Transformer-based sentiment modeling for identifying cross-country fintech perception gaps

Kayla Zhafira Ardinov, Muhardi Saputra, Riska Yanu Fa'rifah
Information System Study Program, School of Industrial Engineering, Telkom University, Bandung, Indonesia

Article Info

Article history:

Received Feb 16, 2026
Revised Jun 12, 2026
Accepted Jun 27, 2026

Keywords:

Aspect-based sentiment analysis
Cross-country comparison topic modeling
DistilBERT
Fintech
Perception gap
ShopeePay

ABSTRACT

Conventional sentiment analysis lacks granularity for capturing detailed user experiences and cross-country comparative insights in digital finance. This study identifies and maps perception gaps among ShopeePay users in Indonesia and Thailand using a topic-informed sentiment analysis pipeline inspired by aspect-based sentiment analysis (ABSA) principles. Adopting the knowledge discovery in databases (KDD) framework, over 170,000 web-scraped reviews were preprocessed, automatically labeled through pseudo-labeling, and balanced using random oversampling. To mitigate pseudo-label reinforcement, manual validation on 500 reviews per country achieved agreement rates of 98.80% (Indonesia) and 99.20% (Thailand) with Cohen's Kappa above 0.97. The fine-tuned DistilBERT model achieved accuracies of 97.65% (Indonesia) and 98.38% (Thailand), though these figures should be interpreted within the pseudo-labeled evaluation context. Significant perception gaps were revealed: Thai users showed lower satisfaction with transactions (67.6% negative) while Indonesian users were more positive (93.0% positive). Process time received negative dominance in both countries, with Indonesia at 78.2% and Thailand at 50.8% negative. These findings demonstrate that user satisfaction is shaped by local infrastructure and cultural contexts, providing strategic insights for regional fintech development.

Corresponding Author:

Muhardi Saputra
Information System Study Program, School of Industrial Engineering, Telkom University
Bandung, Indonesia
Email: muhardi@telkomuniversity.ac.id

1. INTRODUCTION

The integration of internet technology has transformed financial services within the digital economy [1], entering the financial technology (fintech) 4.0 phase with a focus on full asset digitization and payment infrastructure strengthening [2]. These innovations span electronic payment systems to digital wallets (e-wallets) [3], supported by responsive regulatory policies and increased financial literacy [4]. In Southeast Asia, Indonesia and Thailand lead fintech investment growth [5], driving more equitable financial inclusion across ASEAN [6] through fintech-banking collaboration [7]. E-wallets have become essential tools with 80% adoption rates in both markets [8]–[10], and ShopeePay has emerged as the market leader in e-commerce transactions [11]. Despite this leadership, ShopeePay faces persistent user satisfaction challenges [11], including account verification constraints and cybersecurity threats [12]. The pressure of information technology use often triggers technostress, hindering service adoption [13], [14] and manifesting as negative reviews containing technical complaints and privacy concerns [15]. Aspect-based sentiment analysis (ABSA) offers a relevant framework for understanding these perceptions by measuring sentiment on specific review aspects [16], [17]. Unlike conventional sentiment analysis, which assigns a single polarity to entire documents [18], [19], ABSA evaluates sentiment based on aspect categories and opinion terms [16], [20], with advances driven by transformer-based models and topic modeling such as Latent Dirichlet Allocation (LDA) [21]–[23].

However, canonical ABSA as defined in semantic evaluation (SemEval) shared tasks involves joint extraction of aspect terms and polarities at the sentence level. Many applied studies instead adopt a relaxed operationalization where LDA-derived topics serve as proxy aspects and sentiment is classified at the document level [24], [25]. This study follows the latter approach. While we use the term “ABSA” to situate our work within the broader literature, we explicitly acknowledge that our implementation does not perform fine-grained aspect-opinion pair extraction; instead, LDA-derived topics function as service-level aspects with sentiment classification applied per review within each topic cluster. Most ABSA studies remain restricted to single-country or single-language datasets, overlooking regional comparative nuances [26].

Cross-country ABSA research in fintech is particularly scarce, leaving a significant gap: existing studies cannot explain why the same digital service receives fundamentally different perceptions across markets. This gap is critical in Southeast Asia, where rapid fintech adoption coexists with diverse regulatory environments, infrastructure maturity levels, and socio-cultural norms. Indonesia and Thailand were selected for three reasons: (1) both represent the largest e-wallet markets in Southeast Asia by transaction volume [5], [9], [10]; (2) they operate under different regulatory frameworks Indonesia’s Bank Indonesia licensing versus Thailand’s Bank of Thailand sandbox model; and (3) prior literature has focused on each market independently [12], [13] without systematic cross-country aspect-level comparison.

This study addresses this gap through a unified topic-informed sentiment pipeline inspired by ABSA principles, combining LDA for automatic topic extraction as proxy service aspects [27] with a fine-tuned DistilBERT model for computational efficiency [28], [29]. This knowledge distillation mechanism captures the unique characteristics of e-wallet reviews [24], [30]. The framework’s importance lies in uncovering granular perception gaps that generic sentiment metrics overlook, providing a strategic blueprint for navigating regional socio-cultural complexities in fintech. The main contributions are fourfold: (i) a structured cross-country topic-informed sentiment pipeline with explicit distinction from canonical ABSA; (ii) automatic topic extraction using LDA optimized with coherence scores as proxy service aspects; (iii) sentiment classification with pseudo-label reliability validated through manual annotation using Cohen’s Kappa ($\kappa > 0.97$); and (iv) perception gap quantification supported by a documented cross-country topic alignment protocol for strategic insights in Southeast Asian fintech. Section 2 describes the methodology; section 3 presents results and discussion; section 4 presents conclusions.

2. METHOD

2.1. Research framework overview

This study adopts the knowledge discovery in databases (KDD) framework, encompassing systematic stages from data acquisition to knowledge interpretation. The KDD approach ensures structured information extraction through selection, preprocessing, transformation, data mining, and evaluation stages [31]. The framework is iterative, allowing researchers to return to previous stages if evaluation results do not meet quality standards. The entire research flow is shown in Figure 1.

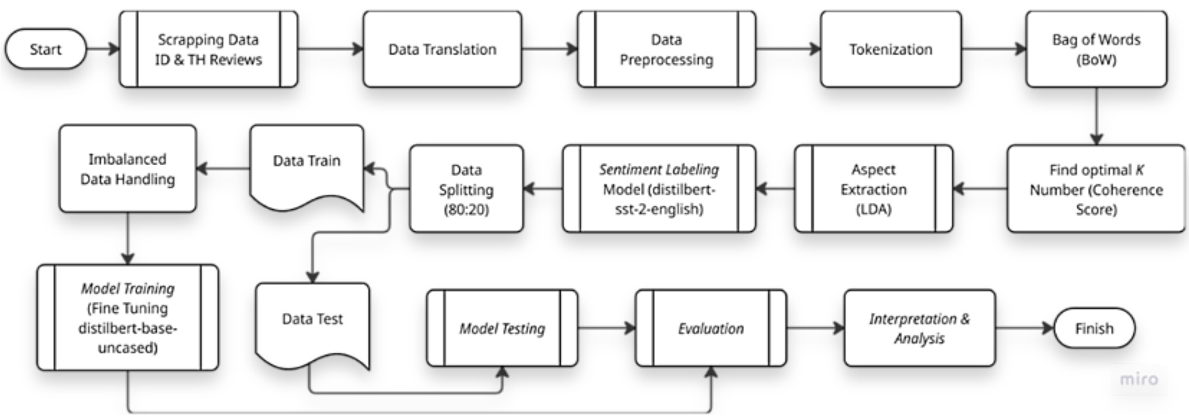

Figure 1. The proposed research framework for cross-country sentiment analysis of fintech services using KDD and LDA methodology

2.2. Data collection

The data collection process used web scraping in Python with the Google Play Scraper library to extract data from ShopeePay applications in Indonesia and Thailand. The data collected on October 20, 2025, yielded a massive volume of reviews, ranging from 80,000 to 90,000 per country. The data attributes collected included reviewId, userName, content, score, and review date (at). However, this study focused its analysis on the content attribute due to potential bias in numerical ratings, which often do not align with users' textual expressions [32]. All raw reviews were then saved in XLSX format for processing in the preprocessing stage.

2.3. Translation and text preprocessing

All reviews were translated into English to leverage the DistilBERT-base-uncased model, which has been extensively pre-trained on high-quality English corpora. This cross-language approach allows analysis on various target languages without requiring massive labeled data [33]. Text cleaning involved case-folding, removing empty values, HTML entities, symbols, emojis, and numbers. Dataset statistics are presented in Table 1, and cleaning examples in Table 2.

Table 1. Summary of raw and cleaned ShopeePay review datasets for Indonesia and Thailand

Country	Raw data	Cleaned data
Indonesia	91234	89311
Thailand	84100	82260

Table 2. Examples of translated reviews and corresponding cleaned text after preprocessing steps

Country	Translated reviews	Cleaned reviews
Indonesia	Cheap items..discounts to satisfy customers	cheap items discounts to satisfy customers
	I'm still having trouble verifying	im still having trouble verifying
Thailand	Excellent. This app is the best.	excellent this app is the best
	Good products, everything available,	good products everything available affordable

2.4. Tokenization and lemmatization

Text was lemmatized using the spaCy library, and standard English stopwords were removed to reduce semantic noise [34]. The final result of this stage is a collection of final tokens ready to be used as input for forming a topic modeling corpus, as presented in Table 3.

Table 3. Sample lemmatized review representations used as input for topic modeling

Lemmatized reviews
['cheap', 'item', 'discount', 'satisfy', 'customer']
['im', 'still', 'having', 'trouble', 'verifying]
['excellent', 'app', 'good']
['good', 'product', 'available', 'affordable', 'price', 'app']

2.5. Aspect extraction

Clean text was converted to numerical format using Bag-of-Words (BoW). Topic extraction was performed using LDA to identify main review themes [35]. LDA-extracted topics are treated as proxy service aspects rather than fine-grained aspect terms as defined in canonical ABSA frameworks (e.g., SemEval tasks), following applied research that employs unsupervised topic models to approximate aspect categories when labeled aspect-level data is unavailable [24], [25].

The optimal number of topics (K) was determined from the peak of the topic coherence score graph. The LDA procedure produces topic distributions per word and per document, enabling structured identification of public opinion patterns [27]. The Indonesian dataset achieved highest coherence at K=9, while the Thai dataset's optimal point was K=6 (Figure 2 and Figure 3). This variance reflects differing semantic diversity: Indonesian reviews exhibited broader themes (e.g., loan features, admin fee complaints) requiring higher K, while Thai reviews concentrated on core functionalities like registration and payment processing, requiring fewer topics for semantic consistency. Topics were labeled based on dominant keywords, reflecting key digital wallet service areas including transactions/payments, user experience, promotions/incentives, service processes, and account/security.

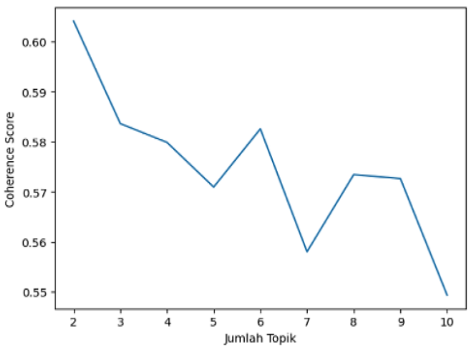

Figure 2. Topic coherence for the Thai dataset. The peak at K=6 indicates optimal semantic consistency for focused and representative aspect identification

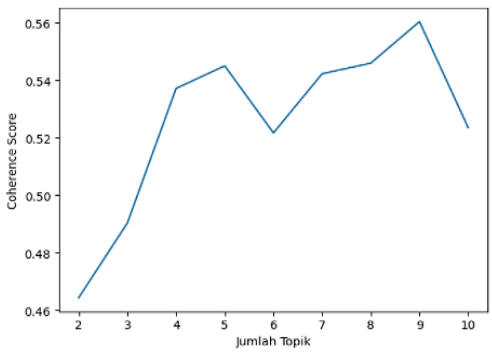

Figure 3. Topic coherence for the Indonesian dataset. The peak at *K*=9 ensures optimal semantic consistency for distinct and representative aspect identification

After determining the optimal number of topics (K), the next step is to annotate each topic generated by the LDA model through the topic labeling stage. This process is done manually by analyzing the most representative keywords (topic keywords) and examining the review examples with the highest probability in each topic to ensure accurate semantic interpretation. The results of aspect extraction for each country are presented in Tables 4 and 5. Based on these results, 9 service aspects were identified for the Indonesian region and 6 for the Thai region.

Table 4. LDA-based topic keywords and manually assigned aspect labels for Indonesian ShopeePay reviews

Topic no.	Topic labeling Indonesia	
	Topic keywords	Labels
0	easy, transaction, fast, payment, simple, practical, safe, process	Transaction and payment
1	use, well, download, app, get, application, balance, coin	App usage
2	great, pay, awesome, shopee, loan, limit, spinjam, paylater, spaylater	Loan feature
3	excellent, like, transfer, admin, free, bank, fee, wallet	Admin fees and bank transfer
4	cool, lot, satisfied, star, shopee, buy, discount, shopeepay, voucher	Promo, discount, and voucher
5	try, want, account, time, bad, application, money, long	Process time
6	helpful, okay, application, slow, smoothly, run, app, bug	App performance (bug/slow)
7	thank, shopeepay, shopee, help, easy, make, hopefully, shopping	General positive feedback
8	good, application, useful, ok, job, app, service	Good app/service

Table 5. LDA-based topic keywords and manually assigned aspect labels for Thailand ShopeePay reviews

Topic no.	Topic labeling Thailand	
	Topic keywords	Labels
0	money, account, bank, transfer, add, baht, game, app, want, deduct	Transaction and payment
1	like, card, number, app, try, enter, verify, identity, say, code	Registration
2	convenient, pay, great, like, bill, shopee, thank, ok, buy, app	Bill payment and merchant
3	use, easy, convenient, fast, work, like, lot, discount, promotion, app	User experience (UX) and promotion
4	time, long, problem, star, bad, take, service, system, wait, contact	Service process time
5	good, app, service, worth, game, mastercard, activity, useful, exchange	General positive feedback

For cross-country comparison analysis, this study focuses its evaluation on overlapping aspects in both countries. These aspects include the dimensions of transaction & payment, promotion/voucher, and process time. This alignment is intended to ensure that the perception gap is measured using service parameters that are equivalent and functionally valid across both fintech ecosystems. These selected aspects will then be processed for automatic sentiment labeling before entering the classification phase using the DistilBERT model

Since LDA was conducted separately on the Indonesian and Thai corpora, an aspect alignment process was required to identify comparable service dimensions across both countries. First, Jaccard Similarity was calculated based on the top 10 keywords of each topic, resulting in a maximum similarity score of 0.20. This relatively low score indicates lexical differences arising from the use of different languages and terminologies, suggesting that keyword overlap alone is insufficient to establish topic equivalence. Therefore, a manual alignment process was performed by examining topic keywords, representative reviews, and the functional meaning of each topic. The resulting aspect pairs were subsequently validated through expert judgment by a domain expert to ensure conceptual and functional equivalence across the two fintech ecosystems. Through this process, three overlapping aspects were identified, namely transaction & payment, promotion/voucher, and process time.

2.6. Sentiment pseudo-labeling

Initial sentiment labeling was performed using a pseudo-labeling approach with the pretrained DistilBERT SST-2 model to generate binary positive/negative classes. This stage enabled automatic labeling of large-scale datasets without requiring comprehensive manual annotation. This model was developed through knowledge distillation techniques to transfer knowledge from large BERT models to smaller models without significant loss of accuracy [28].

Although the number of layers was reduced, the Multi-Head Attention block structure, which is crucial for language processing, was retained [36]. The model was trained using a transformer architecture that captures bidirectional language context [23]. In the classification process, the model extracts features from the special classification token [CLS], which are then processed through the Softmax function to generate class probabilities [37]. The following equation defines in (1):

$$P(y_i|x) = \frac{\{e^{\{z_i\}}\}}{\sum_{\{j=1\}}^{\{C\}} e^{\{z_j\}}} \tag{1}$$

where $P\ (yi\ |x)$ represents the likelihood of input x being classified into class yi, z is the model's output logit score for that class, and C is the total number of sentiment classes. The pseudo-labeling results (Table 6) show class imbalance in both datasets: Indonesian positive sentiment at 78.7% (70,288 reviews) versus 21.3% negative (19,023), and Thailand at 72.1% positive (56,190) versus 27.9% negative (21,756).

Table 6. Summary of sentiment polarity distribution across countries highlighting class imbalance

Country	Positive count	Negative count	Positive (%)	Negative (%)
Indonesia	70,288	19,023	78.7	21.3
Thailand	56,190	21,756	72.1	27.9

The significant imbalance in distribution between the positive and negative classes in both countries confirms the presence of data imbalance. If this data is used directly for fine-tuning without special handling, there is a concern that the model will be biased towards the majority class (positive) and perform poorly at recognizing user complaints (negative class). Therefore, these labeling results provide a strong basis for conducting the Imbalanced Data Handling stage using the oversampling technique before starting model training.

To mitigate the risk of circular evaluation, a stratified random sample of 500 reviews per country was independently annotated by the authors. Results (Table 7) show Cohen's Kappa of 0.9760 (Indonesia) and 0.9840 (Thailand), indicating "Almost Perfect" agreement. Disagreements occurred primarily in short or sarcastic reviews with inherently ambiguous polarity. These findings confirm that the pseudo-labels generated by the DistilBERT SST-2 model are exceptionally reliable for training purposes. While the final classification accuracies (97–98%) reported in Section 3 partly reflect the model's alignment with the pseudo-labeler's output, this manual validation provides robust evidence that such alignment is closely synchronized with human emotional interpretation. The results of the manual validation, comparing the automated pseudo-labels against human judgment, are summarized in Table 7.

Table 7. Manual validation results of pseudo-label reliability

Country	Subset size	Agreement count	Agreement rate (%)	Cohen's Kappa (κ)	Agreement level
Indonesia	500	494	98.80	0.9760	Almost perfect
Thailand	500	496	99.20	0.9840	Almost perfect

2.7. Data splitting and class imbalance handling

The dataset was split 80:20 for training and testing, with the fixed split selected due to the scale (170,000+ reviews) and computational cost (~24 hours per fine-tuning cycle). Random oversampling was applied exclusively to the training set; the test set retained its natural imbalanced distribution. Exact data counts are in Table 8.

Table 8. Distribution of positive and negative sentiment samples across training and testing sets

Country	Sentiment	Data split summary		
		Data total	Data train	Data test
Indonesia	Positive	70288	56230	14058
	Negative	19023	15218	3805
Thailand	Positive	56190	44951	11239
	Negative	21756	17405	4351

The labeling results revealed striking class imbalance in both datasets (e.g., Indonesian reviews: 78.7% positive vs. 21.3% negative). To address this, random oversampling was applied specifically to the training set, replicating minority class (negative) samples to equalize distribution. This method was selected over synthetic minority over-sampling technique (SMOTE), which generates synthetic samples through vector-space interpolation that often produces 'hallucinated' semantic meanings absent in real-world language. For a transformer-based model like DistilBERT that relies on bidirectional context, duplicated but authentic reviews preserve natural linguistic structures more effectively than distorted synthetic data. Random oversampling was also preferred over loss-adjustment techniques such as class weighting or focal loss, as it provides a more stable training distribution for fine-tuning large-scale transformers, directly optimizing recall for negative sentiment to ensure high sensitivity in detecting technostress and critical user complaints.

2.8. Fine-tuning DistilBERT for sentiment classification

The DistilBERT-base-uncased model, pre-trained on the SST-2 dataset, was fine-tuned for domain-specific sentiment classification. DistilBERT was chosen over larger models such as BERT-base or RoBERTa due to computational constraints: with over 170,000 reviews and approximately 24 hours per training cycle, a lighter architecture was essential. DistilBERT offers a 40% size reduction and 60% speed increase while retaining ~97% of BERT's performance [28]. Fine-tuning was conducted using the ShopeePay review dataset to adjust the model's language representation to e-wallet service terminology and user complaint patterns. The configuration included a learning rate of 1×10^{-5}, 3 epochs to prevent overfitting while ensuring convergence, and a validation-set-based evaluation strategy to select the best model checkpoint.

2.9. Evaluation methods

The evaluation stage assesses the classification model's reliability and the quality of the features derived [38]. The first component is the evaluation of aspect extraction results using the topic coherence metric. This metric measures semantic consistency within each topic generated by the LDA model; a higher coherence value indicates that the words in that topic are strongly semantically connected. The optimal number of aspects, K is determined from the peak of the coherence score graph before a significant decline. The second component is the evaluation of the fine-tuned sentiment classification model's performance. Testing was conducted on test data measured using a confusion matrix to obtain accuracy, precision, recall, and F1-score values [39]. Given the data balancing efforts in the previous stage, the F1-score is crucial as a more balanced metric for assessing model performance across positive and negative classes [40]. The formula is defined as (2).

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{2}$$

To ensure evaluation transparency: (i) evaluation was conducted exclusively on the unmodified test set (20% split) retaining its natural class distribution; (ii) the reported accuracy reflects alignment with pseudo-labels, not human ground truth, though manual validation on a 500-review subset (section 2.6) provides an external reliability check; and (iii) high classification accuracy alone does not guarantee real-world robustness in this translated, cross-country setting, though the model's reliability is further supported by consistent F1-scores across both sentiment classes and the narrow confidence intervals, as detailed in the section 3. Beyond global metrics, aspect-level evaluation assesses model consistency across defined ShopeePay service aspects. If results meet quality standards, the process proceeds to knowledge interpretation; otherwise, the system iterates back to data mining or transformation stages.

3. RESULTS AND DISCUSSION
3.1. General sentiment distribution

After fine-tuning the DistilBERT-base-uncased model, sentiment classification performance was evaluated on test datasets from both countries. Evaluation metrics included accuracy, precision, recall, and F1-score to assess the model's ability to classify reviews as positive or negative. A summary of the global classification performance for the Indonesian and Thai datasets is presented in Table 9. For the Indonesian dataset, the model achieved an accuracy of 97.65% with strong detection of both negative sentiment (precision = 0.9517, recall = 0.9372) and positive sentiment (precision = 0.9831, recall = 0.9871). The Thailand dataset achieved 98.38% accuracy with well-balanced metrics across both classes (negative F1 = 0.9709, positive F1 = 0.9887).

However, these figures require caveats: the 97–98% accuracy primarily reflects agreement with the pseudo-label teacher model rather than human ground truth, though manual validation (Section 2.6) confirmed ~99% human agreement ($\kappa > 0.97$). Translation to English may also reduce sentiment-bearing nuance, potentially simplifying classification for an English-centric model. Nonetheless, this performance is consistent with similar transformer fine-tuning studies [40], [41], where binary classification typically exceeds 95%. Model performance is further validated through Confusion Matrix analysis (Figure 4 and Figure 5), comparing predicted labels against actual labels on the test data.

Table 9. Global classification performance metrics of the fine-tuned DistilBERT model for Indonesian and Thai datasets

Country	Class	Precision	Recall	F1-score	Support
Indonesia	Negative	0.9517	0.9372	0.9444	3805
	Positive	0.9831	0.9871	0.9851	14058
	Accuracy (global)			0.9765	17863
Thailand	Negative	0.9708	0.971	0.9905	4351
	Positive	0.9888	0.9887	0.9905	11239

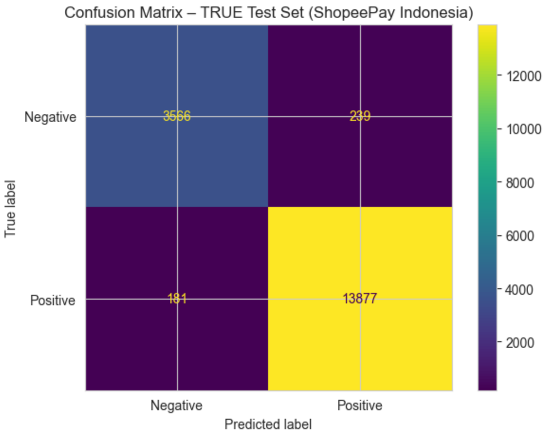

Figure 4. Confusion matrix for the Indonesian dataset. The dominant diagonal values validate the model's reliability in distinguishing sentiment polarities

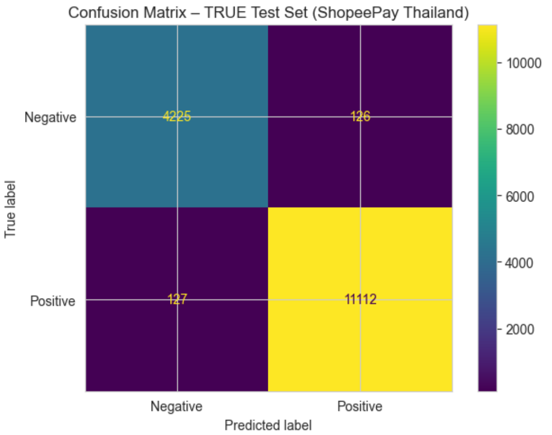

Figure 5. Confusion matrix for the Thai dataset. Minimal off-diagonal values ensure classification accuracy, supporting the validity of the perception gap findings

Beyond global metrics, to fulfill the standards of Aspect-Based Sentiment Analysis (ABSA), classification performance was evaluated at the aspect level. To maintain alignment with the subsequent cross-country comparison, Table 10 presents the detailed classification report specifically for the three overlapping service aspects across both datasets. The model demonstrates high consistency and robustness in both countries. For example, in the "promotion/voucher" aspect, the model achieved exceptional F1-scores of 0.985 (Indonesia) and 0.993 (Thailand). While the "process time" aspect exhibited a slightly lower precision (0.930) in the Indonesian dataset, likely due to the highly nuanced or sarcastic language often employed by users when expressing latency frustrations, the overall aspect-level metrics confirm that the fine-tuned DistilBERT model's predictive capability is highly reliable for identifying the cross-country perception gaps.

Table 10. Aspect-level classification performance on overlapping aspects

Aspect	Country	Samples	Accuracy	Precision	Recall	F1-score
Transaction and payment	ID	1,891	0.977	0.985	0.99	0.987
	TH	2,695	0.988	0.974	0.947	0.961
Promotion/voucher	ID	1,254	0.976	0.983	0.988	0.985
	TH	1,757	0.989	0.992	0.993	0.993
Process time	ID	1,787	0.968	0.93	0.925	0.928
	TH	1,352	0.993	0.989	0.986	0.987

3.2. Cross-country perception gap analysis

The final stage of knowledge interpretation within the KDD framework is to conduct a comparative sentiment analysis on overlapping aspects of service between Indonesia (ID) and Thailand (TH). This analysis aims to identify the perception gap among users regarding the quality of ShopeePay services across two markets. A visualization of the comparison of sentiment proportions using a Grouped 100% stacked bar chart is presented in Figure 6.

In transaction and payment, a significant gap was observed: Indonesian users showed 93.0% positive sentiment, while Thai users expressed only 32.4% positive sentiment (67.6% negative). This indicates notable friction with payment features in Thailand, aligning with research on digital payment friction in developing markets where e-wallet and banking infrastructure interoperability significantly affects satisfaction [42]. Thailand's fragmented payment landscape, despite the national PromptPay system, may create integration friction that Indonesian users operating within a more consolidated e-commerce ecosystem do not experience to the same degree.

In promotion/voucher, both countries showed predominantly positive sentiment, with Thai users at 90.6% and Indonesian users at 80.8% positive. The remaining gap may reflect Indonesia's highly competitive e-wallet ecosystem (GoPay, OVO, DANA, ShopeePay), which generates higher promotional expectations [43].

In process time, Indonesia showed 78.2% negative sentiment compared to Thailand's more balanced split (50.8% negative, 49.2% positive). This may be linked to variable internet connectivity across the Indonesian archipelago and higher server loads during peak shopping events [44]. Overall, these findings provide empirical evidence that user expectations are deeply rooted in local infrastructure and cultural contexts. By bridging the gap between high-level sentiment and specific service attributes, this study offers a data-driven framework for regional fintech providers to refine operational strategies across diverse Southeast Asian ecosystems.

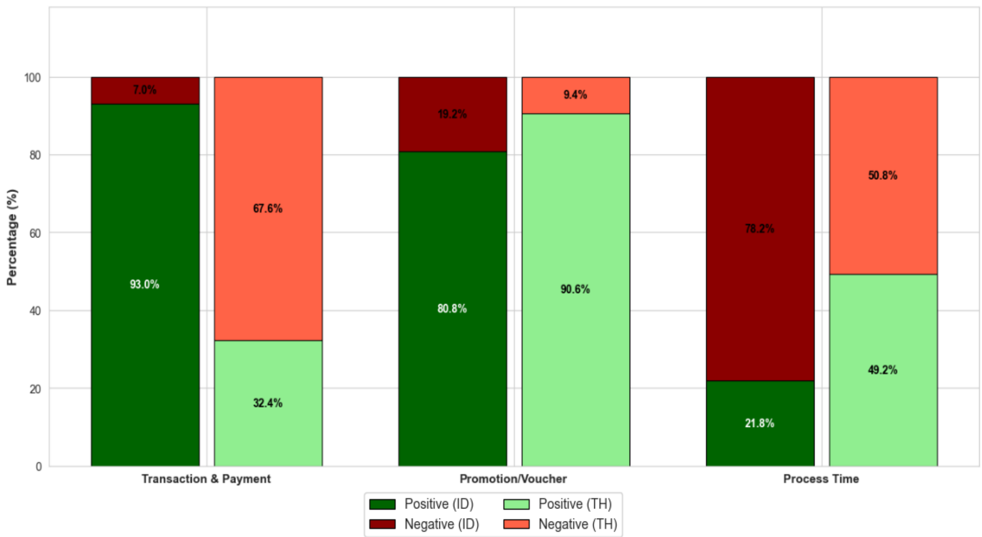

Figure 6. Comparison of positive and negative sentiment proportions on overlapping service aspects between Indonesian and Thai ShopeePay users

3.3. Technical strength of DistilBERT and model evaluation

The DistilBERT model achieved 97–98% accuracy, demonstrating the transformer architecture's effectiveness for informal fintech review language. Generalization was ensured by limiting fine-tuning to three epochs with a conservative learning rate (1×10^{-5}) and selecting checkpoints based on lowest validation loss. Consistent F1-scores across all sentiment classes confirm that random oversampling addressed class imbalances without artificial bias. The narrow confidence intervals in Table 11 further provide strong evidence that the performance is statistically robust and not an artifact of a favorable data split.

Table 11. Statistical performance and robustness metrics

Country	Test sample size (*n*)	Testing accuracy	95% Confidence interval (CI)
Indonesia	17,863	97.65%	[97.43%, 97.87%]
Thailand	15,590	98.38%	[98.18%, 98.58%]

Although *k*-fold cross-validation and ablation studies (e.g., comparing against BERT-base, RoBERTa, or logistic regression) were not conducted due to computational constraints (~24 hours per cycle), the pipeline's reliability is supported by consistent performance across two linguistically distinct datasets and alignment with cross-lingual sentiment analysis standards [33], [41]. Future work should incorporate baseline comparisons to better contextualize the model's contribution.

3.4. Strategic business implications and recommendations

The following are directional hypotheses informed by sentiment patterns, not causal claims. The sentiment data identifies what users are dissatisfied with but cannot establish why without supplementary data.

3.4.1. Optimization of Thailand's payment ecosystem

The transaction and payment gap, 67.6% negative in Thailand versus 7.0% in Indonesia indicates significant transaction experience obstacles in Thailand. The negative sentiment suggests potential friction points in payment or know your customer (KYC) processes, though causal attribution cannot be conclusively established. This reflects technostress, where users perceive technology burden as exceeding their adaptive capacity [45]. Directional recommendation: ShopeePay should investigate whether deeper PromptPay integration and local banking application programming interface (API) connectivity could minimize transaction failures, and whether simplifying the e-KYC journey through automated optical character recognition (OCR) for Thai ID cards may reduce frustration and lower churn rates toward local competitors.

3.4.2. Enhancing voucher competitiveness in Indonesia

In terms of Promotion/Vouchers, both countries show positive sentiment, though Thai users are more satisfied (90.6% positive) than Indonesian users (80.8% positive). Directional recommendation: to bridge the satisfaction gap, ShopeePay Indonesia could consider adopting a 'transparency-first' promotion strategy. This might involve simplifying the terms and conditions (T&C) visible to users and implementing real-time voucher availability notifications. These recommendations are based on the observed sentiment patterns and would benefit from further validation through user surveys or A/B testing before implementation.

3.4.3. Infrastructure scaling and latency management process time

Process time is a critical issue for both countries, but the urgency is higher in Indonesia, where 78.2% of respondents expressed negative sentiment compared to Thailand's 50.8%. Directional recommendation: the technical team should investigate backend server scalability and latency mitigation, specifically for top-up and real-time payment confirmations. The sentiment data alone cannot identify the specific technical root causes, but the volume and intensity of complaints suggest that infrastructure capacity and query performance warrant priority investigation.

3.5. Theoretical implications

Theoretically, this study offers a new perspective on the transactional model of technostress. Our findings suggest that technostress is not a universal issue; rather, it is shaped by regional infrastructure and local cultural contexts. For instance, while Thai users experience higher stress levels regarding transaction reliability, Indonesian users are more affected by service speed and latency. This demonstrates that a user's experience with technology-related stress is deeply tied to their specific digital environment. By comparing these two markets, this study highlights that user behavior theories cannot follow a 'one-size-fits-all' approach, particularly within the diverse Southeast Asian fintech landscape.

3.6. Limitations and future work

This study has several limitations. First, the pipeline employs topic-informed sentiment classification rather than canonical aspect-term-level ABSA, limiting granularity. Second, pseudo-labeling introduces circular evaluation risk, though mitigated by manual validation ($\kappa > 0.97$). Third, English translation may flatten regional nuances such as Indonesian slang or Thai emotional particles. Fourth, semi-manual topic alignment ($\kappa = 0.92$) may not generalize without adaptation. Fifth, no baseline model comparisons were conducted due to computational constraints.

Future research should explore multilingual models (mBERT, XLM-RoBERTa) for direct analysis on original texts, incorporate larger human-annotated evaluation sets as primary benchmarks, and expand to other Southeast Asian markets (Vietnam, Malaysia) to validate generalizability. Additionally, integrating this pipeline into live fintech dashboards and exploring large language model (LLM)-based prompting for simultaneous aspect-polarity extraction [20] represent strategic opportunities.

4. CONCLUSION

This study identified ShopeePay perception gaps between Indonesian and Thai users by applying a KDD framework integrated with fine-tuned DistilBERT, achieving accuracies of 97.65% and 98.38% respectively. These results demonstrate that transformer architecture with random oversampling provides a consistent approach for sentiment classification in digital finance, though reliance on pseudo-labeling and English translation may introduce semantic biases. Manual validation confirmed approximately 99% label agreement (Cohen's Kappa > 0.97), establishing robust reliability. Beyond technical performance, the primary contribution is empirical evidence that Technostress is a localized phenomenon moderated by regional infrastructure maturity and socio-cultural expectations, not a uniform response. By revealing divergent perceptions of the same platform, this study adds a critical comparative dimension to behavioral technology adoption models, demonstrating that single-market evaluations are insufficient for emerging Southeast Asian markets.

ACKNOWLEDGMENTS

The authors would like to express their sincere gratitude to the Information System Study Program, School of Industrial Engineering, Telkom University, for providing the essential facilities, academic environment, and administrative support that made this research possible.

FUNDING INFORMATION

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Kayla Zhafira Ardinov	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓		✓	✓
Muhardi Saputra	✓			✓		✓				✓		✓		✓
Riska Yanu Fa'rifah	✓	✓		✓						✓		✓		✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. Authors state no conflict of interest.

DATA AVAILABILITY

The dataset used in this study consists of user reviews scraped from the Google Play Store platform on October 20, 2025. The raw data and cleaned tokens are archived and can be made available by the corresponding author upon reasonable request for academic purposes.

REFERENCES

[1] A. R. Latiff, M. Z. Alqudah, H. Samara, and N. Alslaibi, "Empowering the financial sector: the role of fintech research development trends," *Future Business Journal*, vol. 11, no. 1, p. 92, Apr. 2025, doi: 10.1186/s43093-025-00512-y.

[2] K. Banerjee and V. V. R. Kumar, "Inclusive fintech: a systematic review and agenda for future research," *Cogent Business & Management*, vol. 12, no. 1, Dec. 2025, doi: 10.1080/23311975.2025.2549511.

[3] L. N. Nasution, R. Rusiadi, and D. P. Nasution, "ASEAN-5 economic analysis based on financial inclusion and financial technology," *Proceeding International Pelita Bangsa*, vol. 1, no. 01, pp. 114–133, 2023, doi: 10.37366/pipb.v1i01.3112.

[4] A. H. Y. Zheng, R. Ab-Rahim, and A. H. Y. Jing, "Examining the fintech ecosystem of ASEAN-6 countries," *Asia-Pacific Social Science Review*, vol. 22, no. 2, pp. 1–13, Jun. 2022, doi: 10.59588/2350-8329.1417.

[5] QED Investors and Boston Consulting Group, "Global fintech 2024: Prudence, profits, and growth." Accessed: Dec. 26, 2025. [Online]. Available: https://www.qedinvestors.com/blog/global-fintech-2024-prudence-profits-and-growth.

[6] D. Fernandez and M. Rakotomalala, "Financial technology and inclusion in ASEAN," LKCSB Research Working Paper, Singapore Management University, 2020.

[7] UOB, PwC Singapore, and Singapore FinTech Association, "FinTech in ASEAN 2025: navigating the new realities." 2025. Accessed: Dec. 26, 2025. [Online]. Available: https://go.uob.com/fintech2025.

[8] M. Yang, A. Al Mamun, M. Mohiuddin, N. C. Nawi, and N. R. Zainol, "Cashless transactions: a study on intention and adoption of e-wallets," *Sustainability*, vol. 13, no. 2, p. 831, Jan. 2021, doi: 10.3390/su13020831.

[9] I. Wafa, "E-wallets become the preferred digital payment method in 2025 (E-Wallet jadi metode pembayaran digital favorit 2025)," GoodStats. Accessed: Jan. 01, 2026. [Online]. Available: https://data.goodstats.id/statistic/e-wallet-jadi-metode-pembayaran-digital-favorit-2025-V6gYH.

[10] Antom, "Thailand's payment market: Things you may not know," Accessed: Jan. 01, 2026. [Online]. Available: https://knowledge.antom.com/thailands-payment-market-things-you-may-not-know.

[11] I. E. Carballo, "Indonesia: 2025 analysis of payments and ecommerce trends," Payments & Commerce Market Intelligence (PCMI). Accessed: Jan. 01, 2026. [Online]. Available: https://paymentscmi.com/insights/indonesia-ecommerce-payments-trends-2025/.

[12] M. I. J. Ugli and M. Sari, "Global trends in mobile payment adoption: A systematic literature review with insights for Indonesia," *Innovation Science and Technology*, vol. 1, no. 5, pp. 57–64, 2025, doi: 10.5281/zenodo.17443591.

[13] Y.-K. Lee, "Impacts of digital technostress and digital technology self-efficacy on fintech usage intention of Chinese Gen Z consumers," *Sustainability*, vol. 13, no. 9, p. 5077, Apr. 2021, doi: 10.3390/su13095077.

[14] A.-M. Cazan *et al.*, "Technostress and time spent online: a cross-cultural comparison for teachers and students," *Frontiers in Psychology*, vol. 15, Jul. 2024, doi: 10.3389/fpsyg.2024.1377200.

[15] U. Kishnani, N. Noah, S. Das, and R. Dewri, "Privacy and security evaluation of mobile payment applications through user-generated reviews," in *Proceedings of the 21st Workshop on Privacy in the Electronic Society*, New York, NY, USA: ACM, Nov. 2022, pp. 159–173. doi: 10.1145/3559613.3563196.

[16] I. Piriyakul, S. Kunathikornkit, and R. Piriyakul, "Determining optimal marketing mix configurations: an integration of opinion-oriented language and necessary condition analysis," *Journal of Vacation Marketing*, Jun. 2025, doi: 10.1177/13567667251351025.

[17] Y. Chilukuri *et al.*, "An aspect-based sentiment classification method using the Pachinko allocation model," *Engineering, Technology & Applied Science Research*, vol. 15, no. 4, pp. 25844–25850, Aug. 2025, doi: 10.48084/etasr.11555.

[18] S. Karmaker, "Exploring the psychological factors of consumer purchase behavior in Bangladesh on the adoption of digital payment apps," M.S. thesis, Department of Business Administration, United International University, 2024.

[19] S. M. Mohammad, "Chapter 11 - Sentiment analysis: automatically detecting valence, emotions, and other affectual states from text," in *Emotion Measurement (Second Edition)*, Second Edi., H. L. Meiselman, Ed., Woodhead Publishing, 2021, pp. 323–379. doi: https://doi.org/10.1016/B978-0-12-821124-3.00011-9.

[20] S. Baral, "Using large language models for aspect based sentiment analysis," M.S thesis, Economics and Business Administration (Business Data Science), Aalborg University, Aalborg, Denmark, 2024.

[21] M. Wankhade, A. C. S. Rao, and C. Kulkarni, "A survey on sentiment analysis methods, applications, and challenges," *Artificial Intelligence Review*, vol. 55, no. 7, pp. 5731–5780, Oct. 2022, doi: 10.1007/s10462-022-10144-1.

[22] M. S. Jahan, H. U. Khan, S. Akbar, M. U. Farooq, S. Gul, and A. Amjad, "Bidirectional language modeling: a systematic literature review," *Scientific Programming*, vol. 2021, pp. 1–15, May 2021, doi: 10.1155/2021/6641832.

[23] M. D. Deepa and A. Tamilarasi, "Bidirectional encoder representations from transformers (BERT) language model for sentiment analysis task: Review," *Turkish Journal of Computer and Mathematics Education*, vol. 12, no. 7, pp. 1708–1721, 2021.

[24] A. N. K. Kambayo, S. S. Berutu, J. Jatmika, and A. Nshimiyimana, "Aspect-based sentiment analysis using latent Dirichlet allocation (LDA) and DistilBERT on Threads app reviews," *Infact: International Journal of Computers*, vol. 9, no. 01, pp. 25–34, Mar. 2025, doi: 10.61179/infact.v9i01.707.

[25] S. M. Sofi and A. Selamat, "Aspect based sentiment analysis: feature extraction using latent dirichlet allocation (LDA) and term frequency-inverse document frequency (TF-IDF) in machine learning (ML)," *Malaysian Journal of Information and Communication Technology (MyJICT)*, vol. 8, no. 2, pp. 169–179, Dec. 2023, doi: 10.53840/myjict8-2-102.

[26] J. Šmíd, "Cross-lingual aspect-based sentiment analysis (Mezijazyčná aspektově orientovaná analýza sentimentu)," M.S. thesis, University of West Bohemia, 2023.

[27] P. Madzík and L. Falát, "State-of-the-art on analytic hierarchy process in the last 40 years: literature review based on latent dirichlet allocation topic modelling," *PLOS ONE*, vol. 17, no. 5, p. e0268777, May 2022, doi: 10.1371/journal.pone.0268777.

[28] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter," *arXiv preprint arXiv: 1910.01108*, 2020, [Online]. Available: http://arxiv.org/abs/1910.01108.

[29] M. Jojoa, P. Eftekhar, B. Nowrouzi-Kia, and B. Garcia-Zapirain, "Natural language processing analysis applied to COVID-19 open-text opinions using a distilBERT model for sentiment categorization," *AI & SOCIETY*, vol. 39, no. 3, pp. 883–890, Jun. 2024, doi: 10.1007/s00146-022-01594-w.

[30] C. Lu *et al.*, "Knowledge distillation of transformer-based language models revisited," *arXiv preprint arXiv: 2206.14366*, 2022, [Online]. Available: http://arxiv.org/abs/2206.14366.

[31] Sonam and Jyoti, "A narrative review of data mining techniques for user behaviour recognition with illustrative application of the Apriori algorithm," *Journal of Computers, Mechanical and Management*, vol. 4, no. 2, 2025, doi: 10.57159/jcmm.4.2.25214.

[32] T. Aljrees *et al.*, "Contradiction in text review and apps rating: Prediction using textual features and transfer learning," *PeerJ Computer Science*, vol. 10, p. e1722, Jan. 2024, doi: 10.7717/peerj-cs.1722.

[33] M. S. U. Miah, M. M. Kabir, T. Bin Sarwar, M. Safran, S. Alfarhood, and M. F. Mridha, "A multimodal approach to cross-lingual sentiment analysis with ensemble of transformer and LLM," *Scientific Reports*, vol. 14, no. 1, p. 9603, Apr. 2024, doi: 10.1038/s41598-024-60210-7.
[34] N. Khamphakdee and P. Seresangtakul, "An efficient deep learning for Thai sentiment analysis," *Data*, vol. 8, no. 5, p. 90, May 2023, doi: 10.3390/data8050090.
[35] H. Nach, "An analysis of the news coverage of fintech in Africa: a natural language processing approach," *SSRN Electronic Journal*, 2024, doi: 10.2139/ssrn.4823123.
[36] J. Li, Y. Zhu, and K. Sun, "A novel iteration scheme with conjugate gradient for faster pruning on transformer models," *Complex & Intelligent Systems*, vol. 10, no. 6, pp. 7863–7875, Dec. 2024, doi: 10.1007/s40747-024-01595-w.
[37] Y. Boreshban, S. M. Mirbostani, G. Ghassem-Sani, S. A. Mirroshandel, and S. Amiriparian, "Improving question answering performance using knowledge distillation and active learning," *Engineering Applications of Artificial Intelligence*, vol. 123, p. 106137, Aug. 2023, doi: 10.1016/j.engappai.2023.106137.
[38] G. Menghani, "Efficient deep learning: a survey on making deep learning models smaller, faster, and better," *ACM Computing Surveys*, vol. 55, no. 12, pp. 1–37, Dec. 2023, doi: 10.1145/3578938.
[39] A. Dankar, A. Jassani, and K. Kumar, "Improving knowledge distillation for BERT models: loss functions, mapping methods, and weight tuning," *arXiv preprint arXiv: 2308.13958*, 2023, [Online]. Available: https://arxiv.org/pdf/2308.13958.
[40] R. I. Perwira, V. A. Permadi, D. I. Purnamasari, and R. P. Agusdin, "Domain-specific fine-tuning of IndoBERT for aspect-based sentiment analysis in Indonesian travel user-generated content," *Journal of Information Systems Engineering and Business Intelligence*, vol. 11, no. 1, pp. 30–40, Mar. 2025, doi: 10.20473/jisebi.11.1.30-40.
[41] C. Zhou *et al.*, "A comprehensive evaluation of large language models on aspect-based sentiment analysis," *arXiv preprint arXiv: 2412.02279*, 2024, [Online]. Available: http://arxiv.org/abs/2412.02279.
[42] P. Kurniawan and D. Achjari, "Dynamic cross-border payment preferences: a qualitative study of Indonesian expatriates in Thailand and Malaysia," *Banks and Bank Systems*, vol. 19, no. 2, pp. 115–125, May 2024, doi: 10.21511/bbs.19(2).2024.09.
[43] R. M. Azhar and D. Azmawati, "Indonesia's efforts in supporting ASEAN digital economy by expanding the use of QRIS in Southeast Asia," *SHS Web of Conferences*, vol. 204, p. 02004, Nov. 2024, doi: 10.1051/shsconf/202420402004.
[44] A. Rachman, N. Julianti, and S. Arkoyah, "Challenges and opportunities for QRIS implementation as a digital payment system in Indonesia," *EkBis: Jurnal Ekonomi dan Bisnis*, vol. 8, no. 1, pp. 1–13, Jun. 2024, doi: 10.14421/EkBis.2024.8.1.2134.
[45] M. Salo, H. Pirkkalainen, C. E. H. Chua, and T. Koskelainen, "Formation and mitigation of technostress in the personal use of IT," *MIS Quarterly*, vol. 46, no. 2, pp. 1073–1108, Jun. 2022, doi: 10.25300/MISQ/2022/14950.

BIOGRAPHIES OF AUTHORS

Kayla Zhafira Ardinov is currently a final-year information systems student at the School of Industrial Engineering, Telkom University, having commenced her studies in 2022. She is a BNSP-certified Associate Data Scientist with a specialized interest in data analytics, particularly in the fields of data engineering, automation, and fraud detection. Throughout her academic journey, she has been involved in various projects that utilize advanced analytical techniques. Her final undergraduate research focuses on an empirical analysis of technostress among Indonesian e-wallet users, employing a structured aspect-based sentiment analysis (ABSA) pipeline with fine-tuned DistilBERT and LDA modeling. She can be contacted at email: kaylazhafira@student.telkomuniversity.ac.id.

Muhardi Saputra is an assistant professor and lecturer at Telkom University's Information Systems Department and a Doctoral Candidate in IT at Universitas Gadjah Mada. A recipient of LPDP (2015) and BPI (2022) scholarships, he has over 10 years of experience as an IT Consultant at Sharing Vision Indonesia, specializing in electronic payments, banking tech, and digital business. He is an active member of ITB's SCCIC, a BNSP-certified Deputy ICT Project Manager, and a Certified Data Science Manager (CDSM). His research focuses on digital business perspectives influencing technology acceptance. He can be contacted at email: muhardi@telkomuniversity.ac.id.

Riska Yanu Fa'rifah is an assistant professor (Lektor) and researcher in the information systems study program at Telkom University. She holds a B.S. (S.Si.) and M.S. (M.Si.) in Statistics from ITS and is a certified Big Data Analyst. As an active member of the Digital Enterprise and Technology expertise group, her research focuses on computational statistics, data analytics, ABSA, machine learning for fintech, and health informatics. She has led initiatives regarding e-wallet technostress and predictive modeling for medical data. Her pedagogical duties include teaching data mining, discrete mathematics, and probability and statistics, while also serving as an academic advisor and lab coordinator. She can be contacted at email: riskayanu@telkomuniversity.ac.id.