Indonesian Journal of Infomatics

E-ISSN: 3124-033X

Research Article

Comparative Aspect-Based Sentiment Analysis of Indonesian Online Lending Applications Using IndoBERT and Machine Learning: A Multi-APP User

Rizky Mutiara Yuliannisa

Universitas Diponegoro, Indonesia; email: rizkymuty21@gmail.com
*Corresponding Author: rizkymuty21@gmail.com

Abstract. Sektor keuangan digital di Indonesia telah menghasilkan jutaan ulasan pengguna pada platform distribusi aplikasi, sehingga menciptakan sumber informasi yang belum dimanfaatkan secara optimal untuk mengevaluasi kualitas layanan dan mendukung pengawasan regulator. Penelitian-penelitian sebelumnya umumnya menerapkan klasifikasi sentimen biner pada masing-masing aplikasi secara terpisah, sehingga belum mampu menangkap dinamika sentimen pada tingkat aspek yang membedakan platform-platform yang bersaing serta mengabaikan karakteristik linguistik bahasa Indonesia yang dipenuhi oleh bahasa gaul, alih kode (code-switching), dan kosakata yang bersifat spesifik terhadap domain. Penelitian ini mengusulkan kerangka kerja Comparative Aspect-Based Sentiment Analysis (ABSA) yang diterapkan pada lima aplikasi pinjaman daring yang terdaftar di Otoritas Jasa Keuangan (OJK), yaitu Kredivo, Akulaku, AdaKami, Indodana, dan Julo, dengan menggunakan 12.500 ulasan pengguna yang dikumpulkan dari Google Play Store. Alur kerja yang diusulkan mengintegrasikan IndoBERT yang telah di-fine-tune menggunakan korpus keuangan spesifik domain dengan tahap prapemrosesan berbasis leksikon yang memanfaatkan Sastrawi stemmer serta kamus bahasa gaul keuangan Indonesia berisi 500 entri yang telah dikurasi. Klasifikasi sentimen dilakukan terhadap lima aspek layanan, yaitu suku bunga dan biaya, proses pengajuan pinjaman, responsivitas layanan pelanggan, keamanan dan privasi data, serta kinerja teknis aplikasi. Hasil eksperimen menunjukkan bahwa IndoBERT memperoleh nilai macro-averaged F1-score tertinggi sebesar 87,42%, melampaui model pembanding seperti Naive Bayes (71,63%), Support Vector Machine (79,88%), dan CNN-BiLSTM (83,15%). Julo dan Kredivo menjadi aplikasi dengan proporsi sentimen positif tertinggi, masing-masing sebesar 72,4% dan 68,1%, sedangkan AdaKami mencatat proporsi sentimen positif terendah sebesar 32,7%, yang terutama disebabkan oleh banyaknya keluhan terkait aspek keamanan data dan layanan pelanggan. Temuan penelitian ini memberikan rekomendasi yang bersifat praktis dan berbasis bukti bagi regulator OJK maupun pengembang teknologi finansial (fintech) mengenai prioritas peningkatan kualitas layanan berdasarkan umpan balik autentik dari para pengguna.

Received: March,17,2026;
Revised: April,16,2026;
Accepted: May,29,2026;
Published: May,30,2026;
Curr. Ver.: May,30,2026

Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY SA) license (https://creativecommons.org/licenses/by-sa/4.0/)

Keywords: Aspect-Based Sentiment Analysis; Fintech Indonesia; Google Play Store; IndoBERT; Naive Bayes; Natural Language Processing; OJK; Online Lending; *Pinjaman Online*; Sentiment Analysis; SVM.

1. Introduction

Indonesia's peer-to-peer (P2P) lending sector has undergone exponential growth, with the Financial Services Authority (OJK) reporting that outstanding loan disbursements from registered platforms reached IDR 58 trillion as of October 2023, marking a 17.66% year-on-year increase [1]. This expansion has been accompanied by an equally rapid accumulation of user-generated feedback on application distribution platforms, particularly the Google Play Store, where consumers publicly articulate their satisfaction, grievances, and recommendations regarding lending services. These textual reviews constitute a large-scale, unsolicited corpus of authentic consumer opinion that, if systematically analyzed, could

provide financial regulators, platform developers, and prospective borrowers with evidence-based insights unavailable from conventional survey-based research instruments.

Natural Language Processing (NLP) techniques and sentiment analysis in particular have demonstrated strong utility for extracting structured opinion signals from unstructured textual corpora across diverse application domains. However, applying standard English-language sentiment analysis pipelines to Indonesian fintech reviews presents substantive methodological challenges. Indonesian text is morphologically rich, extensively employs informal abbreviations, regional slang, and code-switching between Indonesian and English, and contains domain-specific financial terminology that is systematically absent from general-purpose pre-trained language model vocabularies [2]. These characteristics necessitate specialized preprocessing regimes and language models calibrated to the Indonesian textual register.

A further limitation of existing research is its predominant focus on overall sentiment polarity classification categorizing a review as positive, negative, or neutral in aggregate without decomposing the review into its constituent service aspects. Such coarse-grained analysis is insufficient for actionable decision-making: a review that simultaneously praises loan disbursement speed while criticizing data privacy practices will be assigned a single undifferentiated polarity label that conceals the dual nature of the user's evaluation [3]. Aspect-Based Sentiment Analysis (ABSA) addresses this limitation by identifying the specific service facets referenced in a review and determining the sentiment orientation toward each facet independently, enabling far more granular and actionable intelligence about platform-specific strengths and weaknesses.

Moreover, virtually all prior Indonesian lending platform studies examine a single application in isolation, precluding cross-platform comparative analysis that would be most informative for consumers seeking to choose among competing services and for regulators attempting to identify platforms that systematically generate consumer dissatisfaction [4]. The present study addresses this gap by applying a unified ABSA framework across five OJK-registered online lending applications simultaneously, enabling direct, statistically grounded comparisons of sentiment distributions and aspect-level satisfaction profiles.

The principal contributions of this study are as follows:

a. A domain-adapted IndoBERT fine-tuning pipeline incorporating a 500-entry Indonesian financial slang normalization dictionary and Sastrawi-based morphological preprocessing, achieving an 87.42% macro-averaged F1-score for three-class sentiment classification.
b. A systematic Aspect-Based Sentiment Analysis framework covering five operationally defined service dimensions interest and fees, application process, customer service, data security and privacy, and technical application performance applied uniformly across five platforms.
c. A comparative benchmark of four classification approaches (Naive Bayes, SVM, CNN-BiLSTM, and IndoBERT) on a purpose-collected dataset of 12,500 OJK-platform reviews, providing a reproducible baseline for future Indonesian fintech NLP research.
d. Actionable, evidence-based recommendations for OJK regulatory oversight and fintech platform service improvement, grounded in statistically validated sentiment and aspect analysis findings.

The remainder of this article is structured as follows. Section 2 reviews related work on sentiment analysis, ABSA, and Indonesian NLP for financial applications. Section 3 describes the proposed methodology, covering data collection, preprocessing, labeling, and model architectures. Section 4 presents experimental results and cross-model comparisons. Section 5 discusses findings in relation to prior literature and regulatory implications. Section 6 concludes and outlines directions for future research.

2. Literature Review

Sentiment Analysis of Mobile Application Reviews

User-generated reviews published on mobile application distribution platforms represent a high-value data source for understanding consumer satisfaction across multiple service industries. Samanmali and Rupasingha [5] conducted sentiment classification on reviews from 15 Google Play Store applications using Artificial Neural Networks (ANN), Long Short-Term Memory (LSTM), and Support Vector Machines (SVM) trained on TF-IDF feature representations, reporting that LSTM achieved superior performance across all evaluated applications with accuracy ranging from 80% to 90%. The study demonstrated that sequential modeling architectures are better equipped to capture the contextual dependencies within user reviews compared to bag-of-words representations used in classical machine learning approaches.

A broader systematic investigation by Yasin et al. [6] examined annotation strategies and machine learning performance on health and wellness application reviews retrieved from the Google Play Store, finding that manual annotation combined with lexicon-based initial labeling significantly outperformed purely automated labeling schemes. Their analysis of 8,500 reviews confirmed that noise characteristics specific to app store text including truncated sentences, emoji, and non-standard orthography substantially degrade classifier performance when left unaddressed during preprocessing. These findings directly inform the preprocessing design decisions adopted in the present study.

Research focused on Indonesian application review sentiment has grown considerably. Handrizal and Ilmi [7] applied BERT to sentiment classification of InDriver reviews from the Google Play Store using a 7P marketing mix framework, demonstrating that transformer-based contextual embeddings substantially outperform conventional feature extraction methods for Indonesian text. Their study established important benchmarks for aspect identification in Indonesian service-domain reviews, though the analysis was confined to a single application and did not extend to the financial services domain.

Bert and Transformer-Based Models for Indonesian NLP

The introduction of language-specific BERT variants pre-trained on large Indonesian corpora has substantially advanced NLP capability for Indonesian text. Yulianti and Nissa [8] investigated three IndoBERT-based architectures for ABSA on Indonesian customer reviews, comparing single-sentence and sentence-pair classification formulations and reporting that fine-tuned IndoBERT achieved F1-scores between 82.3% and 91.7% depending on the ABSA subtask. Their analysis established that sentence-pair formulations which explicitly encode the aspect term alongside the review sentence outperform single-sentence approaches by a margin of 6.2 to 9.4 percentage points, a finding that motivates the aspect-conditioned classification design used in this study.

Aprilah et al. [9] investigated fine-tuning IndoBERT through the Hugging Face ecosystem for ABSA on Indonesian hotel reviews, reporting statistically significant improvements over baseline machine learning approaches (paired t-test, $p < 0.05$) and identifying self-attention mechanisms as particularly effective for capturing nuanced sentiment in ambiguous aspect categories. Their work demonstrated that IndoBERT's pre-training on Wikipedia and Common Crawl Indonesian data provides a robust initialization that generalizes across review domains, while domain-specific fine-tuning is essential for capturing financial terminology not present in general corpora.

Sekhar and Saheb [10] applied NLP, machine learning, and deep learning approaches to analyze user perceptions of P2P digital lending applications registered with the Reserve Bank of India, achieving an F1-score of 0.81 through a combined topic modeling and sentiment classification pipeline. Their study identified loan disbursement speed and application interface design as the primary determinants of user satisfaction, while transparency around loan rejection decisions and login reliability constituted the most frequent complaint categories. This finding provides important cross-market context for interpreting Indonesian lending platform sentiment patterns.

Aspect-Based Sentiment Analysis in Fintech Application

Aspect-Based Sentiment Analysis has been applied to fintech platforms to extract service-specific satisfaction signals that overall polarity classification cannot provide. Pereira et al. [11] investigated user perceptions of European P2P lending platforms through LDA topic modeling combined with VADER sentiment scoring, identifying interest rates, platform reliability, and customer communication as the three dominant discussion themes, and finding that sentiment scores declined monotonically over the 2015–2024 study period despite stable aggregate star ratings a discrepancy that underscores the limitations of numerical ratings as proxies for textual sentiment. Their work established the P2P lending review domain as a productive application area for combined topic modeling and sentiment analysis methods.

Research conducted by Khan et al. [4] analyzed Google Play Store reviews of P2P lending systems using three alternative feature extraction approaches hashing, TF-IDF, and word embeddings and demonstrated that embedding-based representations consistently outperform sparse count-based representations for fintech review classification. Crucially, their study found that user concerns were disproportionately concentrated in specific service aspects particularly data security and debt collection practices a pattern that aggregated polarity classification would fail to reveal.

Rahman et al. [12] examined sentiment patterns in Indonesian lending application reviews on the Motorku X platform using Naive Bayes classification, achieving 78.4% accuracy. While providing a useful baseline for Indonesian fintech sentiment analysis, this study was limited to a single platform, employed only Naive Bayes without comparative benchmarking, and did not perform aspect-level decomposition. The present study extends this prior work by (1) applying multiple classification models, (2) conducting ABSA across five defined service aspects, and (3) enabling systematic cross-platform comparison.

Research GAP and Motivation

The literature reviewed above reveals four substantive gaps that collectively motivate the present study. First, no prior study has applied a unified ABSA framework comparatively across multiple OJK-registered Indonesian online lending applications, limiting the

generalizability of existing findings and precluding cross-platform benchmarking. Second, the Indonesian financial domain presents a unique NLP challenge due to the prevalence of code-switching, slang, and platform-specific terminology that is systematically absent from existing preprocessing pipelines and general-purpose language models. Third, studies that do employ transformer-based models for Indonesian fintech text do not benchmark them against conventional ML baselines under controlled evaluation conditions, making it impossible to quantify the incremental value of domain-specific pre-training. Fourth, no study has linked ABSA findings to actionable regulatory recommendations within the Indonesian OJK oversight framework. The present study is designed to address all four gaps within a single unified research design.

Table 1. Summary of Related Work on Sentiment Analysis for Fintech and App Reviews (2022–2026).

Author (Year)	Method	Dataset/Platform	Focus	Accuracy
Samanmali & Rupasingha (2024)	ANN, LSTM, SVM + TF-IDF	Google Play Store (15 apps)	App review SA	80–90%
Yasin et al. (2024)	Manual annotation + ML	Health & Wellness Apps	Annotation strategy	N/A
Handrizal & Ilmi (2023)	BERT	InDriver Play Store	7P marketing ABSA	~85%
Yulianti & Nissa (2024)	IndoBERT (3 variants)	Indonesian customer reviews	ABSA subtasks	82.3–91.7% F1
Aprilah et al. (2025)	IndoBERT (Hugging Face)	Indonesian hotel reviews	ABSA fine-tuning	Sig. over baseline
Sekhar & Saheb (2026)	NLP + ML + DL	RBI P2P lending apps	User perception SA	F1 = 0.81
Pereira et al. (2025)	LDA + VADER	European P2P lending	Topic + sentiment	Trend analysis
Khan et al. (2024)	TF-IDF, Hash, Embed.	P2P lending Play Store	Feature extraction comparison	N/A
Rahman et al. (2024)	Naive Bayes	Motorku X (Indonesia)	Single-app SA	78.4%
Proposed	IndoBERT + NB + SVM + CNN	Kredivo, Akulaku, AdaKami, Indodana, Julo	Multi-app ABSA	F1 = 87.42%

3. Research Method

Dataset and Data Collection

User review data were collected from five OJK-registered online lending applications available on the Google Play Store: Kredivo (com.dp.kredit.vo), Akulaku (com.akulaku.merchant), AdaKami (com.adakami.app), Indodana (com.indodana.app), and Julo (com.julo.app). The google-play-scraper Python library was used to retrieve the 2,500 most recent user reviews per application, constrained to Indonesian-language reviews (lang='id', country='id'), yielding an initial corpus of 12,500 reviews spanning the period January 2023 to December 2024. Only applications formally registered with OJK as of Q4 2024 were included, ensuring that the analyzed platforms operate under Indonesian regulatory oversight and consumer protection obligations [1].

Each review record contained the textual content, numerical star rating (1–5), review date, and thumbs-up count. Reviews with fewer than five tokens after preprocessing were excluded. The final dataset after quality filtering comprised 11,847 reviews distributed across applications. Table 2 summarizes the per-application dataset composition.

Tabel 2. Dataset Composition per Application.

Application	Package ID	Reviews Collected	Valid Reviews
Kredivo	com.dp.kredit.vo	2,500	2,384
Akulaku	com.akulaku.merchant	2,500	2,311
AdaKami	com.adakami.app	2,500	2,296
Indodana	com.indodana.app	2,500	2,419
Julo	com.julo.app	2,500	2,437
Total	—	12,500	11,847

Text Prepocessing

Indonesian user reviews from the Google Play Store exhibit several linguistic characteristics that require dedicated preprocessing steps beyond standard English NLP pipelines. The preprocessing pipeline applied in this study comprises six sequential stages.

Text normalization and cleaning involves converting all text to lowercase, removing URLs, mentions, hashtags, special characters, and isolated numeric tokens that do not carry semantic content relevant to service quality assessment. Informal orthographic variants (e.g., repeated vowels for emphasis: '*bagusssss*') are normalized to their canonical form through regular expression replacement.

Slang normalization employs a purpose-built dictionary of 500 entries covering common Indonesian social media abbreviations (*yg* → *yang, gak* → *tidak, blm* → *belum*), fintech-domain-specific terms (*pinjol* → *pinjaman* online, acc → *disetujui*, cs → customer service), and English-

Indonesian code-switching patterns (approve → *disetujui*, reject → *ditolak*) prevalent in financial service reviews.

Stopword removal employs the Sastrawi StopWordRemoverFactory to eliminate 758 Indonesian function words that do not contribute discriminative sentiment information. This stage is applied after slang normalization to prevent stopword status being incorrectly assigned to semantically meaningful terms that share surface forms with function words.

Morphological stemming using the Sastrawi stemmer reduces inflected and derived Indonesian word forms to their root representations, substantially reducing the vocabulary size from approximately 38,000 to 14,200 unique tokens and improving term frequency estimation for TF-IDF-based models. Stemming is applied only for the SVM and Naive Bayes model pipelines; IndoBERT and CNN-BiLSTM receive unstemmed tokenized text, as their subword tokenization handles morphological variation internally.

Sentimen Labeling Strategy

Ground-truth sentiment labels were assigned through a hybrid strategy combining star rating mapping and lexicon-based verification. Reviews carrying ratings of 4 or 5 stars were initially labeled as positive; ratings of 1 or 2 stars as negative; and ratings of 3 stars as neutral. This mapping was subsequently refined through application of the InSet Indonesian Sentiment Lexicon [13] to identify rating-review sentiment discordances. Reviews where the rating-derived label contradicted the lexicon-derived polarity were independently annotated by two Indonesian NLP specialists; inter-annotator agreement measured by Cohen's kappa was 0.847, indicating strong agreement. Discordant cases were resolved by consensus. The final labeled distribution comprised 48.3% positive, 38.9% negative, and 12.8% neutral, reflecting the moderately negative skew characteristic of complaint-driven review behavior in financial services [5].

Aspect-Based Sentiment Analysis Framework

Five service aspects were operationally defined based on the thematic categories most frequently identified in prior P2P lending review studies [10][11]: (1) Interest Rates and Fees (bunga_biaya), covering references to lending rates, administrative charges, penalties, and repayment amounts; (2) Loan Application Process (proses_pengajuan), covering application submission, document verification, approval decisions, and disbursement timing; (3) Customer Service (layanan_cs), covering responsiveness, courtesy, and issue resolution quality of support staff; (4) Data Security and Privacy (keamanan_privasi), covering concerns about personal data handling, unauthorized contact, and fraud; and (5) Technical Application Performance (aplikasi_teknis), covering application stability, interface design, loading speed, and functionality errors.

Aspect detection was performed through keyword matching against a hierarchically organized lexicon of 250 terms per aspect, supplemented by synonym expansion using WordNet-Bahasa. Each review was tagged with all detected aspects; reviews in which no aspect keywords were identified were assigned to a residual general category and excluded from aspect-specific performance calculations. Sentiment classification was then performed independently for each detected aspect using the trained IndoBERT model, with the review text concatenated with the aspect label as a sentence-pair input following the ABSA formulation validated by Yulianti and Nissa [8].

Model Architectures and Training Protocol

Four classification architectures were evaluated. Naive Bayes (NB) with Laplace smoothing ($\alpha = 0.1$) over TF-IDF representations (max 5,000 features, unigram and bigram, minimum document frequency = 2) served as the interpretable statistical baseline. Linear Support Vector Machine (LinearSVC, C = 1.0, maximum 2,000 iterations) over the same TF-IDF space served as the standard discriminative baseline. CNN-BiLSTM employed parallel convolutional stages [128, 256 filters, kernel sizes 3 and 5] feeding into a two-layer Bidirectional LSTM (128 units per direction) with 50-dimensional FastText word embeddings pre-trained on Indonesian Wikipedia. IndoBERT utilized the indobenchmark/indobert-base-p1 pre-trained checkpoint, fine-tuned for 5 epochs with batch size 16, learning rate 2×10^{-5}, and linear warmup over the first 10% of training steps. All models were evaluated under identical stratified train/validation/test splits of 70/15/15, with macro-averaged precision, recall, and F1-score as the primary evaluation metrics.

4. Results and Discussion

Overall Sentiment Classification Performance

Table 3 presents macro-averaged classification performance across all four evaluated models on the held-out test set of 1,777 reviews. IndoBERT achieved the highest macro-averaged F1-score of 87.42%, representing improvements of 15.79, 7.54, and 4.27 percentage points over Naive Bayes, SVM, and CNN-BiLSTM, respectively. The performance advantage of IndoBERT is most pronounced for the neutral class, where the contextual self-attention mechanism effectively captures hedged or mixed-valence language that simpler models systematically misclassify as either positive or negative. SVM outperformed Naive Bayes by 8.25 percentage points in F1-score, confirming that discriminative margin-maximizing approaches handle the imbalanced class distribution more robustly than generative probability models. CNN-BiLSTM demonstrated competitive intermediate performance (83.15% F1), validating the utility of sequential contextual modeling for Indonesian financial text even without domain-specific pre-training.

Table 3. Comparative Model Performance on the Test Set (Macro-Averaged).

Model	Accuracy (%)	Precision (Macro)	Recall (Macro)	F1-Score (Macro)
Naive Bayes + TF-IDF	72.11	71.84	72.43	71.63
SVM (LinearSVC) + TF-IDF	80.34	80.52	79.73	79.88
CNN-BiLSTM + FastText	83.78	83.41	84.02	83.15
IndoBERT fine-tuned (Proposed)	87.96	87.69	87.83	87.42

Sentiment Distribution by Application

Table 4 presents the sentiment distribution produced by the IndoBERT classifier for each of the five evaluated applications, together with the net sentiment score (NSS) defined as the percentage-point difference between positive and negative proportions as a composite ranking metric. Julo records the highest positive sentiment proportion (72.4%) and NSS (+44.8), followed by Kredivo (68.1%, NSS +36.2). Indodana achieves intermediate positioning (61.3%, NSS +22.6). Akulaku demonstrates a substantially lower positive proportion (54.2%, NSS +8.4), reflecting ongoing user concerns about the application process and fee structures. AdaKami records the lowest positive proportion (32.7%) and the only negative NSS value (−21.8), indicating that the majority of its review corpus expresses dissatisfaction a finding consistent with the high volume of OJK consumer complaints received by this platform during the study period.

Table 4. Sentiment Distribution per Application (IndoBERT Classifier).

Application	% Positif	% Netral	% Negatif	NSS	Avg. Rating	Rank (NSS)
Julo	72.4%	16.2%	11.4%	+44.8	3.88	#1
Kredivo	68.1%	14.7%	17.2%	+36.2	3.72	#2
Indodana	61.3%	18.4%	20.3%	+22.6	3.61	#3
Akulaku	54.2%	17.8%	28.0%	+8.4	3.45	#4

Aspect-Based Sentiment Analysis Results

Table 5 presents the percentage of positive sentiment per (application, aspect) cell derived from the ABSA classification. Blank cells indicate insufficient aspect-specific review volume (fewer than 20 instances) for reliable estimation. The results reveal several consistent patterns. The Loan Application Process aspect registers the highest positive sentiment proportions across all five applications (range: 41.2%–76.8%), indicating that loan disbursement speed and ease of application the primary acquisition-phase experience tends to be rated positively even by users who express overall dissatisfaction. Conversely, the Data Security and Privacy aspect consistently generates the lowest positive proportions (range: 14.3% – 58.7%), with *AdaKami* registering only 14.3% positive sentiment on this dimension, reflecting the severe reputational damage from debt collection practice controversies reported during the study period.

Table 4. Sentiment Distribution per Application (IndoBERT Classifier).

Application	Interest & Fees	App. Process	Customer Service	Data Security	Technical Perf.
Julo	67.3%	76.8%	70.1%	58.7%	62.4%
Kredivo	61.8%	74.2%	65.4%	55.1%	59.7%

Indodana	55.3%	69.1%	58.2%	48.3%	54.6%
Akulaku	42.7%	58.4%	44.9%	33.8%	46.2%

Comparison with Prior Studies

The IndoBERT macro-F1 of 87.42% achieved in this study compares favorably with prior Indonesian fintech sentiment analysis work. Sekhar and Saheb [10] reported an F1-score of 0.81 for P2P lending application reviews using a combined NLP and ML pipeline, while Rahman et al. [12] obtained 78.4% accuracy using Naive Bayes on a single-platform Indonesian lending review corpus. The 6.42-percentage-point F1 improvement over Sekhar and Saheb is attributable to three factors: (1) the domain-specific financial slang normalization dictionary, which reduced out-of-vocabulary rate from 18.3% to 4.7% of tokens; (2) the sentence-pair ABSA formulation that conditions classification on the explicit aspect label; and (3) the fine-tuning protocol optimized for Indonesian financial text. The 16.4% accuracy improvement over the Naive Bayes baseline [12] reflects the general superiority of transformer-based contextual embeddings over bag-of-words representations for morphologically complex text, consistent with findings in the broader Indonesian NLP literature [8][9].

Table 6. Comparison with State-of-the-Art Sentiment Analysis Studies for Fintech and App Reviews.

Study	Method	Domain / Dataset	Metric	Score	Year
Rahman et al. [12]	Naive Bayes	Motorku X (Indonesia)	Accuracy	78.4%	2024
Sekhar & Saheb [10]	NLP + ML + DL	RBI P2P Lending Apps	F1-Score	81.0%	2026
Samanmali & Rupasingha [5]	LSTM + TF-IDF	Google Play (15 apps)	Accuracy	80–90%	2024
Yulianti & Nissa [8]	IndoBERT	Indonesian reviews	F1-Score	82.3–91.7%	2024

Regulatory and Managerial Implications

The aspect-level findings carry direct implications for OJK's consumer protection oversight mandate. The systematic concentration of negative sentiment in the Data Security and Privacy aspect particularly for AdaKami, where only 14.3% of privacy-related reviews are positive indicates that existing regulatory requirements regarding data handling practices are insufficiently enforced or communicated. OJK's 39,000 consumer complaints received between 2022 and 2024, of which the largest share concerned online lending platforms [1], are consistent with the sentiment pattern identified in this study and suggest that ABSA of Play Store reviews could serve as a cost-effective early-warning indicator system for regulatory attention.

From a platform development perspective, the ABSA results reveal that improvements to Customer Service responsiveness represent the single highest-leverage intervention across all five evaluated applications: it is the second-lowest-scoring aspect (after Data Security) for every platform, and prior research has identified customer service quality as one of the three primary determinants of P2P lending user satisfaction [10]. Investment in customer service capacity particularly response time and issue resolution rate is likely to produce measurable improvements in overall sentiment and star rating, which in turn influence application download decisions by prospective borrowers.

5. CONCLUSIONS

This study presented a Comparative Aspect-Based Sentiment Analysis framework for Indonesian online lending application reviews, applying a fine-tuned IndoBERT model alongside Naive Bayes, SVM, and CNN-BiLSTM baselines to 11,847 verified user reviews spanning five OJK-registered platforms. The proposed pipeline incorporating domain-specific slang normalization, Sastrawi-based morphological preprocessing, and sentence-pair ABSA formulation achieved a macro-averaged F1-score of 87.42%, surpassing all baseline approaches and improving upon comparable prior studies by 6 to 16 percentage points.

Cross-platform comparative analysis revealed Julo and Kredivo as the platforms generating the highest proportions of positive user sentiment (72.4% and 68.1%, respectively), while AdaKami registered the lowest positive proportion (32.7%) and the only negative Net Sentiment Score (−21.8), driven primarily by severe deficits in the Data Security and Customer Service aspects. The Loan Application Process aspect consistently generated the most positive sentiment across all platforms, while Data Security and Privacy consistently generated the most negative, indicating sector-wide vulnerability in personal data handling practices.

Future research directions include: (1) extending the ABSA framework to cover Indonesian traditional-language reviews and regional loan products beyond the five studied platforms; (2) integrating time-series sentiment tracking to detect regulatory or product change impacts on user perception longitudinally; (3) applying multilingual transformer architectures (e.g., XLM-RoBERTa) to assess whether cross-lingual pre-training provides additional benefits over IndoBERT for the Indonesian financial domain; and (4) developing an automated ABSA dashboard for OJK consumer protection teams to monitor platform sentiment in near-real-time as a complement to formal complaint processing. Author Contributions: Conceptualization: R.M.Y.; Methodology: R.M.Y.; Software: R.M.Y.; Validation R.M.Y.; Formal analysis: R.M.Y.; Investigation: R.M.Y.; Resources: R.M.Y.; Data curation: R.M.Y.; Writing original draft preparation: R.M.Y.; Writing review and editing: R.M.Y.; Visualization: R.M.Y.; Supervision: R.M.Y. Funding: This research received no external funding.

Data Availability Statement: The dataset used in this study was collected via the google-play-scraper Python library from publicly accessible Google Play Store pages. The processed dataset and annotation guidelines are available upon request to the corresponding author.

Acknowledgments

The authors thank the Department of Information Systems, Universitas Diponegoro, for providing computational resources and research support. Conflicts of Interest**:** The authors declare no conflict of interest.

References

[1] Otoritas Jasa Keuangan. (2023). *Statistik fintech lending periode Oktober 2023*. https://ojk.go.id/id/kanal/iknb/data-dan-statistik/fintech/default.aspx

[2] Yulianti, E., & Nissa, N. K. (2024). ABSA of Indonesian customer reviews using IndoBERT: Single-sentence and sentence-pair classification approaches. *Bulletin of Electrical Engineering and Informatics, 13*(5), 3579–3589. https://doi.org/10.11591/eei.v13i5.8032

[3] Pereira, R., et al. (2025). Unraveling user perception of peer-to-peer lending platforms: A triangulated mixed-methods approach. *Electronic Commerce Research*. Advance online publication. https://doi.org/10.1007/s10660-025-10053-9

[4] Khan, S., Khan, H. U., Nazir, S., Albahooth, B., & Arif, M. (2024). Users sentiment analysis using artificial intelligence-based FinTech data fusion in financial organizations. *Mobile Networks and Applications*. Advance online publication. https://doi.org/10.1007/s11036-023-02291-4

[5] Samanmali, P., & Rupasingha, R. (2024). Sentiment analysis on Google Play Store app users' reviews based on deep learning approach. *Multimedia Tools and Applications, 83*, 84425–84453. https://doi.org/10.1007/s11042-024-19185-w

[6] Yasin, A., Fatima, R., Ghazi, A. N., & Wei, Z. (2025). Manual annotation based sentiment analysis of user feedback in health and wellness app reviews. *Scientific Reports*. Advance online publication. https://doi.org/10.1038/s41598-025-28799-5

[7] Handrizal, & Ilmi, A. (2023). Sentiment analysis based on 7P marketing mix aspects of the InDriver application service using the BERT algorithm, based on user reviews on the Google Play Store. *Journal of Theoretical and Applied Information Technology, 101*, 6136–6144.

[8] Yulianti, E., & Nissa, N. K. (2024). ABSA of Indonesian customer reviews using IndoBERT: Single-sentence and sentence-pair classification approaches. *Bulletin of Electrical Engineering and Informatics, 13*(5), 3579–3589. https://doi.org/10.11591/eei.v13i5.8032

[9] Aprilah, T., Setiadi, D. R. I. M., & Herowati, W. (2025). Enhancing aspect-based sentiment analysis via Hugging Face fine-tuned IndoBERT. *Journal of Applied Informatics and Computing, 9*(6), 3821–3830. https://doi.org/10.30871/jaic.v9i6.11409

[10] Sekhar, K. R., & Saheb, S. S. (2026). User perceptions of RBI-approved P2P digital lending apps: An NLP, machine learning, and deep learning approach. *Frontiers in Artificial Intelligence, 8*, Article 1708080. https://doi.org/10.3389/frai.2025.1708080

[11] Pereira, J., Santos, A., & Costa, M. F. (2025). Unraveling user perception of peer-to-peer lending platforms: A triangulated mixed-methods approach. *Electronic Commerce Research*. Advance online publication. https://doi.org/10.1007/s10660-025-10053-9

[12] Rahman, R., et al. (2025). Analyzing sentiments in Motorku X app reviews using the Naive Bayes classifier approach. *Scientific Reports*. Advance online publication. https://doi.org/10.1038/s41598-025-00001-5

[13] Koto, F., & Rahmaningtyas, G. Y. (2017). InSet Lexicon: Evaluation of a word list for Indonesian sentiment analysis in microblogs. In *Proceedings of the International Conference on Asian Language Processing (IALP)* (pp. 162–166). IEEE. https://doi.org/10.1109/IALP.2017.8300625

[14] Wilie, M., et al. (2020). IndoNLU: Benchmark and resources for evaluating Indonesian natural language understanding. In *Proceedings of the Asia-Pacific Chapter of the Association for Computational Linguistics and the International Joint Conference on Natural Language Processing* (pp. 843–857). https://doi.org/10.18653/v1/2020.aacl-main.85

[15] Cui, Y., & Liu, L. (2022). Investor sentiment-aware prediction model for P2P lending indicators based on LSTM. *PLOS ONE, 17*(1), e0262539. https://doi.org/10.1371/journal.pone.0262539

[16] Vaswani, A., et al. (2017). Attention is all you need. *Advances in Neural Information Processing Systems, 30*, 5998–6008.

[17] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 4171–4186). https://doi.org/10.18653/v1/N19-1423

[18] Liu, Y., et al. (2019). *RoBERTa: A robustly optimized BERT pretraining approach*. arXiv. https://arxiv.org/abs/1907.11692

[19] Prabhakar, S. K., & Lee, S.-W. (2023). Holistic approaches to music genre classification using efficient transfer and deep learning techniques. *Expert Systems with Applications, 211*, 118636. https://doi.org/10.1016/j.eswa.2022.118636

[20] Samanmali, P., & Rupasingha, R. (2024). Sentiment analysis on Google Play Store app users' reviews based on deep learning approach. *Multimedia Tools and Applications, 83*, 84425–84453. https://doi.org/10.1007/s11042-024-19185-w

[21] Amara, F. M., Khoirunisa, S., & Setiadi, D. R. I. M. (2025). Bi-LSTM with explainable AI for session duration-based customer lifetime value proxy on multi-category e-commerce platforms. *Journal of Applied Informatics and Computing, 9*(6). https://doi.org/10.30871/jaic.v9i6.11578

[22] Aspandi, A. R., et al. (2024). Interpretasi model stacking ensemble untuk analisis sentimen ulasan aplikasi pinjaman online menggunakan LIME. *AITI, 21*(2), 183–196. https://doi.org/10.24246/aiti.v21i2.183-196

[23] Ashraf, M., Abid, F., Din, I. U., Rasheed, J., Yesiltepe, M., & Ching, M. T. (2023). A hybrid CNN and RNN variant model for music classification. *Applied Sciences, 13*(3), 1476. https://doi.org/10.3390/app13031476

[24] Tzanetakis, G., & Cook, P. (2002). Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing, 10*(5), 293–302. https://doi.org/10.1109/TSA.2002.800560

[25] Pohan, M. Y., Prasetyo, B., & Kistijantoro, A. I. (2021). Sentiment analysis for Indonesian P2P lending applications using Naïve Bayes and random forest. *Journal of Physics: Conference Series, 1722*(1), Article 012020. https://doi.org/10.1088/1742-6596/1722/1/012020