International Proceeding
Universitas Tulungagung
2026

CONSUMER EXPERIENCE AND E-SERVICE QUALITY IN DIGITAL PAYMENTS: LATENT DIRICHLET ALLOCATION TOPIC MODELING OF DANA E-WALLET REVIEWS

Farida Islamiah[1]*, Muhamad Ikbal Thola[2]

[1]Makassar State University, Indonesian
[2]National Institute of Public Administration, Indonesian
*farida.islamiah@unm.ac.id

Abstract

The DANA digital wallet serves hundreds of millions of accounts in Indonesia, so its service quality affects a large volume of daily transactions. This study aims to identify the main topics in DANA user reviews on Google Play, interpret them through the E-S-QUAL and E-RecS-QUAL dimensions of electronic service quality, and test whether ratings differ across topics. The data consist of 4,000 Indonesian-language reviews posted between April and October 2026. After near-identical template reviews were filtered out and the text was preprocessed in Indonesian (slang normalization, stopword removal, and Sastrawi stemming), 3,979 reviews were modeled with Latent Dirichlet Allocation in the Gensim library. The number of topics was selected using *c_v* and *u_mass* coherence and topic diversity across three random seeds. The selected model contains ten topics with a *c_v* coherence of 0.417. The largest topic is transaction and system disruptions (24.3%), followed by the DANA Cicil installment feature (15.7%) and deducted balances and refunds (13.3%). The customer service and chatbot topic has the lowest mean rating (1.71), whereas DANA Cicil has the highest (3.01). Ratings differ significantly across topics (Kruskal-Wallis $H = 242.58$, $p < .001$), with a weak-to-moderate association (Cramér's $V = .181$). Most reviews relate to the system availability and fulfillment dimensions, while service recovery problems receive the most negative ratings. These findings suggest that digital wallet providers should prioritize system stability, clear refund processes, and escalation paths from chatbots to human agents.

Keywords: Digital wallet; user reviews; topic modeling; Latent Dirichlet Allocation; electronic service quality

INTRODUCTION

Digital payments in Indonesia are growing rapidly, and digital wallets (e-wallets) have become one of the main channels for cashless transactions. DANA is one of the largest providers. According to the company, DANA reached 170 million users in 2023, up 23 percent from the previous year. In the same period, QRIS transactions grew by 272 percent and the *Kirim Uang* (money transfer) feature grew by 147 percent (CGS International Sekuritas, 2024). The company's sustainability report later recorded about 240 million DANA accounts, an average of 58 million transactions per day, and around 2.6 million micro, small, and medium enterprises (MSMEs) registered as merchant partners (DANA, 2025). On the user side, a Jakpat survey of 1,945 digital wallet users in November 2025 ranked DANA as the most widely used

platform among Generation Z (75%), millennials (70%), and Generation X (68%) (Karimani, 2026).

At this scale, service quality concerns tens of millions of transactions every day. Disruptions such as failed transactions, balances that have not been credited, or account verification problems can affect many users at once, and some of them report these experiences in Google Play reviews. App reviews are user-generated content: they are unsolicited, abundant, and record post-use experiences directly. Reviews have therefore been widely used to identify the determinants of satisfaction with digital financial services (Çallı, 2023; Kar, 2021; Oh & Kim, 2022; Taufik et al., 2025). Reviews also complement company-reported data with the users' perspective. For example, DANA reports zero material data breach incidents (DANA, 2025), whereas users' perceptions of security can only be learned from the users themselves.

The large volume of reviews and their informal language make manual analysis impractical. Topic modeling with Latent Dirichlet Allocation (LDA) is the most widely used approach for extracting latent themes from user reviews, particularly for digital banking apps (Mustaqim & Suryono, 2025). In Indonesia, LDA has been applied to OVO user comments (Widiantoro & Harnadi, 2022, 2023), to a comparison of OVO, DANA, LinkAja, and Doku (Widiantoro et al., 2023), and to studies of technostress among digital wallet users (M. Saputra et al., 2023; Sumidra et al., 2026). Other studies have compared LDA with BERTopic (Islamiyah et al., 2026), used BERTopic as the main method (Hanifah et al., 2025; S. N. A. Zahra et al., 2025), or used LDA as the basis for aspect-based sentiment analysis (Aprilia et al., 2025).

These studies share three limitations that serve as the starting point of the present research. First, some studies determine the number of topics through repeated trials (Widiantoro & Harnadi, 2023) or a single coherence measure, and they rarely report how sensitive the results are to the model's random initialization. Yet standard coherence scores can be misleading for short texts (Weisser et al., 2023). Comparisons between models are also inconsistent: BERTopic yields higher coherence than LDA on digital wallet reviews (Islamiyah et al., 2026) but lower coherence on mobile banking reviews (Halim et al., 2026). Second, topics are often mapped onto conceptual frameworks without validation. Aprilia et al. (2025) showed that aspect labels derived automatically from LDA keywords matched manual annotations in only 11.46% of cases. Third, Indonesian studies generally stop at listing topics and recommending app improvements. N. J. Saputra & Safitri (2025) did use topics as features to predict GoPay user satisfaction, but for predictive purposes. Meanwhile, international studies have already tested the relationship between review topics and app ratings and performance (Oh & Kim, 2022; Yi et al., 2024).

Based on these limitations, this study has three objectives: (1) to identify the main topics in DANA user reviews on Google Play using Gensim-based LDA with Indonesian-language preprocessing; (2) to interpret these topics through the electronic service quality dimensions of E-S-QUAL and E-RecS-QUAL (Parasuraman et al., 2005); and (3) to test whether rating distributions differ across topics. The number of topics used in this study is 10. The novelty of the study lies in a transparent modeling protocol for informal Indonesian-language reviews that covers three elements: filtering near-identical template reviews, selecting the number of topics

based on several measures and several random seeds, and statistically testing the relationship between topics and ratings. Theoretically, the study extends the application of E-S-QUAL to digital wallet services using review data. Practically, it provides a list of service improvement priorities for digital wallet providers. Among the ten review topics, transaction and system disruptions form the largest topic, whereas customer service is the most negatively rated topic. These findings indicate that improvement priorities lie in system stability and in service recovery when transactions fail.

LITERATURE REVIEW

Electronic Service Quality in Digital Wallets

Parasuraman et al. (2005) developed E-S-QUAL to measure the quality of services delivered through electronic channels. The scale consists of four core dimensions: efficiency (the ease and speed of accessing and using the service), system availability (the correct technical functioning of the system), fulfillment (the extent to which service promises are kept), and privacy (the security and protection of customer information). For situations in which customers encounter problems, they developed E-RecS-QUAL with three service recovery dimensions: responsiveness (effective handling of problems), compensation (redress for the problems experienced), and contact (the availability of assistance by telephone or online representatives).

This framework is relevant to digital wallets because their core services, namely payments, transfers, and balance top-ups, are delivered entirely through the app. Handling problems such as balances that are not credited or funds that are held up is also directly related to the recovery dimensions. S. Zahra & Alamsyah (2022) classified 53,968 DANA reviews into these seven dimensions. They found positive sentiment above 69% in all dimensions except compensation, which was below 42%. Nurcahyo et al. (2025) applied SERVQUAL to Twitter conversations and found that DANA excelled in the economic benefit dimension, whereas LinkAja excelled in responsiveness. Both studies set the dimensions a priori through classification, so themes outside the predefined framework are difficult to detect. Topic modeling takes the opposite direction: themes are derived from the data and then interpreted with a theoretical framework.

Another framework applied to digital wallet reviews in Indonesia is technostress, the psychological strain caused by suboptimal interaction with technology (M. Saputra et al., 2023; S. Zahra & Alamsyah, 2022). Islamiyah et al. (2026) mapped review topics for OVO, DANA, and ShopeePay onto the techno-complexity and techno-uncertainty dimensions. Sumidra et al. (2026) mapped complaints in 100,000 DANA reviews onto four technostress dimensions. For web-based services, the WebQual framework is also widely used. Islamiah et al. (2022) found that usability, information quality, and service interaction significantly affected user satisfaction with an academic information system, with service interaction, which covers security and trust, having the strongest effect. This study adopts E-S-QUAL because its dimensions refer directly to service aspects that providers can control, including service recovery when problems occur.

User Reviews as a Data Source on Service Quality

App store reviews provide a large record of user experience. Kar (2021) built the Digital Service Usage Satisfaction Model from Twitter data on mobile payments in India. The results showed that cost, usefulness, trust, social influence, credibility, information privacy, and responsiveness determine usage satisfaction. Oh & Kim (2022) derived quality dimensions from 96,140 reviews of banking apps in the United States using LDA. Using weekly panel data, they showed that complaints about security, customer service, and complexity lowered app ratings. Yi et al. (2024) analyzed 230,940 reviews of nine mobile payment apps and identified four quality factors, namely payment convenience, security, customer service, and app design, that affect market performance through ratings. Raza et al. (2026) derived constructs from 23,830 digital wallet reviews in Saudi Arabia through LDA and expert construct mapping, and then tested them with regularized regression.

Hussain et al. (2023) showed that positive and negative reviews discuss different dimensions. Positive reviews highlight security, convenience, and ease of use, whereas negative reviews highlight system availability, responsiveness, problematic updates, login problems, and reliability. In Indonesia, social media data have also been used to map public perceptions of QRIS (Ulya et al., 2023), the perceived value of OVO customers on Instagram (Adiningtyas & Auliani, 2024), and public responses to tax policy through YouTube comments (Islamiah et al., 2026). Google Play reviews have been used to classify user sentiment toward the sharia fintech app Ammana (Islamiah & Muthahharah, 2023) and to assess the digital service quality of the Bank Jago app (Taufik et al., 2025). Ruslan et al. (2025) extended this work by comparing 36,000 reviews of the apps of three digital banks and one conventional bank. These three studies focus on sentiment classification by comparing machine learning algorithms, and SVM achieved the highest accuracy in two of them. This approach measures the direction of sentiment, whereas identifying the themes that users discuss requires topic modeling. Review data, however, have limitations. Ratings tend to be positively biased (Perea-Khalifi et al., 2024), and reviews contain no demographic information about reviewers, so differences between user groups cannot be tested (Hussain et al., 2023). The results of review analysis should therefore be read as a picture of the experiences of users who choose to speak up, not of the entire user population.

Latent Dirichlet Allocation and Its Implementation in Gensim

LDA is a generative probabilistic model that treats each document as a mixture of topics and each topic as a probability distribution over the vocabulary (Blei et al., 2003). The per-document topic distributions and the per-topic word distributions are given Dirichlet priors with parameters α and η, respectively. A single review can therefore address several issues at once, for example a failed transaction and a slow customer service response. Gensim provides an LDA implementation based on online variational inference that is efficient for large corpora (Řehůřek & Sojka, 2010), and LDA has been widely used in applied research across many fields (Jelodar et al., 2019).

The quality of topic models is generally assessed with coherence. The c_v measure combines a sliding window, normalized pointwise mutual information, and cosine similarity. Among the measures tested by Röder et al. (2015), c_v had the highest correlation with human

judgments. The u_mass measure counts word co-occurrence within documents of the same corpus (Mimno et al., 2011). Topic diversity measures the proportion of unique words among the top 25 words of all topics to detect overlapping topics (Dieng et al., 2020). Nevertheless, a model's statistical fit does not always align with how easily humans can interpret its topics (Chang et al., 2009). For short texts, standard coherence scores can even be poor evaluation metrics (Weisser et al., 2023) (Weißer et al., 2022).

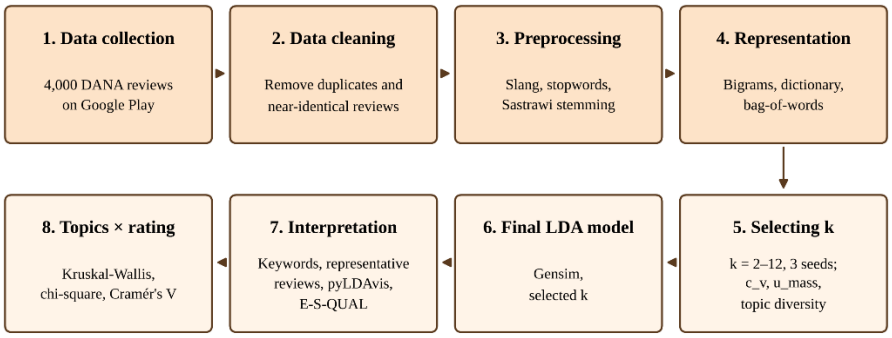
Maier et al. (2018) emphasized four steps that determine the validity and reliability of LDA: appropriate preprocessing, parameter selection including the number of topics, evaluation of model reliability, and valid interpretation of topics. For short texts, LDA remains among the methods that produce meaningful topics (Albalawi et al., 2020), but it should be accompanied by manual inspection of keywords and representative documents. LDAvis visualization supports interpretation by displaying inter-topic distances and the relevance of words to each topic (Sievert & Shirley, 2014).

RESEARCH METHOD

Research Design

This study uses a quantitative descriptive-exploratory design with a text mining approach. The unit of analysis is an individual review of the DANA app. The research stages are shown in Figure 1.

Figure 1. Research flow



Research Data

The data consist of 4,000 Indonesian-language reviews of the DANA app on Google Play, collected with the google-play-scraper library. Each review contains the review text, a star rating of 1-5, the date, the number of likes (thumbs up), and the app version. The reviews were written between April 13 and October 20, 2026, with 95.7% of them posted in July-October 2026, and they cover 36 app versions (2.120.1 to 2.149.1). Reviewer identities were not used in the analysis, and review excerpts in this article are presented without reviewer names.

Data Cleaning and Text Preprocessing

Data cleaning was carried out in two steps. First, empty reviews and exact duplicates were removed. Second, near-identical template reviews were filtered out. An initial inspection revealed five-star reviews with almost identical sentences that differed by only a few words; if left in the data, such reviews could form spurious topics. Reviews of at least 100 characters

were represented with character-level TF-IDF 3-5-grams. From each pair with a cosine similarity of at least 0.85, only one review was retained.

All preprocessing was performed in Indonesian without translation so that the meaning of informal expressions and service-specific terms, such as DANA Cicil or *verifikasi wajah* (face verification), would not be altered by machine translation. The steps were: (1) case folding; (2) removal of URLs, mentions, numbers, punctuation, and emoji; (3) reduction of repeated letters; (4) tokenization; (5) normalization of nonstandard words with an 82-entry slang dictionary; (6) removal of 798 combined stopwords from the Indonesian NLTK list, Sastrawi, and a domain list (greetings, fillers, and app names); (7) stemming with Sastrawi; and (8) a second round of stopword removal on the stemmed words. These steps are consistent with the preprocessing practices of previous studies on Indonesian-language app reviews (Islamiah & Muthahharah, 2023; Widiantoro & Harnadi, 2023). Words that are meaningful in the digital wallet context, such as *masuk* (credited), *hilang* (lost), *kembali* (returned), *gagal* (failed), and *lama* (long), were excluded from the stopword list. Reviews with fewer than three tokens after preprocessing were excluded. Frequently co-occurring phrases were merged into bigrams with Gensim's Phrases. Words appearing in fewer than five reviews or in more than 50% of reviews were removed from the dictionary before the bag-of-words corpus was built.

Topic Modeling and Selection of the Number of Topics

The LDA model was built with Gensim's LdaModel class. The parameters α and η were estimated from the data (auto), allowing the model to assign different prior weights to each topic. The number of topics k was tested over the range of 2 to 12. Each value of k was trained with three random seeds (42, 7, and 2024) to assess the stability of the results. For each model, c_v coherence (Röder et al., 2015), u_mass coherence (Mimno et al., 2011), and topic diversity (Dieng et al., 2020) were calculated. The value of k was chosen based on the highest mean c_v with a small standard deviation, and the resulting topics were then checked for interpretability. The final model was retrained with a larger number of iterations. The modeling parameters are summarized in Table 2.

Table 2. Modeling parameters

Parameter	**Value**
Range of the number of topics (k)	2–12
Random seeds for selecting k	42, 7, 2024
Passes / iterations (selecting k)	10 / 200
Passes / iterations / *chunksize* (final model)	50 / 400 / 2,000
Priors α and η	*auto* (estimated from the data)
Dictionary filter	minimum 5 reviews; maximum 50% of reviews
Bigrams (*Phrases*)	*min_count* = 5; *threshold* = 10
Evaluation measures	c_v, u_mass, topic diversity (top 25 words)
Keywords per topic	15

Source: research configuration

Topic Interpretation and Validation

Each topic was interpreted from its 15 highest-weighted keywords, the five reviews with the highest topic probability, and the pyLDAvis visualization (Sievert & Shirley, 2014). Topic labels were assigned based on the content of the representative reviews and then mapped onto the E-S-QUAL and E-RecS-QUAL dimensions. Topic labeling and mapping to the service quality dimensions were performed by the researchers. To allow readers to check the interpretation, representative reviews are quoted for each topic.

Analysis of the Relationship between Topics and Ratings

Each review was assigned a dominant topic, that is, the topic with the highest probability in that review. The distribution of ratings per topic is presented as percentages and means. Differences in ratings across topics were tested with the Kruskal-Wallis test because ratings are ordinal. The association between topics and rating classes (negative for 1-2 stars, neutral for 3 stars, and positive for 4-5 stars) was tested with the chi-square test, and its strength was measured with Cramér's V. Ratings were treated as a potentially biased proxy for satisfaction (Perea-Khalifi et al., 2024). All analyses were run in Python on Google Colab using the Gensim, Sastrawi, NLTK, scikit-learn, SciPy, and pyLDAvis libraries.

RESULT AND DISCUSSION

Data Profile and Preprocessing Results

Table 3 summarizes the amount of data at each stage. No exact duplicates were found, but 14 near-identical template reviews were filtered out, leaving 3,986 reviews. After preprocessing, seven reviews with fewer than three tokens were excluded, so 3,979 reviews were analyzed. The median review length after preprocessing was 14 tokens, so the data can be classified as short texts. The most frequent stemmed words were *transaksi* (transaction; 1,711 occurrences), *saldo* (balance; 1,246), *uang* (money; 1,087), *bayar* (pay; 899), *cicil* (installment; 891), and *akun* (account; 845). The most frequent bigrams included *fitur_cicil* (installment feature; 196), *layan_langgan* (customer service; 157), *kirim_uang* (send money; 116), *verifikasi_wajah* (face verification; 95), *ganti_nomor* (change number; 74), and *kode_otp* (OTP code; 71).

Table 3. Data profile and preprocessing results

Stage or measure	Value
Initial reviews	4,000
After removing exact duplicates	4,000
After filtering near-identical reviews	3,986
After preprocessing (at least 3 tokens)	3,979
Total tokens after preprocessing	61,139
Median tokens per review	14
Unique vocabulary (before bigrams)	5,061
Vocabulary including bigrams, before / after dictionary filtering	5,258 / 1,316
Star ratings 1 / 2 / 3 / 4 / 5 (number of reviews)	2,280 / 451 / 328 / 225 / 695

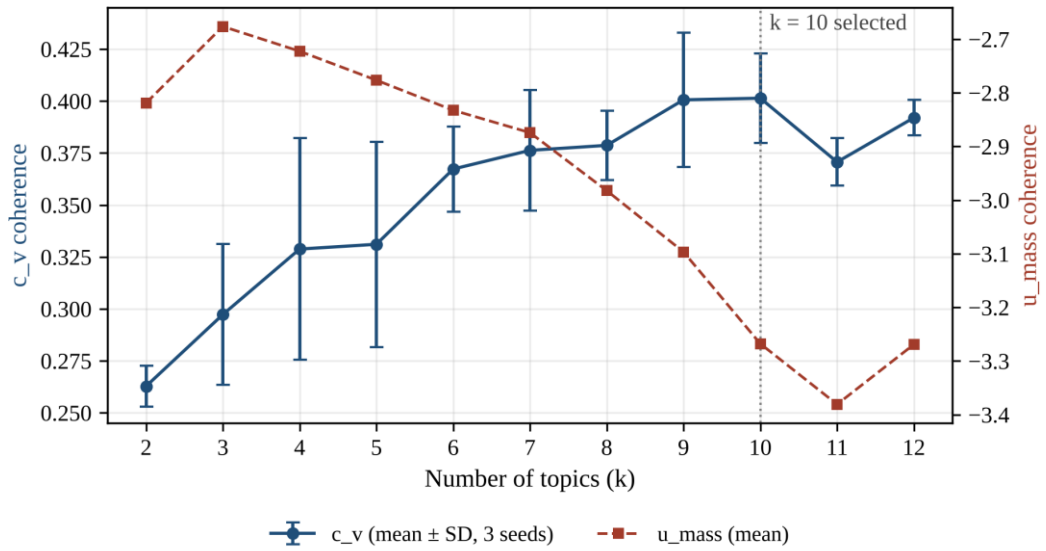
Source: data processing results

The rating distribution is strongly skewed toward negative ratings: 57.30% of reviews gave one star and 11.33% gave two stars, whereas four- and five-star reviews accounted for only 23.1%. The mean rating was 2.15. This pattern is the opposite of the positive bias reported by Perea-Khalifi et al. (2024) for P2P payment apps. The topics found therefore describe the structure of complaints more than the satisfaction level of all DANA users.

Determining the Number of Topics

Figure 2 shows that the mean c_v increased from 0.263 at k = 2 and peaked at k = 9 (0.4006 ± 0.0324) and k = 10 (0.4014 ± 0.0216) before declining at k = 11. The value k = 10 was selected because it had the highest mean c_v with a smaller standard deviation than k = 9, and each of its topics could be clearly interpreted from representative reviews. In contrast, u_mass worsened as k increased, from -2.68 at k = 3 to -3.27 at k = 10. Topic diversity also decreased from 0.873 at k = 2 to 0.563 at k = 10. This pattern indicates a trade-off: adding topics increases the semantic coherence measured by c_v but also increases vocabulary overlap between topics.

Figure 2. c_v and u_mass coherence by number of topics (mean of three random seeds)



The final model, trained with 50 passes, yielded a c_v of 0.4169, a u_mass of -3.4871, a topic diversity of 0.604, and a perplexity of 75.73. Per-topic coherence ranged from 0.313 (T10) to 0.598 (T6). This c_v value is comparable to those reported for LDA on short Indonesian-language texts, for example 0.488 and 0.384 for tweets about QRIS (Ulya et al., 2023). The mean probability of the dominant topic per review ranged from 0.48 to 0.64, indicating that many reviews contain more than one topic.

Topics in DANA User Reviews

Table 4 presents the ten topics, ordered by the proportion of reviews in which each is the dominant topic, together with the electronic service quality dimension that fits each topic best. The review excerpts below are quoted verbatim in Indonesian, including their original spelling, followed by the authors' English translation in square brackets.

Transaction and system disruptions (T1) is the largest topic (24.28%). Reviews in this topic complain about "*sistem sibuk*" [system busy] messages, failed QRIS payments and transfers, pages that keep loading, and disruptions after app updates without prior notice. One user wrote: "*ga ada pemberitahuan gangguan, tiba tiba gangguan ... qris gak bisa transfer juga ga bisa*" [there was no notice of any disruption, suddenly there is a disruption ... QRIS does not work, and transfers do not work either] (1 star). DANA Cicil (T2), DANA's paylater feature, is the second-largest topic (15.73%). This topic has two sides: praise for its ease and speed (keywords *mudah* [easy], *instan* [instant], *bantu* [help], and *bagus* [good]) and complaints about limits that do not appear or activations that fail, for example "*tidak mendapatkan limit pinjaman setelah pelunasan dan saya sudah menunggu 3 bulan lamanya*" [I did not get a loan limit after paying off my installments, and I have waited for three months] (2 stars).

The next two topics both concern balances but with different emphases. Deducted balance and refunds (T3; 13.32%) covers failed transactions in which the balance had already been deducted, transfers reported as successful that never arrived, and slow refunds: "*transfer berhasil, tetapi rekening yang di tuju tidak ada saldo yang masuk*" [the transfer was successful, but no balance arrived in the destination account] (1 star). Missing balance and security (T4; 10.18%) covers transactions that users did not recognize and a sense of insecurity: "*saya baru baru top up selang berapa lama uang saya hilang ada transaksi lain yang mengatas namakan tiktok koin..padahal saya tidak melakukannya*" [I had just topped up, and shortly afterward my money was gone; there was another transaction for TikTok coins, even though I did not make it] (1 star). The proximity of the two topics is visible in the inter-topic distance map (Figure 3), where the T3 and T4 circles overlap.

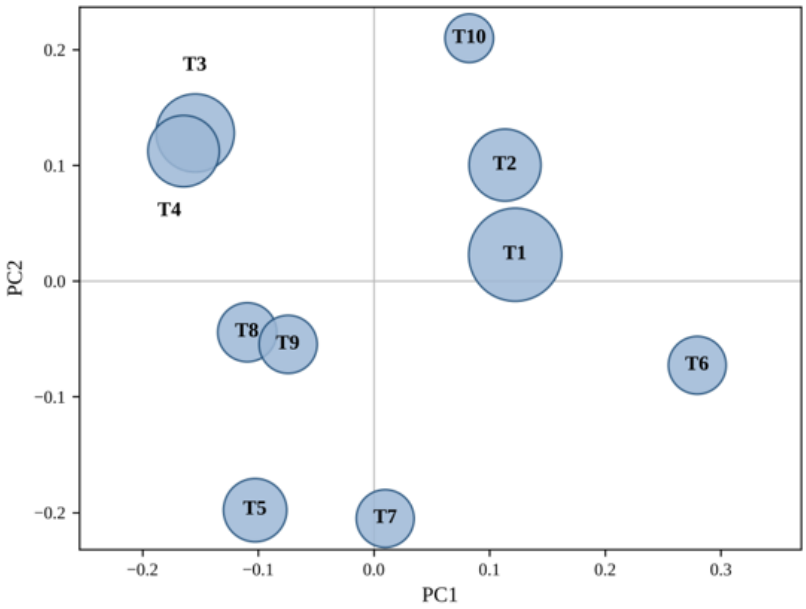
Table 4. Topics in DANA user reviews

Topic	Label	Top 10 keywords	% of reviews (n)	Primary dimension (secondary)
T1	Transaction and system disruptions	transaction, error, transfer, disruption, long, update, people, pay, open, slow	24.28 (966)	System availability (Efficiency)
T2	DANA Cicil: limit and activation	installment, transaction, limit, easy, feature, installment_feature, appear, help, pay, good	15.73 (626)	Fulfillment (Efficiency)
T3	Deducted balance and refunds	balance, credited, deduct, pay, transaction, return, money, top (up), business, success	13.32 (530)	Fulfillment (Compensation)
T4	Missing balance and security	money, lost, safe, balance, transaction, disappointed, new, credited, top (up), times	10.18 (405)	Privacy (Compensation)
T5	Login, number change, and OTP	login, number, enter, account, lost, phone, old, face_verification, new, PIN	8.19 (326)	Efficiency (Privacy)
T6	Events, points, and rewards	reward, event, network, point, open, error, wifi, signal, problem, run out	7.46 (297)	System availability (Fulfillment)

Topic	Label	Top 10 keywords	% of reviews (n)	Primary dimension (secondary)
T7	Premium account upgrade and identity verification	account, premium, register, try, new, long, ID card, data, delete, verification	6.16 (245)	Efficiency (Fulfillment)
T8	Delayed transfers and access to support	money, credited, transfer, problem, number, withdraw, complicated, customer_service, account, data	6.01 (239)	Fulfillment (Contact)
T9	Customer service and chatbot	customer_service, *ana*, *gua*, solution, response, customer, problem, help, resolved, pay	5.50 (219)	Responsiveness (Contact)
T10	Bill payments and purchases	pay, menu, buy, thousand, electricity, bill, automatic, open, display, times	3.17 (126)	Fulfillment (Efficiency)

Source: data processing results. Keywords are English translations of the Indonesian stemmed tokens; underscores mark bigrams. *ana* derives from *Diana*, the name of the customer service virtual assistant mentioned by users; *gua* is slang for 'I' that the normalization dictionary did not capture.

Figure 3. Inter-topic distance map (from pyLDAvis). Circle area is proportional to the topic's token share



Two topics relate to account access and verification. Login, number change, and OTP (T5; 8.19%) is dominated by users who lost or changed their phone numbers while OTP codes were still sent to the old number: "*nomor hp sudah tidak aktif lagi ... selanjutnya di kirim SMS kode verifikasi ke nomor HP lama*" [my phone number is no longer active ... then the verification code is sent by SMS to the old phone number] (1 star). This topic also includes failed face verification. Premium account upgrade and identity verification (T7; 6.16%) covers failed uploads of the national identity card (KTP) or family card and premium status that does not appear: "*disuruh upgrade premium tapi ngga pernah bisa pas di foto KTP selalu kesalahan*

sistem" [I was told to upgrade to premium, but it never works; whenever I photograph my ID card there is a system error] (2 stars).

Events, points, and rewards (T6; 7.46%) has the highest coherence (0.598) and lies apart from the other topics in Figure 3 because of its distinctive vocabulary: *hadiah* (reward), *event*, *point*, and *tukar* (redeem). Complaints arise when point redemption fails during claim hours, with an unstable network or a busy system given as the reason: "*setiap masuk jam klaim hadiah pasti ada kendala alasan jaringan tak stabil lah sistem sibuk*" [every time the reward claim hour starts there is always a problem, with the excuse of an unstable network or a busy system] (1 star). Some users even call the program a scam. Delayed transfers and access to support (T8; 6.01%) combines delays in cross-platform transfers with difficulty in obtaining help: "*uang belum masuk sama sekali ke gopay, udh komplain lewat CS Diana, disuruh nunggu 3 hari*" [the money has not reached GoPay at all; I already complained through CS Diana and was told to wait three days] (1 star). Customer service and chatbot (T9; 5.50%) focuses on responses from customer service, including the virtual assistant Diana, which users perceive as slow and repetitive: "*sangat slow respon terutama untuk AI DIANA nya sangat menyebalkan karna selalu menanyakan pertanyaan yg sama terus menerus*" [very slow response, especially the AI DIANA, which is very annoying because it keeps asking the same questions over and over] (1 star). The smallest topic, bill payments and purchases (T10; 3.17%), covers payments for electricity, phone credit, data packages, and automatic bills. This topic contains complaints about menus that do not work as well as praise, for example for free transfer fees "*10x ke semua bank*" [10 times to all banks] each month (5 stars).

Relationship between Topics and Ratings

Table 5 shows that mean ratings range from 1.71 for customer service and chatbot (T9) to 3.01 for DANA Cicil (T2). The percentage of 1–2-star reviews ranges from 46.6% (T2) to 82.2% (T9). After T9, the most negatively rated topic is events, points, and rewards (T6; mean 1.80; 78.1% with 1-2 stars). Only T2 and T10 have a share of 5-star reviews above 25%, consistent with both topics containing complaints as well as praise.

Table 5. Ratings by dominant topic

Topic	Label	n	Mean	Median	% 1–2 stars	% 5 stars
T1	Transaction and system disruptions	966	2.04	1	69.6	11.0
T2	DANA Cicil: limit and activation	626	3.01	3	46.6	38.2
T3	Deducted balance and refunds	530	1.98	1	73.2	14.3
T4	Missing balance and security	405	1.99	1	73.1	15.6
T5	Login, number change, and OTP	326	1.95	1	74.2	13.8
T6	Events, points, and rewards	297	1.80	1	78.1	11.4
T7	Premium account upgrade and identity verification	245	1.96	1	73.1	12.7
T8	Delayed transfers and access to support	239	2.01	1	73.6	18.0
T9	Customer service and chatbot	219	1.71	1	82.2	11.0

Topic	Label	n	Mean	Median	% 1–2 stars	% 5 stars
T10	Bill payments and purchases	126	2.56	2	58.7	27.0
	All reviews	3,979	2.15	1	68.6	17.5

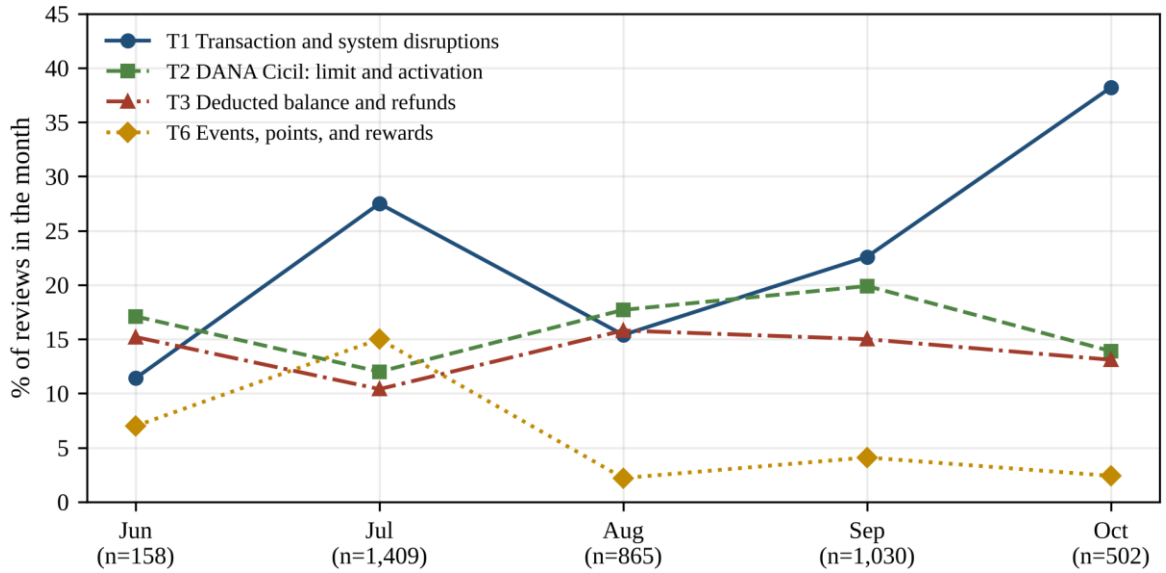
Source: data processing results

The Kruskal-Wallis test showed that rating distributions differed significantly across topics ($H(9) = 242.58$, $p < .001$). The chi-square test between topics and rating classes was also significant ($\chi^2(18) = 261.14$, $p < .001$), with a Cramér's V of .181, indicating a weak-to-moderate association. In other words, the type of problem discussed is related to the rating given, although topics are not the only determinant of ratings.

Monthly Topic Dynamics

Figure 4 shows the monthly proportions of four main topics from June to October 2026. April and May are not shown because they contained only 7 and 8 reviews, respectively. The proportion of transaction and system disruptions (T1) fluctuated: 27.5% in July, falling to 15.4% in August, then rising to 22.6% in September and 38.2% in October. Events, points, and rewards (T6) peaked in July (15.0%) and then remained below 4.2% in the following months. One of its representative reviews mentions an "*event bola*" [football event]. This pattern suggests that some complaints were concentrated in particular disruption or campaign periods. However, the monthly composition is also affected by how the data were collected, so this finding is descriptive.

Figure 4. Monthly proportion of dominant topics for the four main topics



Discussion

When each topic is mapped to its primary dimension (Table 4), 38.23% of reviews relate to fulfillment (T2, T3, T8, T10), 31.74% to system availability (T1, T6), 14.35% to efficiency (T5, T7), 10.18% to privacy (T4), and 5.50% to responsiveness (T9). About 70% of reviews thus relate to two core E-S-QUAL dimensions, fulfillment and system availability (Parasuraman et al., 2005). This finding is consistent with Hussain et al. (2023), who found

that negative mobile banking reviews discuss system availability, problematic updates, login problems, and responsiveness. In Indonesia, Ruslan et al. (2025) also found that login and registration, app updates, and complicated transaction processes were the main complaints about both digital and conventional bank apps. The finding is also in line with previous studies of DANA reviews, which highlighted app performance, failed transactions, and identity verification (Sumidra et al., 2026), as well as network stability, OTP, and face verification for OVO, DANA, and ShopeePay (Islamiyah et al., 2026).

The service recovery dimensions of E-RecS-QUAL received the most negative ratings. The customer service and chatbot topic (T9) has the lowest mean rating, and complaints about the difficulty of getting help also appear in T8 and T3. This result is in line with S. Zahra & Alamsyah (2022), who found that the compensation dimension had the lowest positive sentiment in DANA reviews; with Nurcahyo et al. (2025), who found that DANA lagged behind LinkAja in responsiveness; and with Kar (2021), who identified responsiveness as a determinant of mobile payment satisfaction. Reviews in T8 and T9 show that users who have already experienced a transaction problem are disappointed a second time when the chatbot gives repetitive answers and there is no escalation to a human agent. This recovery failure appears to lower ratings most sharply.

Findings on security should be interpreted with caution. DANA reports zero material data breach incidents as well as ISO/IEC 27001 and PCI-DSS certification (DANA, 2025), whereas 10.18% of reviews in this study (T4) report missing balances or unrecognized transactions. These two facts are not necessarily contradictory. Balance losses reported by users may result from account takeover, social engineering, or third-party charges rather than from system data breaches. For user trust, however, what matters is the perceived experience Oh & Kim (2022) showed that security complaints lower the ratings of financial apps, and Yi et al. (2024) identified security as one of the quality factors affecting the market performance of payment apps. The login and OTP topic (T5) illustrates a trade-off between security and ease of access: strict verification protects accounts but makes access difficult for legitimate owners who lose their phone numbers.

Two topics point to themes that E-S-QUAL does not explicitly cover. First, DANA Cicil (T2) is the highest-rated topic but also contains complaints about opaque credit limits and eligibility. Users value its convenience but do not understand why a limit does not appear after repayment. Second, events and points (T6) is the topic with the highest coherence and very negative ratings. This suggests that a loyalty program that fails to deliver on its promises can damage evaluations of the core service. The categories used in previous studies, such as system ease, system problems, and customer service problems (Widiantoro & Harnadi, 2023), do not separate these two themes. The measurement of digital wallet service quality could therefore be extended with dimensions for credit feature transparency and loyalty program reliability.

Theoretically, these results support the relevance of E-S-QUAL and E-RecS-QUAL for digital wallets. They also show that the type of problem is statistically related to ratings, as found for financial apps in other countries (Oh & Kim, 2022; Yi et al., 2024). Practically, the findings yield five improvement priorities for digital wallet providers: (1) maintaining system capacity and stability during peak hours, events, and after updates, with proactive disruption notices; (2) setting clear time limits and status updates for refunds; (3) providing an escalation

path from the chatbot to human agents for balance and account cases; (4) offering alternative account recovery for users who lose their phone numbers; and (5) explaining the DANA Cicil limit criteria transparently. These priorities matter given the scale of DANA's services, which, according to the company, recorded 272% growth in QRIS transactions in 2023 (CGS International Sekuritas, 2024). When transaction volume grows this fast, system disruptions directly affect merchants and buyers at the point of payment.

CONCLUSION

This study identified ten topics in 3,979 DANA user reviews on Google Play using Gensim-based LDA with Indonesian-language preprocessing. The selected model has a c_v coherence of 0.417. The largest topic is transaction and system disruptions, followed by DANA Cicil and by deducted balance and refunds. When mapped to E-S-QUAL, about 70% of reviews relate to the fulfillment and system availability dimensions. Ratings differ significantly across topics: customer service and chatbot receives the lowest ratings, whereas DANA Cicil receives the highest. These findings indicate that improving system stability and service recovery is the main priority, and that credit features and loyalty programs deserve attention as additional dimensions of digital wallet service quality.

This study has several limitations. First, the data come from a single platform and a single collection, with 95.7% of reviews posted in July–October 2026. The data are also dominated by negative reviews (68.6% with 1–2 stars), so the results do not represent all users. Second, the reviews contain no demographic data, and ratings are only a proxy for satisfaction. Third, topic labeling and mapping to the E-S-QUAL dimensions were performed by the researchers without a word intrusion test or independent raters. Fourth, the normalization dictionary does not capture all slang (e.g., *gua*), and LDA's bag-of-words approach ignores word order and negation.

Future research is recommended to (1) use multi-year and multi-platform data, including social media; (2) validate the topics with human judgment, for example through a word intrusion test with several raters; (3) compare LDA with models designed for short texts, such as BERTopic or GSDMM; (4) link topic prevalence to ratings in panel data, as Oh and Kim (2022) did; and (5) test the dimensions of credit feature transparency and loyalty program reliability through user surveys.

REFERENCES

Adiningtyas, H., & Auliani, A. S. (2024). Customer Perceived Value: A Study Based on Customer Perception on Social Media. *Procedia Computer Science*, *234*, 1451–1458. https://doi.org/10.1016/j.procs.2024.03.145

Albalawi, R., Yeap, T. H., & Benyoucef, M. (2020). Using Topic Modeling Methods for Short-Text Data: A Comparative Analysis. *Frontiers in Artificial Intelligence*, *3*. https://doi.org/10.3389/frai.2020.00042

Aprilia, E. F., Arifiyanti, A. A., & Sembilu, N. (2025). Aspect-Based Sentiment Analysis on User Perceptions of OVO using Latent Dirichlet Allocation and Support Vector Machine. *Aviation Electronics, Information Technology, Telecommunications, Electricals, and Controls (AVITEC)*, *7*(2), 163. https://doi.org/10.28989/avitec.v7i2.3035

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, *3*(Jan), 993–1022.

Çallı, L. (2023). Exploring mobile banking adoption and service quality features through user-generated content: the application of a topic modeling approach to Google Play Store reviews. *International Journal of Bank Marketing*, *41*(2), 428–454. https://doi.org/10.1108/IJBM-08-2022-0351

CGS International Sekuritas. (2024, January 31). *Dana Indonesia Catatkan 170 Juta Pengguna Pada Tahun 2023*. https://itrade.cgsi.co.id/dana-indonesia-catatkan-170-juta-pengguna-pada-tahun-2023

Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J., & Blei, D. (2009). Reading Tea Leaves: How Humans Interpret Topic Models. *Advances in Neural Information Processing Systems*, *22*.

DANA. (2025). *Laporan Keberlanjutan DANA 2025* . https://a.m.dana.id/danaweb/cms/2025_Sustainabillity_Report_Template_5_FINAL_9_3730638a56.pdf

Dieng, A. B., Ruiz, F. J. R., & Blei, D. M. (2020). Topic Modeling in Embedding Spaces. *Transactions of the Association for Computational Linguistics*, *8*, 439–453. https://doi.org/10.1162/tacl_a_00325

Halim, D., Budi, I., Mubina, B. F., Budi Santoso, A., & Kresna Putra, P. (2026). Comparative Topic Modelling of Mobile Banking User Reviews Using LDA and BERTopic: A Case Study of wondr by BNI. *The Indonesian Journal of Computer Science*, *15*(2). https://doi.org/10.33022/ijcs.v15i2.5118

Hanifah, A., Saputra, M., & Fa'rifah, R. Y. (2025). Exploring User Sentiments and Adoption Barriers in E-Wallet Services in Indonesia using BERTopic on Play Store Reviews. *2025 4th International Conference on Electronics Representation and Algorithm (ICERA)*, 286–291. https://doi.org/10.1109/ICERA66156.2025.11086632

Hussain, A., Hannan, A., & Shafiq, M. (2023). Exploring Mobile Banking Service Quality Dimensions in Pakistan: a Text Mining Approach. *International Journal of Bank Marketing*, *41*(3), 601–618. https://doi.org/10.1108/IJBM-08-2022-0379

Islamiah, F., Hamzah, Muh. Q., Az Zahira, F., Pailang, R. A., & Agand, N. (2026). Analisis Sentimen Publik terhadap Pernyataan Menteri Keuangan tentang Kebijakan Pajak meggunakan Support Vector Machine. *Jurnal Teknologi Dan Manajemen Industri Terapan*, *5*(2), 993–999. https://doi.org/10.55826/jtmit.v5i2.1752

Islamiah, F., & Muthahharah, I. (2023). Sentiment Analysis Approach to User Perceptions of Ammana Sharia Fintech Based on Machine Learning Algorithms. *Jurnal Ekonomi Ichsan Sidenreng Rappang*, *2*(1), 359–369.

Islamiah, F., Rusmiati, R., & Wijaya, R. (2022). Penilaian Kepuasan Pengguna Website Sistem Informasi Akademik Menggunakan Metode Website Quality. *METIK JURNAL*, *6*(2), 133–139. https://doi.org/10.47002/metik.v6i2.381

Islamiyah, S., Permanasari, A. E., Pratama, A. R., & Saputra, M. (2026). Exploring User Experience and Technostress in Indonesian E-Wallets: A Comparative Topic Modeling Analysis Using LDA and Bertopic. *2026 International Seminar on Intelligent Business and Edge-Computing Research (ISIBER)*, 296–301. https://doi.org/10.1109/ISIBER68248.2026.11470035

Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. *Multimedia Tools and Applications*, *78*(11), 15169–15211. https://doi.org/10.1007/s11042-018-6894-4

Kar, A. K. (2021). What Affects Usage Satisfaction in Mobile Payments? Modelling User Generated Content to Develop the "Digital Service Usage Satisfaction Model."

Information Systems Frontiers, *23*(5), 1341–1361. https://doi.org/10.1007/s10796-020-10045-0

Karimani, G. S. (2026, April 15). *5 E-Wallet Favorit Publik RI untuk Bertransaksi*. https://data.goodstats.id/statistic/5-e-wallet-favorit-publik-ri-untuk-bertransaksi-lYpyi

Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., Pfetsch, B., Heyer, G., Reber, U., Häussler, T., Schmid-Petri, H., & Adam, S. (2018). Applying LDA Topic Modeling in Communication Research: Toward a Valid and Reliable Methodology. *Communication Methods and Measures*, *12*(2–3), 93–118. https://doi.org/10.1080/19312458.2018.1430754

Mimno, D., Wallach, H., Talley, E., Leenders, M., & McCallum, A. (2011). Optimizing Semantic Coherence in Topic Models. *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, 262–272.

Mustaqim, I. Z., & Suryono, R. R. (2025). A Systematic Literature Review of Topic Modeling Techniques in User Reviews. *Journal of Information Systems Engineering and Business Intelligence*, *11*(2), 238–253. https://doi.org/10.20473/jisebi.11.2.238-253

Nurcahyo, R., Prabuwono, A. S., Dewi, I. N., Farizal, Wibowo, N., Habiburrahman, M., Hindriyandhito, K., & Jang, Y. M. (2025). Enhancing Digital Wallet Services Based on Electronic Word of Mouth Using Social Media Reviews. *Sage Open*, *15*(3). https://doi.org/10.1177/21582440251358713

Oh, Y. K., & Kim, J.-M. (2022). What Improves Customer Satisfaction in Mobile Banking Apps? An Application of Text Mining Analysis. *Asia Marketing Journal*, *23*(4). https://doi.org/10.53728/2765-6500.1581

Parasuraman, A., Zeithaml, V. A., & Malhotra, A. (2005). E-S-QUALE-S-QUAL: A multiple-item scale for assessing electronic service quality. *Journal of Service Research*, *7*(3), 213–233. https://doi.org/10.1177/1094670504271156

Perea-Khalifi, D., Irimia-Diéguez, A. I., & Palos-Sánchez, P. (2024). Exploring the determinants of the user experience in P2P payment systems in Spain: a text mining approach. *Financial Innovation*, *10*(1), 2. https://doi.org/10.1186/s40854-023-00496-0

Raza, A., Al-Zaeemi, S. A., Al-Sharafi, M. A., Albashrawi, M., Dwivedi, Y. K., & Chae, I. (2026). A big data and machine learning-based approach to theorising consumers' continuance intentions to use digital wallets. *Behaviour & Information Technology*, 1–22. https://doi.org/10.1080/0144929X.2026.2726454

Řehůřek, R., & Sojka, P. (2010). *Software framework for topic modelling with large corpora*.

Röder, M., Both, A., & Hinneburg, A. (2015). Exploring the Space of Topic Coherence Measures. *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, 399–408. https://doi.org/10.1145/2684822.2685324

Ruslan, A., Aris, V., Taufik, M., Islamiah, F., & Asizah, A. B. M. (2025). Analisis Perbadingan Kualitas Layanan Digital pada Bank Kovensional dan Bank Digital di Indonesia. *RIGGS: Journal of Artificial Intelligence and Digital Business*, *4*(4), 145–156. https://doi.org/10.31004/riggs.v4i4.3188

Saputra, M., Santosa, P. I., & Permanasari, A. E. (2023). Technostress Effects on Indonesian e-Wallet Users: A Case Study with LDA Modeling and Sentiment Analysis. *2023 7th International Conference on Information Technology (InCIT)*, 115–120. https://doi.org/10.1109/InCIT60207.2023.10413062

Saputra, N. J., & Safitri, C. (2025). Enhancing User Satisfaction on a Digital Payment Platform Using Topic Modeling and Machine Learning Techniques. *2025 International Conference on Information and Communication Technology (ICoICT)*, 1–6. https://doi.org/10.1109/ICoICT66265.2025.11193077

Sievert, C., & Shirley, K. (2014). LDAvis: A method for visualizing and interpreting topics. *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*, 63–70.

Sumidra, I. K. S. C., Saputra, M., & Fa'rifah, R. Y. (2026). From User Reviews to Technostress Dimensions: a Data Driven Study of Indonesian E-Wallet Applications. *2026 18th International Conference on Electronics, Computer, and Computation (ICECCO)*, 1–6. https://doi.org/10.1109/ICECCO67619.2026.11488786

Taufik, M., Aris, V., Ruslan, A., Islamiah, F., & Asizah, A. B. M. (2025). Analisis Kualitas Layanan Digital pada Sektor Perbankan Digital. *RIGGS: Journal of Artificial Intelligence and Digital Business*, *4*(1), 360–367. https://doi.org/10.31004/riggs.v4i1.417

Ulya, D. I., Kunaefi, A., Rolliawati, D., & Nugroho, B. A. (2023). Unpacking Public Perceptions of Qris with Twitter Data: A Vader And LDA Methodology. *Jurnal ELTIKOM : Jurnal Teknik Elektro, Teknologi Informasi Dan Komputer*, *7*(2), 145–159. https://doi.org/10.31961/eltikom.v7i2.742

Weisser, C., Gerloff, C., Thielmann, A., Python, A., Reuter, A., Kneib, T., & Säfken, B. (2023). Pseudo-document simulation for comparing LDA, GSDMM and GPM topic models on short and sparse text using Twitter data. *Computational Statistics*, *38*(2), 647–674. https://doi.org/10.1007/s00180-022-01246-z

Widiantoro, A. D., & Harnadi, B. (2022). Review of User Comments for the OVO Fintech application using LDA. *2022 6th International Conference on Information Technology (InCIT)*, 326–330. https://doi.org/10.1109/InCIT56086.2022.10067746

Widiantoro, A. D., & Harnadi, B. (2023). Analysis of User Comments Based on Topic Modeling using LDA on OVO E-Wallet. *SISFORMA*, *9*(2), 99–107. https://doi.org/10.24167/sisforma.v9i2.5343

Widiantoro, A. D., Mustafid, M., & Sanjaya, R. (2023). Analyzing User Comments on Fintech E-wallet Applications in Indonesia Using the LDA Algorithm. *2023 7th International Conference on Information Technology (InCIT)*, 169–174. https://doi.org/10.1109/InCIT60207.2023.10412996

Yi, J., Kim, J., & Oh, Y. K. (2024). Uncovering the quality factors driving the success of mobile payment apps. *Journal of Retailing and Consumer Services*, *77*, 103641. https://doi.org/10.1016/j.jretconser.2023.103641

Zahra, S., & Alamsyah, A. (2022). Digital Wallet Service Quality Analysis using Multiclass Classification and Sentiment Analysis. *2022 5th International Conference on Information and Communications Technology (ICOIACT)*, 256–261. https://doi.org/10.1109/ICOIACT55506.2022.9971936

Zahra, S. N. A., Saputra, M., & Fa'rifah, R. Y. (2025). A Data-Driven Analysis of Understanding Technostress in Indonesian E-Wallets Using BERTopic. *2025 4th International Conference on Electronics Representation and Algorithm (ICERA)*, 298–303. https://doi.org/10.1109/ICERA66156.2025.11087262