavioral intentions in online environments (Attar *et al.*, 2022; Helmi *et al.*, 2024; Sulistyowati & Husda, 2023).

Analyzing user reviews, however, presents several challenges due to their unstructured nature, informal language, and semantic ambiguity. Manual analysis is inefficient and prone to bias, necessitating the use of automated sentiment analysis techniques to classify opinions into positive, negative, and neutral categories. Various machine learning and deep learning approaches have been applied to this task. Traditional methods such as Naïve Bayes are known for their simplicity and robustness, particularly when dealing with small datasets, while deep learning models such as Long Short-Term Memory (LSTM) and Bidirectional LSTM (BiLSTM) are designed to capture sequential dependencies and contextual information in textual data. Prior studies indicate, however, that deep learning models do not always outperform traditional methods — especially when datasets are limited or exhibit class imbalance, which may lead to biased predictions toward majority classes (Alzoubi *et al.*, 2024; Shah & Patel, 2025). Techniques such as the Synthetic Minority Oversampling Technique (SMOTE) have been widely used to address this problem, yet the integration of data balancing techniques with both machine learning and deep learning models remains underexplored in the context of fintech application reviews. This study therefore conducts a comparative performance analysis of Naïve Bayes, LSTM, and BiLSTM models with data balancing techniques for sentiment analysis of EasyCash application reviews, contributing empirical evidence on model effectiveness under imbalanced data conditions and offering practical guidance for classification method selection in fintech-related sentiment analysis.

2. Related Work

2.1 Sentiment Analysis in Fintech Applications

Sentiment analysis has become a prominent research area within natural language processing (NLP), particularly in analyzing user-generated content such as online reviews. In the fintech domain, sentiment analysis plays a critical role in understanding user perceptions of digital financial services, including mobile lending applications. Unlike structured survey data, user reviews capture spontaneous and unfiltered opinions — making them a uniquely valuable source for assessing service quality, user satisfaction, and potential operational risks. As fintech platforms continue to grow in scale and user base, the volume of such reviews has expanded considerably, creating both an opportunity and a challenge for systematic analysis. Prior research has demonstrated that sentiment analysis on fintech platforms, such as Google Play Store reviews, can effectively support decision-making and service evaluation processes (Amrie *et al.*, 2022). The ability to automatically classify user sentiment at scale represents a meaningful step toward more responsive and user-centered financial service development.

2.2 Machine Learning Approaches for Sentiment Classification

Traditional machine learning methods, including Naïve Bayes and Support Vector Machine (SVM), have been widely applied in sentiment analysis due to their simplicity, computational efficiency, and relatively strong performance on structured textual data. These models are particularly well-suited for scenarios where labeled data is limited and training resources are constrained. Mongkito *et al.* (2024) showed that Naïve Bayes and SVM can effectively classify user reviews of online lending applications, producing reliable results even when the dataset size is modest. Similarly, Fullah (2025) achieved high classification performance using SVM in analyzing sentiment related to data misuse issues in fintech services, further confirming the method's applicability in domain-specific contexts. Collectively, these findings indicate that traditional machine learning models remain reliable and competitive — a point worth emphasizing given the prevailing tendency in recent literature to favor deep learning approaches regardless of data conditions.

2.3 Deep Learning Approaches in Sentiment Analysis

In contrast to traditional methods, deep learning models have gained considerable traction due to their capacity to capture contextual and sequential relationships within textual data. Long Short-Term Memory (LSTM) and Bidirectional LSTM (BiLSTM) are among the most widely used architectures for sentiment classification tasks, owing to their ability to model temporal dependencies and retain information across long sequences. Noveandini *et al*. (2025) demonstrated that LSTM models can effectively analyze sentiment in mobile application reviews, producing competitive accuracy scores across multiple evaluation metrics. Winarni *et al*. (2025) extended this line of inquiry by showing that combining LSTM with transformer-based techniques such as BERT further improves classification performance, particularly in capturing nuanced semantic context. These findings suggest that deep learning architectures hold genuine promise for sentiment analysis — provided that sufficient data and computational resources are available to support their training requirements.

2.4 Challenges of Imbalanced Data in Sentiment Analysis

Despite the advantages of both machine learning and deep learning methods, sentiment analysis frequently encounters challenges related to imbalanced datasets. In many real-world cases, certain sentiment classes — such as neutral or positive — dominate the dataset, producing skewed classification results that favor the majority class. This is not merely a statistical inconvenience; it directly undermines the practical utility of a model, particularly when the minority class carries the most actionable information, as is often the case with negative sentiment in service reviews. Shah & Patel (2025) found that deep learning models do not always outperform traditional approaches under these conditions, especially when datasets are small or heavily imbalanced. Alzoubi *et al*. (2024) further argued that class imbalance significantly degrades model performance, as classifiers tend to neglect minority classes in favor of those with greater representation — ultimately reducing classification reliability across the board.

2.5 Data Balancing Techniques for Improving Model Performance

To address class imbalance, various data balancing techniques have been proposed, with oversampling methods — particularly the Synthetic Minority Oversampling Technique (SMOTE) — receiving considerable attention in the literature. Rather than simply duplicating existing minority samples, SMOTE generates synthetic instances by interpolating between neighboring data points, producing a more diverse and representative minority class distribution. Assyifa & Luthfiarta (2024) demonstrated that SMOTE-based approaches can substantially improve classification performance by strengthening the representation of underrepresented classes, enabling models to learn more balanced decision boundaries. The practical benefit of this technique is well-documented across multiple classification domains. Its application in combination with both machine learning and deep learning models in fintech sentiment analysis, however, remains underexplored — a gap that warrants direct empirical attention.

2.6 Role of User Reviews in Digital Trust and Decision-Making

User reviews occupy a central position in digital ecosystems as indicators of trust, perceived risk, and information quality. On online platforms, reviews function as a form of electronic word-of-mouth (eWOM), shaping user behavior and influencing decision-making processes in ways that traditional marketing channels cannot easily replicate. The credibility and perceived helpfulness of a review can determine whether a prospective user chooses to adopt or avoid a service — a dynamic that is especially pronounced in fintech, where trust is both difficult to establish and easy to lose. Studies have consistently shown that high-quality and credible reviews can substantially strengthen trust and reduce uncertainty in digital transactions (Attar *et al*., 2022; Helmi *et al*., 2024; Sulistyowati & Husda, 2023). Accurate sentiment analysis of user reviews is therefore not merely an academic exercise — it carries direct implications for service improvement, user retention, and the long-term reputation of fintech platforms.

2.7 Research Gap and Contribution

A review of the existing literature reveals that prior studies have predominantly focused on either machine learning or deep learning approaches in isolation, with limited attention given to systematic comparative analysis across both model types under consistent experimental conditions. The integration of data balancing techniques with both categories of models in fintech-specific sentiment analysis remains insufficiently examined. Most studies either apply balancing techniques without comparing model families, or compare models without adequately controlling for class imbalance — making it difficult to isolate the true source of performance differences. This study addresses that gap directly by conducting a comparative performance analysis of Naïve Bayes, LSTM, and BiLSTM models combined with SMOTE-based data balancing for sentiment analysis of EasyCash application reviews. The research contributes empirical evidence on model effectiveness under imbalanced data conditions and offers practical guidance for selecting appropriate classification methods in similar fintech.

3. Methodology

This study employs a systematic approach to conduct sentiment analysis on user reviews of the EasyCash application, structured across several sequential stages from data collection through to comparative model evaluation. The research process begins with data collection using web scraping techniques from the Google Play Store, targeting reviews submitted by real users of the EasyCash application. These reviews represent diverse user experiences and opinions, making them a suitable source for multi-class sentiment classification. The collected data then undergoes a preprocessing stage designed to improve data quality and ensure consistency across the dataset. Raw user reviews typically contain considerable noise — including irrelevant characters, slang, abbreviations, and inconsistent spelling — all of which can interfere with downstream modeling. To address this, four preprocessing steps were applied sequentially: cleaning to remove irrelevant elements such as punctuation, numbers, and special symbols; tokenization to split text into individual word units; stemming to reduce words to their base forms; and normalization to standardize informal and non-standard language. Each of these steps contributes to producing a cleaner, more uniform textual representation suitable for machine learning algorithms. After preprocessing, the dataset was labeled into three sentiment categories — positive, neutral, and negative — using a lexicon-based or rating-based approach. Given that class imbalance is a common and well-documented challenge in textual datasets, the Synthetic Minority Oversampling Technique (SMOTE) was applied to balance the class distribution before model training. Rather than simply duplicating existing minority samples, SMOTE generates synthetic instances through interpolation, producing a more representative distribution across all sentiment classes. Following data balancing, the dataset was divided into training and testing sets using an 80:20 ratio, ensuring that model evaluation is conducted on data that was not seen during training.

Feature extraction was then performed using the Term Frequency–Inverse Document Frequency (TF-IDF) method, which transforms textual data into numerical vectors by weighting terms according to their frequency within a document relative to their frequency across the entire corpus. These feature vectors served as input for three classification models trained in parallel: Naïve Bayes, Long Short-Term Memory (LSTM), and Bidirectional LSTM (BiLSTM). Naïve Bayes was selected as a representative traditional machine learning baseline, while LSTM and BiLSTM were included to assess the extent to which deep learning architectures — capable of capturing sequential and bidirectional contextual dependencies — offer performance advantages under the given data conditions. The trained models were evaluated using accuracy, precision, recall, F1-score, and ROC curve analysis, providing a multi-dimensional view of classification performance. A comparative analysis was subsequently conducted to determine the most effective model for sentiment classification based on results across all evaluation metrics.

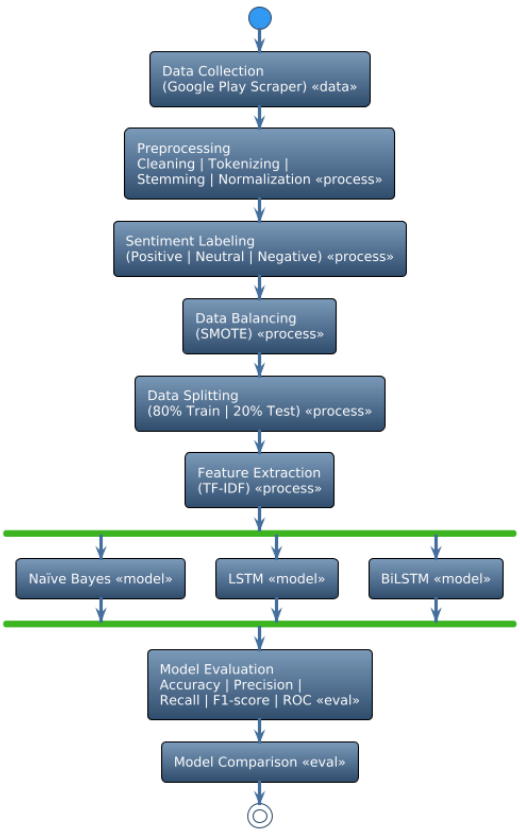

Figure 1. Research Framework Diagram

The diagram illustrates the overall workflow of the research methodology in a structured and sequential manner. It begins with data collection from the Google Play Store, followed by preprocessing steps to clean and standardize the textual data. The processed data is then labeled into sentiment classes and balanced using SMOTE to address class imbalance. After splitting the dataset into training and testing sets, feature extraction is performed using TF-IDF to convert text into numerical representations. Three classification models — Naïve Bayes, LSTM, and BiLSTM — are then trained in parallel. The models are subsequently evaluated using multiple performance metrics, and a comparative analysis is conducted to identify the most effective model for sentiment classification.

4. Result and Discussion

4.1 Results

4.1.1 Dataset Characteristics and Sentiment Distribution

The dataset used in this study consists of user reviews collected from the EasyCash application on the Google Play Store, representing real user experiences and opinions regarding the application's services, usability, and performance. After data collection, the dataset was preprocessed to remove noise and standardize textual content, ensuring suitability for sentiment analysis and machine learning modeling. The final dataset was then labeled into three sentiment categories: positive, neutral, and negative. The distribution of sentiment classes revealed a significant imbalance among categories. The neutral class dominated the dataset, accounting for approximately 62% of the total data, while positive sentiment represented around 25% and negative sentiment accounted for only 13%. This imbalance indicates that most users tend to provide neutral feedback — which may reflect moderate satisfaction or descriptive comments rather than strong opinions. Such skewed distribution poses challenges for classification models, as they may become biased toward the majority class and fail to adequately learn patterns associated with minority classes. Table 1 presents the distribution of sentiment classes before applying data balancing techniques.

Table 1. Sentiment Distribution Before Data Balancing

Sentiment	Reviews	Percentage
Positive	250	25%
Neutral	620	62%
Negative	130	13%

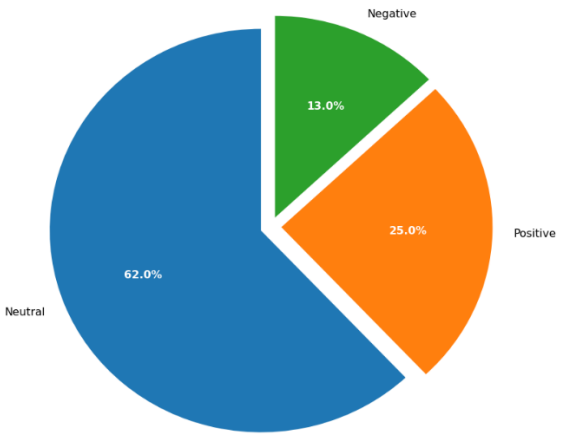

Figure 2. Sentiment Distribution Pie Chart

The pie chart illustrates the imbalance in sentiment distribution, where the neutral class dominates significantly compared to positive and negative classes. This imbalance can negatively affect model performance, particularly in identifying minority classes such as negative sentiment — the class that arguably carries the most actionable information for service improvement. Addressing this issue before proceeding to model training is therefore a necessary step, and it directly justifies the application of SMOTE in this study.

4.1.2 Data Preprocessing and Transformation

The preprocessing stage plays a critical role in preparing textual data for sentiment analysis. Multiple preprocessing techniques were applied sequentially, including cleaning, tokenization, stemming, and

normalization. Cleaning removed irrelevant characters such as punctuation, numbers, and special symbols. Tokenization split sentences into individual words, enabling further analysis at the word level. Stemming reduced words to their root forms, while normalization ensured consistency in word usage across the dataset. Together, these steps substantially improved data quality and reduced noise, making the dataset more suitable for machine learning algorithms. Without proper preprocessing, models may misinterpret irrelevant patterns or fail to capture meaningful linguistic relationships — a risk that is particularly pronounced in user-generated text, which tends to be informal, inconsistent, and noisy. The transformation process also strengthens the effectiveness of feature extraction methods such as TF-IDF, producing a more structured and informative representation for classification tasks.

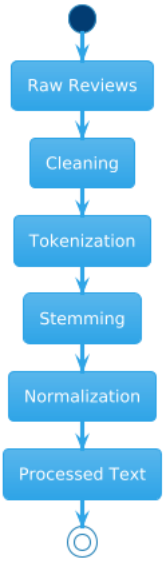

Figure 3. Sequential Preprocessing Steps Applied to the Dataset

The diagram illustrates the sequential preprocessing steps applied to the dataset. Each stage contributes to improving data quality and ensuring that the textual input is suitable for modeling. The transformation from raw text to processed text reflects the importance of preprocessing in achieving accurate sentiment classification, and this stage serves as the foundation for all subsequent feature extraction and modeling processes.

4.1.3 Impact of Data Balancing Using SMOTE

To address the issue of class imbalance, the Synthetic Minority Oversampling Technique (SMOTE) was applied to the dataset. Rather than simply duplicating existing minority samples, SMOTE generates synthetic instances through interpolation between neighboring data points, producing a more diverse and representative class distribution. After applying SMOTE, the dataset became evenly distributed across all three sentiment categories, with each class comprising 400 samples — a substantial improvement over the original skewed distribution. This balancing process is essential for ensuring that classification models can learn patterns from all classes equally, reducing the risk of bias toward the majority class and strengthening performance on minority classes. Consequently, evaluation metrics such as recall and F1-score for minority classes are expected to reflect more reliable results. Table 2 shows the distribution of sentiment classes after data balancing.

Table 2. Sentiment Distribution After Data Balancing

Sentiment	Number of Reviews	Percentage
Positive	400	33%
Neutral	400	33%
Negative	400	33%

The diagram confirms that the dataset becomes evenly distributed after applying SMOTE. This uniform balance allows models to learn more effectively from all sentiment classes, producing classification outcomes that are more reliable and less susceptible to majority-class bias. The use of SMOTE is therefore a necessary step in improving overall model performance, particularly in real-world datasets where class imbalance is the norm rather than the exception.

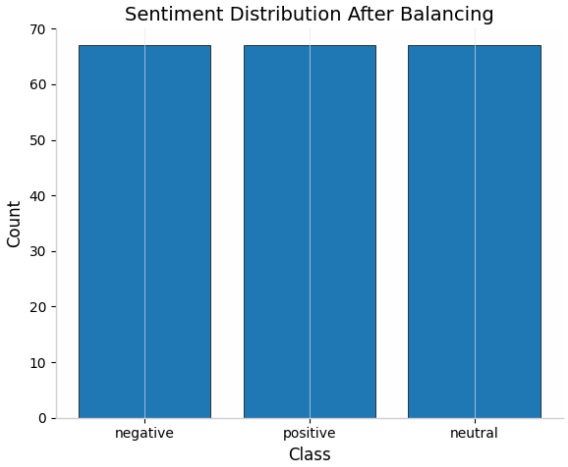

Figure 4. Data After the Balancing Process

4.1.4 Model Performance Evaluation

The performance of the three classification models was evaluated using accuracy, precision, recall, and F1-score — metrics that together provide a thorough assessment of model effectiveness across all sentiment classes. Accuracy measures overall correctness, while precision and recall evaluate the model's capacity to correctly identify specific classes. The F1-score, which balances precision and recall, is particularly informative in the context of multi-class classification where class distributions may still vary after balancing. Naïve Bayes achieved the highest performance among the three models, obtaining an accuracy of 0.70 with balanced precision and recall values. Both LSTM and BiLSTM, by contrast, achieved accuracy scores of only 0.50, with precision values of 0.25 — indicating poor performance in correctly identifying sentiment classes and suggesting that these models were unable to learn discriminative patterns from the available data. The performance gap between Naïve Bayes and the deep learning models is notable, and warrants careful interpretation rather than a straightforward conclusion that one model family is universally superior. Table 3 summarizes the comparative results.

Table 3. Model Performance Comparison

Model	Accuracy	Precision	Recall	F1-Score
Naïve Bayes	0.70	0.65	0.70	0.66
LSTM	0.50	0.25	0.50	0.33
BiLSTM	0.50	0.25	0.50	0.33

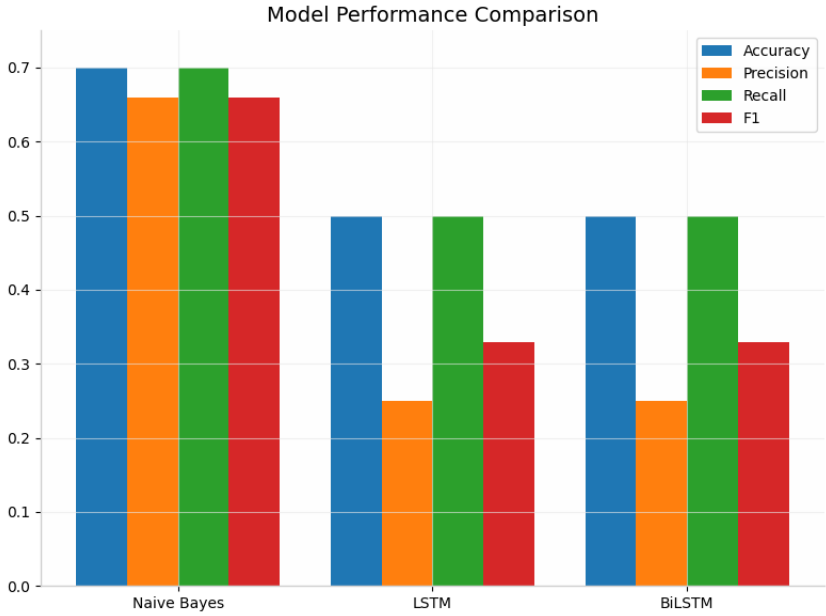

Figure 5. Model Performance Comparison

The bar chart clearly shows that Naïve Bayes outperforms both LSTM and BiLSTM across all evaluation metrics. This result is somewhat counterintuitive given the general perception that deep learning models are

architecturally more capable. The findings suggest, however, that model performance is highly dependent on dataset characteristics — and in this case, the limited dataset size likely constrained the learning capacity of the deep learning models, preventing them from realizing their theoretical advantages.

4.1.5 Discussion of Model Performance

The superior performance of Naïve Bayes can be attributed to its simplicity and effectiveness in handling small datasets. Unlike deep learning models, Naïve Bayes does not require large volumes of data to achieve reasonable performance, as it relies on probabilistic assumptions that allow it to generalize well even under limited data conditions. This property makes it particularly well-suited for text classification tasks where dataset size is a binding constraint. LSTM and BiLSTM, on the other hand, require substantial training data to effectively learn complex patterns and contextual relationships. In this study, the relatively limited dataset size likely hindered the learning process of these models, preventing them from outperforming the simpler Naïve Bayes baseline. Deep learning models are also more sensitive to residual data imbalance and feature sparsity — both of which may have further affected their performance in this setting.

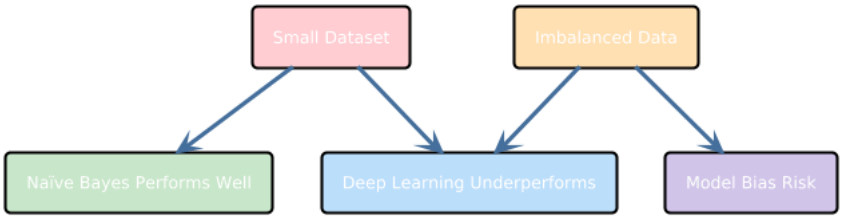

Figure 6. Relationship Between Dataset Characteristics and Model Performance

The diagram illustrates the relationship between dataset characteristics and model performance. Small and imbalanced datasets tend to constrain deep learning models disproportionately, while simpler probabilistic models such as Naïve Bayes remain more robust under these conditions. This finding reinforces the importance of selecting models based on actual dataset characteristics rather than defaulting to the most architecturally complex option available.

4.1.6 Key Insights and Implications

The results of this study yield several important observations about sentiment analysis in fintech applications. Deep learning models do not always outperform traditional machine learning methods — their performance is highly dependent on data size and quality, and architectural complexity alone does not guarantee better results. Data balancing techniques such as SMOTE are essential for improving classification performance in imbalanced datasets, and their application should be treated as a standard step rather than an optional one. The findings also point to the importance of preprocessing and feature extraction: proper data preparation substantially strengthens model performance by reducing noise and improving the quality of textual representation. Taken together, these observations challenge the common assumption that more complex models always yield better results, and suggest that methodological choices should be driven by data realities rather than by prevailing trends in the literature.

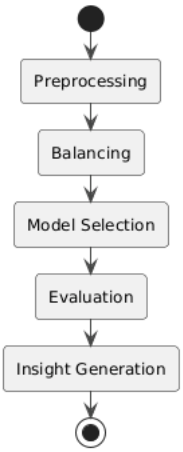

Figure 7. Factors That Contribute to Successful Sentiment Analysis

The diagram summarizes the key stages that contribute to successful sentiment analysis. Each stage plays a critical role in ensuring accurate and reliable classification outcomes, and the insights gained from this study can inform both future research directions and practical applications in fintech sentiment analysis.

4.2 Discussion

The findings of this study indicate that the performance of sentiment classification models is highly influenced by dataset characteristics, particularly size and class distribution. Naïve Bayes outperformed both LSTM and BiLSTM, achieving higher accuracy and more balanced evaluation metrics across all sentiment classes. This result aligns with Shah & Patel (2025), who reported that traditional machine learning models can outperform deep learning approaches when applied to relatively small datasets. Deep learning models typically require large volumes of data to effectively capture contextual patterns, and their performance tends to degrade when data is limited — a condition that was clearly present in this study. The superior performance of Naïve Bayes can therefore be attributed to its efficiency and robustness in handling smaller datasets, rather than to any inherent limitation of the deep learning architectures themselves. These results, however, contrast with findings from Noveandini *et al*. (2025) and Winarni *et al*. (2025), both of whom demonstrated that LSTM-based models can achieve higher accuracy in sentiment analysis tasks. Those studies point to the genuine strength of deep learning models in capturing sequential dependencies and contextual meaning in textual data. The discrepancy between their findings and those of the present study is likely explained by differences in dataset size, feature representation, and preprocessing techniques. In their studies, the use of larger datasets and advanced techniques such as BERT integration contributed to improved performance — conditions that were not replicated here, where the dataset size was relatively limited and thus constrained the learning capacity of both LSTM and BiLSTM.

The impact of class imbalance on model performance is equally apparent from the results. Before applying data balancing techniques, the dataset was dominated by the neutral class, which skewed classification outcomes in ways that would have disproportionately affected minority class detection. This observation is consistent with Alzoubi *et al*. (2024), who argued that imbalanced datasets produce biased predictions toward the majority class. Applying SMOTE successfully improved class distribution and strengthened the models' capacity to learn from minority classes. Assyifa & Luthfiarta (2024) reported comparable findings, confirming that SMOTE-based techniques can substantially improve classification performance by increasing minority class representation — a finding that holds across multiple classification domains, including the fintech context examined here. The role of preprocessing is equally evident from the results. Cleaning, tokenization, stemming, and normalization improved the overall quality of the dataset and facilitated more effective feature extraction using TF-IDF. Amrie *et al*. (2022) made a similar observation, noting that proper preprocessing substantially strengthens sentiment classification performance in Google Play Store review datasets. Effective preprocessing reduces noise and ensures that meaningful textual patterns are captured — a requirement that applies equally to machine learning and deep learning models, regardless of architectural complexity. Taken together, these findings suggest that model selection should be tailored to dataset characteristics rather than driven by assumptions about model superiority. The integration of data balancing techniques and comprehensive evaluation metrics provides a more reliable basis for sentiment analysis in real-world datasets where imbalance and noise are common and persistent challenges.

5. Conclusion and Recommendations

This study evaluated and compared the performance of Naïve Bayes, LSTM, and BiLSTM models for sentiment analysis of EasyCash application reviews, with data balancing techniques applied throughout the experimental pipeline. The findings indicate that Naïve Bayes outperformed both deep learning models in terms of accuracy and overall evaluation metrics — a result that, while counterintuitive at first glance, is consistent with the broader literature on model behavior under limited and imbalanced data conditions. Simpler machine learning models, the evidence suggests, can be more effective than architecturally complex approaches when the dataset is relatively small and class distribution is skewed. The implementation of SMOTE proved effective in addressing class imbalance, producing more uniform class distributions and contributing to more reliable classification outcomes across all sentiment categories. Preprocessing techniques — including cleaning, tokenization, stemming, and normalization — also played a measurable role in improving data quality and enabling more accurate feature representation through TF-IDF. None of these stages operated in isolation; the results reflect the cumulative effect of each methodological decision made throughout the pipeline.

Several directions merit attention in future research. Studies working with deep learning architectures such as LSTM and BiLSTM should prioritize the use of larger and more diverse datasets, given that the performance of these models is highly sensitive to data scale and quality. Exploring more advanced feature representation methods — such as word embeddings or transformer-based models like BERT — may also yield

meaningful improvements in classification accuracy, particularly for capturing semantic nuance in informal user-generated text. The use of hybrid or adaptive data balancing techniques, beyond standard SMOTE, is worth examining as a means of strengthening model robustness in real-world settings where class imbalance takes varied and unpredictable forms. From a practical standpoint, fintech companies stand to benefit from deploying automated sentiment analysis systems to monitor user feedback at scale, identify recurring service issues, and support more responsive customer experience management. When implemented thoughtfully, such systems can serve as an early-warning mechanism for service degradation — and, over time, as a foundation for building more sustained user trust in digital financial services.

References

Adji, Y. B., Muhammad, W. A., Akrabi, A. N. L., & Noerlina. (2023). Perkembangan inovasi fintech di Indonesia. *Business Economic, Communication, and Social Sciences Journal*, *5*(1), 47–58. https://doi.org/10.21512/becossjournal.v5i1.8675

Algamar, M. D., & Ismail, N. (2023). Data subject access request: What Indonesia can learn and operationalise in 2024? *Journal of Central Banking Law and Institutions*, *2*(3), 481–512. https://doi.org/10.21098/jcli.v2i3.171

Alzoubi, S., Aldiabat, K., Al-Diabat, M., & Abualigah, L. (2024). An extensive analysis of several methods for classifying unbalanced datasets. *Journal of Autonomous Intelligence*, *7*(3), 1–9. https://doi.org/10.32629/jai.v7i3.966

Amrie, J. H. W., Kurniawan, S., & Amrie, Y. R. S. (2022). Analysis of Google Play Store's sentiment review on Indonesia's P2P fintech platform. *Proceedings of the IEEE Delhi Section Conference (DELCON)*, 1–5. https://doi.org/10.1109/DELCon54057.2022.9753108

Assyifa, D. S., & Luthfiarta, A. (2024). SMOTE-Tomek re-sampling based on random forest method to overcome unbalanced data for multi-class classification. *Jurnal Ilmiah Teknologi Informasi dan Komunikasi*, *9*(2), 151–160. https://doi.org/10.25139/inform.v9i2.8410

Attar, R. W., Almusharraf, A., Alfawaz, A., & Hajli, N. (2022). New trends in e-commerce research: Linking social commerce and sharing commerce — a systematic literature review. *Sustainability*, *14*(23), Article 16024. https://doi.org/10.3390/su142316024

Feriyanto, Qur'anisa, Z., Herawati, M., Lisvi, & Putri, M. H. (2024). Peran fintech dalam meningkatkan inklusi keuangan di era ekonomi digital. *Gemilang Jurnal Manajemen dan Akuntansi*, *4*(3), 99–114. https://doi.org/10.63200/jebmass.v3i4.204

Fullah, A., Rahayu, S., Pratama, D., & Santoso, B. (2025). Analisis sentimen isu penyalahgunaan data pada layanan pinjaman online menggunakan support vector machine di platform X. *Jurnal Informatika dan Teknik Elektro Terapan*, *13*(3s1), 1038–1047. https://journal.eng.unila.ac.id/index.php/jitet/article/view/7976

Gandasari, M., Hidayat, R. R., & Siswajanthy, F. (2025). Legal theory mengawasi fintech lending sebagai instrumen ekonomi digital. *Indonesian Journal of Islamic Jurisprudence, Economic and Legal Theory*, *3*(1), 399–408. https://mail.shariajournal.com/index.php/ijijel/article/view/941/530

Helmi, M., Alharthi, S., & Habib, S. (2024). Online trust determinants, consumer perception, and purchase intent in Saudi e-commerce: Exploring determinants and evidence. *Humanities and Management Sciences — Scientific Journal of King Faisal University*, 100–107. https://doi.org/10.37575/h/mng/240001

Hesniati, H., & Limgestu, R. (2023). Determinants of intention to use Islamic fintech during Covid-19 pandemic. *Ekuitas: Jurnal Ekonomi dan Keuangan*, *7*(4), 587–604. https://doi.org/10.24034/j25485024.y2023.v7.i4.5860

Kwon, Y., Lee, J. H., & Owens, J. (2023). *Managing fintech risks*. Asian Development Bank. https://doi.org/10.22617/brf230170-2

Laínez, N., & Gardner, J. (2023). Algorithmic credit scoring in Vietnam: A legal proposal for maximizing benefits and minimizing risks. *Asian Journal of Law and Society*, *10*(3), 401–432. https://doi.org/10.1017/als.2023.6

Mehrban, S., Nadeem, M. W., Hussain, M., Ahmed, M. M., Hakeem, O., Saqib, S., Kiah, M. L. M., Abbas, F., Hassan, M., & Khan, M. A. (2020). Towards secure FinTech: A survey, taxonomy, and open research challenges. *IEEE Access*, *8*, 23391–23406. https://doi.org/10.1109/access.2020.2970430

Mongkito, L. O. M. H. A. S., Ransi, N., Surimi, L., Tenriawaru, A., Gunawan, G., & Rauf, B. W. (2024). Analisis sentimen aplikasi peminjaman online berdasarkan ulasan pada Play Store menggunakan metode Naïve Bayes dan support vector machine (studi kasus: Adakami dan EasyCash). *Anoatik Jurnal Teknologi Informasi dan Komputer*, *2*(2), 121–128. https://doi.org/10.33772/anoatik.v2i2.71

Noveandini, R., Wulandari, M. S., & Rasyad, F. (2025). Penerapan model LSTM pada analisis sentimen ulasan pengguna aplikasi Shopee Google Play Store. *Fasilkom*, *15*(2), 290–296. https://ejurnal.umri.ac.id/index.php/jik/article/view/9150

Riska, Sulubara, S. M., & Nurkhalisah. (2025). Analisis hukum peer to peer lending pada platform Shopee Paylater: Perspektif kontrak elektronik dan perlindungan konsumen. *Jurnal Ilmu Sosial dan Ilmu Politik*, *32*(3). Tahta Media Group. https://tahtamedia.co.id/index.php/issj/article/view/1555/1547

Shah, A., & Patel, D. K. K. N. (2025). A comparative study of machine learning and deep learning techniques for multilingual text classification. *International Journal of Applied Mathematics*, *38*(8s), 1278–1283. https://ijamjournal.org/ijam/publication/index.php/ijam/article/view/640/589

Sulistyowati, T., & Husda, N. E. (2023). The trust factor: A comprehensive review of antecedents and their role in shaping online purchase intentions. *Jurnal Ekonomi dan Bisnis Airlangga*, *33*(2), 229–244. https://doi.org/10.20473/jeba.v33i22023.229-244

Tritto, A., He, Y., & Junaedi, V. A. (2020). Governing the gold rush into emerging markets: A case study of Indonesia's regulatory responses to the expansion of Chinese-backed online P2P lending. *Financial Innovation*, *6*(1). https://doi.org/10.1186/s40854-020-00202-4

Winarni, Hindasyah, A., & Sirait, T. S. (2025). Analisis sentimen pada pengguna aplikasi Dana menggunakan metode LSTM dan BERT untuk meningkatkan pengguna aplikasi Dana. *Jurnal Ilmiah Pendidikan Dasar*, *10*(4), 240–250. https://ejurnal.umri.ac.id/index.php/jik/article/view/9150

Wulandari, H. A., Astuti, R. P., & Barokah, M. (2025). Peran teknologi finansial (fintech) dalam meningkatkan efisiensi layanan keuangan di Indonesia. *Jurnal Penelitian Nusantara*, *1*(5), 113–120. https://doi.org/10.59435/menulis.v1i5.240.