STABILITY
Journal of Management & Business
Vol. 8 No.2 Years 2025
http://journal.upgris.ac.id/index.php/stability

CLUSTER ANALYSIS OF DANA AND GOPAY E-WALLET SERVICE REVIEWS USING TF–IDF AND K-MEANS/K-MEDOIDS: AN ECONOMIC AND MANAGEMENT STRATEGY PERSPECTIVE

Fatkhul Amin [1], Fithri Widyanita Yarisma [2], Ira Setiawati [3], Ratih Milati Ilham[4]

Universitas PGRI Semarang, Semarang, Indonesia [1234]

Article Information

Article History:
Submission : December 2025
Accepted : December 2025
Published : December 2025

Keywords:
Customer Trust;
Digital Wallet Adoption;
Perceived Ease of Use;
Perceived Usefulness;
Security and OTP.

INDEXED IN

SINTA - Science and Technology Index
Dimensions
Google Scholar
ReseachGate
Garuda

Abstract

This study aims to (1) identify the thematic structure of Indonesian-language user reviews on digital wallet services, (2) compare the clustering quality produced by alternative unsupervised methods, and (3) derive actionable operational insights for service management and financial performance. A quantitative text-mining design is employed, using an unsupervised learning framework on Indonesian-language Google Play Store reviews collected over the 2021–2025 period, comprising 553.500 reviews for DANA and 265.632 reviews for GoPay. The corpus is obtained via staged web scraping, cleaned and normalized, then represented with TF–IDF uni- and bigram vectors. Two clustering algorithms, K-Means and K-Medoids with cosine distance, are applied separately to the DANA, GoPay, and combined corpora. Cluster quality is evaluated using internal validity indices, including the Silhouette while thematic interpretation relies on dominant term inspection and supporting visualizations. The findings show that DANA's main themes revolve around user experience, ease and reliability of transactions, OTP-based security and access barriers, and customer support, whereas GoPay reviews emphasize transactional operations, balance accessibility, and promotions or perceived service levels. The evaluation results indicate that DANA exhibits more heterogeneous thematic segments than GoPay, implying the need for more differentiated targeting and service strategies, while GoPay can rely on a relatively more homogeneous user base.

Correspondence Address : Universitas PGRI Semarang, Indonesia
Institusion Address : Jl. Sidodadi Timur No.24, Karangtempel, Semarang Timur
E-mail : fatkhulamin@upgris.ac.id

DOI : http://dx.doi.org/ 10.26877/r7hyad25

OPEN ACCESS
ISSN :2621-850X
E-ISSN : 2621-9565

INTRODUCTION

Over the past few years, Indonesia has experienced a rapid expansion of its digital economy, particularly in e-commerce and digital payments (Razaq et al., 2025). Official statistics show that the value of electronic money and e-commerce transactions has increased sharply year-on-year, more than doubling between the late 2010s and the early 2020s as non-cash instruments became deeply embedded in daily transactions. Regional reports such as e-Conomy SEA similarly highlight Indonesia as the largest digital payments market in Southeast Asia, underscoring the strategic role of digital wallets as infrastructure for both micro-level consumer spending and macro-level cash flow (Berlianawati et al., 2024). These developments reflect not only technological diffusion but also broader structural changes in how Indonesian consumers pay, save, and interact with financial services (Jatmika et al., 2023).

The adoption of digital wallets has been accompanied by clear year-to-year shifts in consumer behaviour. Before the COVID-19 pandemic, digital wallets were still a complementary payment option; during and after the pandemic, their usage surged as consumers sought contactless, convenient, and integrated payment experiences (Rahmadan & Zihni, 2025). Recent survey evidence indicates that e-wallets have overtaken several traditional instruments for everyday transactions, signalling a persistent preference for speed, ease of access, and integration with online ecosystems (Anwar et al., 2023). Within this landscape, a small number of providers dominate the Indonesian market, with ShopeePay, OVO, GoPay, DANA, and LinkAja consistently ranked as the most used platforms (Diva & Anshori, 2024). Among them, GoPay and DANA are repeatedly reported as having high penetration and high frequency of use, so focusing on these two wallets is empirically justified to capture user perceptions of mainstream services and to compare their thematic strengths and weaknesses (Anwar et al., 2023).

A growing body of international and Indonesian research has examined the determinants of digital wallet adoption and continued use. Many studies employ the Technology Acceptance Model to analyse how perceived usefulness, ease of use, trust, risk, financial incentives, and social influence shape intention and continuance intention toward e-wallets in countries such as Indonesia, Malaysia, India, and China. Other works apply sentiment analysis or clustering to user-generated content, but typically in different domains, such as e-commerce products, tourism services, or healthcare reviews (Johan et al., 2022). Despite these contributions, there remains a scarcity of studies that (i) use large-scale Indonesian-language customer reviews as a Voice of Customer (VoC) data source for digital wallets, (ii) apply unsupervised clustering to uncover emergent themes without imposing pre-defined categories, and (iii) directly compare clustering algorithms in this specific context. This shortfall of review-based, algorithmically robust thematic mapping limits our understanding of how users experience operational issues, security, and promotional policies across competing e-wallet providers over time (Banutama et al., 2025).

Against this backdrop, the present study addresses several key objectives. First, it seeks to identify the dominant themes in Indonesian-language user reviews of DANA and GoPay on the Google Play Store and to observe how the prominence of these themes evolves year by year (Bimantara & Widiartha, 2023). Second, it aims to compare the clustering quality of two unsupervised algorithms K-Means and K-Medoids when applied to a high-dimensional representation of review texts. The representation is based on term frequency–inverse document frequency (TF–IDF), a

standard text-mining weighting scheme that assigns higher weights to terms that occur frequently in a given document but are relatively rare across the entire corpus, thereby highlighting words that are most informative for distinguishing thematic clusters (Sunarko et al., 2023). Third, the study intends to translate the resulting thematic structures into operational and economic recommendations that can enhance service reliability, user experience, and financial performance for digital wallet providers in Indonesia (Wulandari & Idayanti, 2023).

In doing so, this research makes both theoretical and practical contributions. Theoretically, it enriches the digital payment and VoC literature by demonstrating how unsupervised text mining of a large Indonesian-language review corpus spanning multiple years and focusing on two leading digital wallets can generate a detailed, context-specific thematic map of fintech services (Saputri & Pratama, 2021). Practically, it offers evidence-based guidance for managers of digital wallet platforms, including how to prioritise investments in security and authentication, refine transaction-related user experience, and design more transparent pricing and promotional strategies (Naura Nadra Rizki et al., 2024). By linking large-scale, longitudinal user feedback with clustering based thematic analysis, the study provides a robust empirical basis for continuous monitoring and strategic decision-making in Indonesia's rapidly evolving fintech ecosystem (Pratiwi, 2024).

METHOD

This study employs a quantitative text-mining approach with an unsupervised learning and comparative evaluation design applied to Indonesian-language user reviews of the DANA and GoPay digital wallet applications. The dataset consists of Google Play Store reviews collected over the 2021–2025 period, comprising 553.500 reviews for DANA and 265.632 reviews for GoPay. Reviews are obtained through staged web scraping, then filtered to retain only Indonesian-language content and processed through case-folding, tokenisation, stop-word removal, normalisation, and stemming.

The cleaned corpus is represented using term frequency inverse document frequency (TF–IDF) uni and bigram vectors to highlight terms that are frequent within a document but relatively rare across the corpus, thereby capturing the most informative words for clustering. Two clustering algorithms, K-Means and K-Medoids with cosine distance, are applied separately to the DANA corpus, the GoPay corpus, and the combined corpus. Cluster quality is assessed using internal validity indices, including the Silhouette coefficient while thematic interpretation is conducted by inspecting dominant terms and representative reviews in each cluster, supported by visualisations such as word clouds and temporal trend plots. All procedures are implemented in a Python-based environment.

Data collection and preparation

1. Data are collected through staged web scraping from review pages and their pagination until the target time span is covered. To support longitudinal analysis, each review is stored with its ID, timestamp, and rating.
2. An initial validation of data quality is conducted through random sampling to assess the readability and relevance of review content.
3. Text preprocessing includes case folding, removal of punctuation/numbers/URLs/emoticons, normalization of common spelling

variations, Indonesian-language tokenization, stopword removal, and stemming. Duplicates and near-duplicates are controlled using simple similarity checks to prevent distortion of the clustering process.

Representation and modeling

1. Documents are represented using TF–IDF with an n-gram range of unigrams and bigrams, along with min_df and max_df settings to filter out terms that are too rare or too frequent. Vector normalization uses L2 to improve the stability of inter-document distances.
2. A range of cluster numbers (for example, k = 2–15) is explored, with the primary evaluation based on the Silhouette score using cosine distance. The Calinski–Harabasz and Davies–Bouldin indices are also recorded to support sensitivity analysis and cluster quality comparison.
3. Modeling is performed separately on the DANA corpus, the GoPay corpus, and a combined corpus to assess thematic consistency across applications.

Clustering methods:

1. K-Means: initialization uses a convergence-improving scheme with multiple restarts until a stable solution is obtained.
2. K-Medoids (PAM): the BUILD–SWAP scheme with cosine distance is applied to enhance robustness against outliers and noise, which are common in user-generated reviews.

Cluster quality evaluation:

1. Internal evaluation is conducted using the Silhouette Score to determine the extent to which clusters are clearly separated.
2. A comparative analysis between the two algorithms is carried out to identify differences in thematic proportions, cluster stability, and label consistency under re-initialized parameter settings.
3. Thematic validation is performed by interpreting dominant keywords in each cluster and by using supporting visualizations (word clouds/2D projections), enabling managerial stakeholders to easily comprehend the findings.

Limitations and ethical considerations

1. Reliance on public reviews entails a self-selection bias, such that the results may not fully represent the overall user population.
2. The handling of sensitive data adheres to privacy policies, avoiding any identification of individuals in the reporting of results.

Practical implications

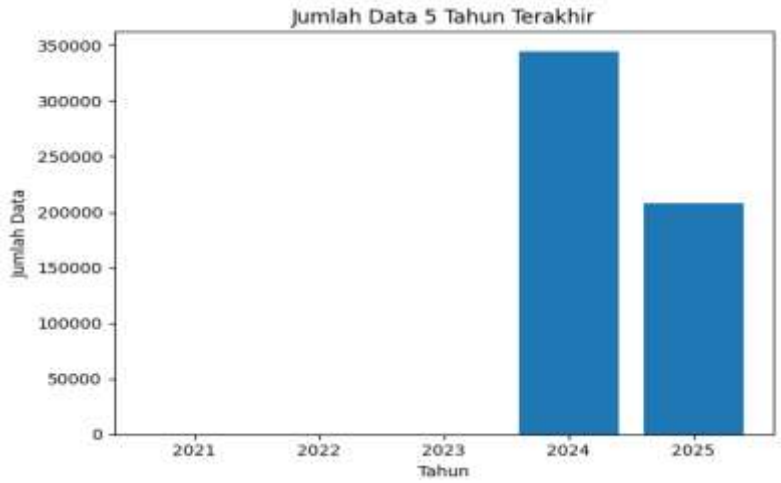
1. Structured thematic findings can serve as a basis for prioritizing operational improvements, particularly in authentication reliability, transaction smoothness, fee transparency, and responsiveness of customer support.
2. The comparative approach using two algorithms provides insights into thematic stability and the reliability of managerial recommendations, thereby supporting the formulation of product policies and marketing strategies.

RESULTS AND DISCUSSION

Results

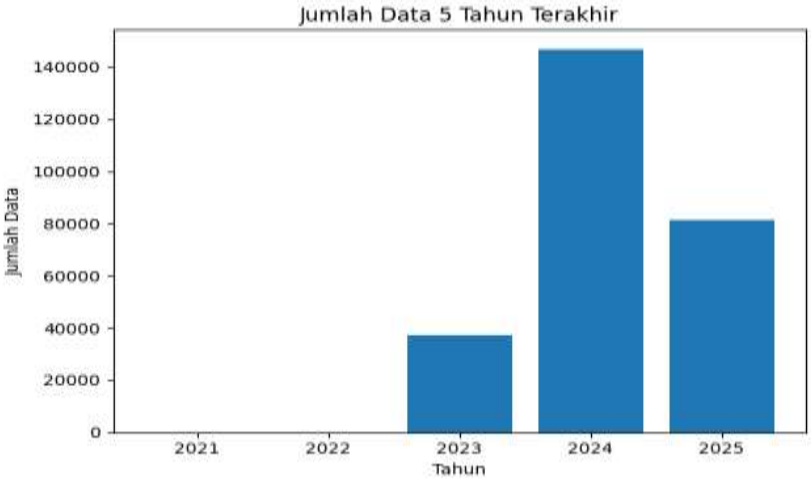
The dataset consists of public Google Play Store reviews for the DANA application in CSV format, nominally covering a five-year window from 2021 to 2025. However, the annual recap shows that no usable entries are available for 2021–2023, while 2024 accounts for 345,038 reviews and 2025 for 208,462 reviews, resulting in a total of 553,500 reviews concentrated in the two active years (2024–2025). The absence of records for 2021–2023 does not indicate that the application had no users during those years; rather, historical reviews from that period could not be retrieved because earlier reviews had been removed or were no longer accessible via the Google Play API at the time of data collection. Consequently, the longitudinal analysis in this study effectively focuses on the 2024–2025 period, when the data are complete and consistently available.

Picture 1. Number of user reviews collected for the DANA application



The review Gopay Application composition over the nominal five-year window (2021–2025) is in practice heavily concentrated in the last three years. No usable reviews could be retrieved for 2021–2022 because earlier records were no longer accessible through the Google Play review interface used in the scraping process, so those years contain zero observations. This absence of data does not indicate that the application had no users during those years; rather, historical reviews from that period could not be retrieved because earlier reviews had been removed or were no longer accessible via the Google Play API at the time of data collection. In contrast, 2023, 2024, and 2025 contribute 37.332 reviews, 146.960 reviews, and 81.340 reviews, respectively, with the composition dominated by 2024 and 2025 and 2023 accounting for a smaller share of the corpus. (The visualization of the annual distribution is presented in picture 2).

Picture 2. Number of user reviews collected for the GoPay application



All entries underwent text cleaning and normalization, including case folding, removal of punctuation/numbers/URLs/emoticons, Indonesian-language tokenization, stopword removal, and stemming. The result is a clean corpus ready for feature extraction. The documents were then represented using TF–IDF to capture the distinctiveness of terms across reviews prior to the clustering modeling stage. This scheme enables an objective comparison of themes between the years under analysis.

The concentration of data in 2024–2025 indicates that the interpretation of trends primarily reflects the most recent period, making the findings relevant for short- to medium-term operational decision-making (improving service reliability and user experience). The TF–IDF bigram results for DANA reviews are presented in Table 1.

Table 1. TF–IDF bigram results for user reviews of the DANA application

```
================================================================================
TOP 30 BIGRAM TERBANYAK (setelah stopword)
--------------------------------------------------------------------------------
dana cicil                      2451
kode otp                        796
saldo dana                      643
pake dana                       635
pakai dana                      561
dana instan                     507
dana dana                       504
dana premium                    460
verifikasi wajah                447
masuk dana                      384
uang dana                       380
tolong dana                     327
tolong perbaiki                 303
kasih bintang                   297
tolong bantu                    238
dana aman                       235
cicil dana                      235
terima kasih                    227
dana transaksi                  226
dana tolong                     225
dana saldo                      224
kirim uang                      216
cicil muncul                    206
dana masuk                      193
muncul dana                     189
transaksi dana                  178
dana membantu                   176
isi saldo                       176
pengguna dana                   171
dana uang                       166

================================================================================
SUKSES! Hasil CSV bersih tersimpan di: /content/data bersih dana.csv
================================================================================
```

Source: Authors' processing, 2025

The TF–IDF bigram analysis shows that word combinations such as “dana cicil”, “kode otp”, “saldo dana”, and “pake dana” appear most frequently in user reviews. This pattern indicates that users predominantly discuss transaction features, security, and the ease of using the application. The occurrence of terms such as “verifikasi wajah” (face verification) and “dana aman” (funds are safe) underscores the importance of security and trust in fintech services, while phrases like “tolong perbaiki” (please fix) and “kasih bintang” (give stars) reflect direct feedback on service quality.

From a business perspective, these findings suggest that DANA’s development and marketing strategies should prioritize enhancing the reliability of digital financial features, transaction security, and user satisfaction as key factors in maintaining customer loyalty and competitive advantage in the fintech industry.

Table 2.TF–IDF bigram results for user reviews of the GoPay application

```
================================================================================
TOP 30 BIGRAM TERBANYAK (setelah stopword)
--------------------------------------------------------------------------------
saldo masuk                    772
pake gopay                     675
saldo gopay                    668
gopay pinjam                   652
pakai gopay                    631
tolong perbaiki                472
masuk masuk                    466
top masuk                      456
kasih bintang                  452
terima kasih                   444
gopay plus                     432
masuk saldo                    429
aktivitas wajar                388
gopay masuk                    384
masuk gopay                    375
isi saldo                      373
top saldo                      370
beli pulsa                     357
top gopay                      343
top game                       314
gopay tabungan                 295
paket data                     294
gopay saldo                    262
transfer bank                  261
token listrik                  234
saldo terpotong                232
the best                       229
upgrade gopay                  223
gopay membantu                 222
terimakasih gopay              212

================================================================================
SUKSES! Hasil CSV bersih tersimpan di: /content/data bersih gopay.csv
================================================================================
```

Source: Authors’ processing, 2025

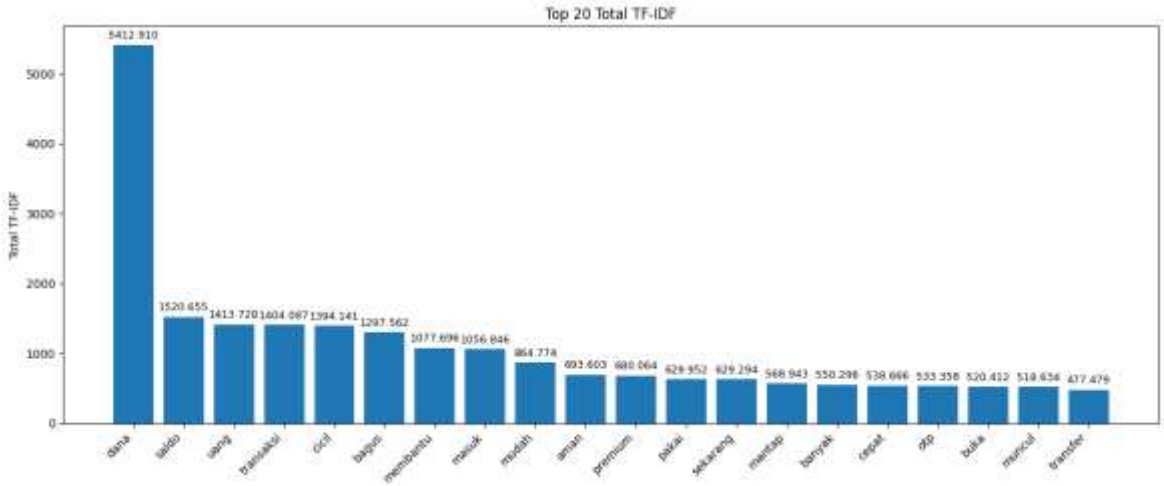
Table 2 presents the results of the TF–IDF bigram analysis, indicating that word combinations such as “saldo masuk”, “pake gopay”, “saldo gopay”, and “gopay pinjam” are among the most frequently occurring terms in user reviews. This pattern reflects that the majority of user interactions are concentrated on digital financial transaction activities, including balance top-ups, the use of balance for payments, and access to lending features.

The emergence of phrases such as “tolong perbaiki” (please fix), “kasih bintang” (give stars), and “terima kasih” (thank you) indicates the presence of direct feedback on service quality and user experience. This suggests that customer satisfaction and service responsiveness remain central concerns in users’ perceptions of GoPay. In addition, terms such as “beli pulsa”, “paket data”, and “token listrik” demonstrate the diversified

use of GoPay across various digital economic services, reinforcing its position as an integrated payment ecosystem.

From a business and managerial perspective, these findings imply that GoPay's development strategy should focus on improving transaction reliability, system security, and ease of access to ancillary services. Such an approach has the potential to strengthen user loyalty, increase customer retention, and expand digital market share amid intensifying competition in the fintech industry.

Picture 3. Histogram of the top 20 TF–IDF word scores for user reviews of the DANA application

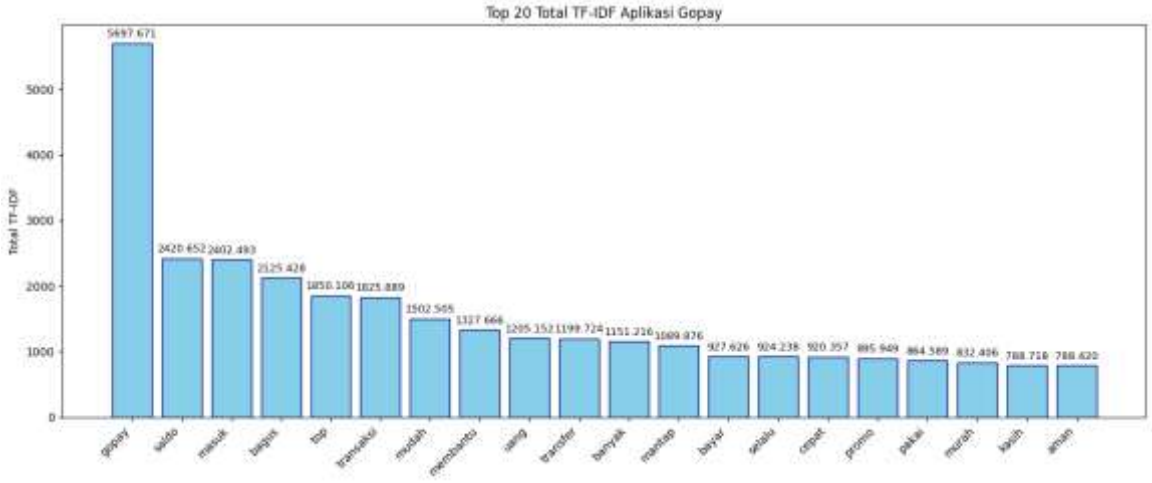


Picture 3 above displays the distribution of the 20 words with the highest TF–IDF values from user reviews of the DANA application on the Google Play Store. The word "dana" occupies a dominant position with a score of 5,412.910, indicating that the DANA brand is the central focus in nearly all reviews. This reflects a high level of brand awareness as well as consistent mention of the brand by users. Furthermore, words such as "saldo" (balance), "uang" (money), and "transaksi" (transaction) rank next, each with TF–IDF values above 1,400, highlighting users' focus on the core function of the application as a digital financial transaction platform.

In addition, terms such as "cicil" (installment), "bagus" (good), and "membantu" (helpful) indicate users' positive perceptions of the ease of access and benefits provided by DANA's services, while the words "aman" (secure) and "mudah" (easy) reinforce the importance of trust and user experience. Words like "cepat" (fast), "mantep" (excellent), and "banyak" (many) reflect users' satisfaction with the application's performance, whereas terms such as "otp", "buka" (open), and "transfer" point to the transaction and verification activities that are frequently carried out.

From a business and managerial perspective, this pattern suggests that DANA users evaluate the application primarily on the basis of security, transaction convenience, and service speed as key factors shaping their usage decisions (Wulandari & Idayanti, 2023). Consequently, DANA's management strategy should focus on improving the quality of the transaction system, strengthening digital security, and optimizing user-oriented service features in order to maintain customer loyalty and competitiveness in the national fintech market.

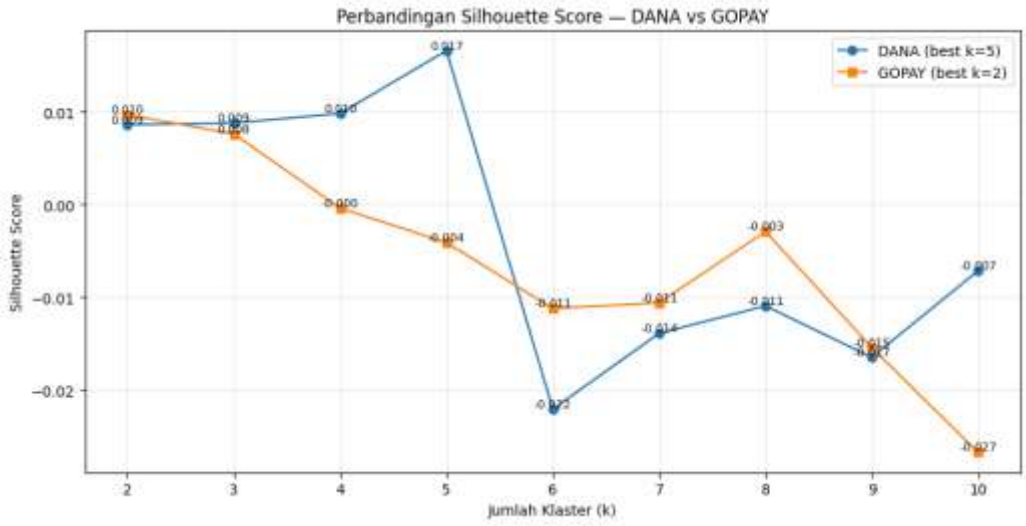
Picture 4. Histogram of the top 20 TF–IDF word scores for user reviews of the GoPay application



Picture 4 above presents the 20 words with the highest TF–IDF values based on an analysis of user reviews of the GoPay application on the Google Play Store. The word "gopay" has the highest score of 5,697.671, indicating the dominance of brand-related discussions and a high level of brand recall among users. Subsequently, the words "saldo" (2,420.652), "masuk" (2,402.493), and "top" (1,850.106) appear most frequently, reflecting users' focus on balance top-up and utilization of digital balance as the core functions of the application. The words "bagus" (good) and "mudah" (easy) indicate positive perceptions of service quality and ease of use, while terms such as "transaksi" (transaction), "transfer", and "bayar" (pay) point to the high intensity of application usage in day-to-day digital payment activities. The presence of words like "promo", "cepat" (fast), and "murah" (cheap/affordable) underscores the importance of promotional strategies and service efficiency in shaping user satisfaction and loyalty. Conversely, terms such as "aman" (secure) and "selalu" (always) capture perceptions of GoPay's reliability and consistent performance as a digital payment platform.

From a business management perspective, these results indicate that GoPay has successfully emphasized a value proposition centered on convenience, security, and transaction speed, which are key factors in maintaining competitiveness in the fintech sector (Azizah & Ernawati, 2024). Consequently, GoPay's product development and marketing communication strategies should continue to focus on strengthening transactional features, enhancing user experience, and optimizing digital promotions grounded in customer loyalty (Jannati et al., 2024).

Picture 5. Silhouette score results for clustering of the DANA and GoPay applications



The graph shown in Picture 5 presents a comparison of the Silhouette Scores between the DANA and GoPay applications in the user clustering analysis. The highest value for DANA is achieved at k = 5 (0.017), while for GoPay it occurs at k = 2 (0.010). This indicates that the user structure of DANA is more heterogeneous, whereas GoPay's users are more homogeneous. These results suggest that DANA has a more complex market segmentation, thereby requiring differentiated marketing strategies for each segment. Conversely, GoPay can adopt a more standardized approach because its user behavior patterns are relatively uniform. From a managerial standpoint, this information is crucial for decision-making related to product development, promotional strategies, and the enhancement of customer loyalty (Safitri Juanita & Cahyono, 2024).

Picture 6. Word cloud visualization results for the DANA application

The word cloud visualization for cluster 2 of the DANA application, as shown in Figure 7, depicts the main theme “helpful / easy / good”. The largest words appearing in the cloud—such as “mudah” (easy), “membantu” (helpful), “bagus” (good), “mantap” (excellent), “transaksi” (transaction), and “aman” (secure)—indicate that the majority of users evaluate DANA positively, particularly in terms of ease of use, transaction speed, and system reliability. This reflects a high level of user satisfaction with respect to service efficiency and overall user experience. From a managerial perspective, these findings are important because they demonstrate DANA’s success in creating added value through the facilitation of digital financial transactions, which has the potential to enhance customer retention and brand loyalty.

Picture 7. Word cloud visualization results for the GoPay application

The word cloud visualization for the GoPay application shown in Picture 8 exhibits a dominant theme of “gopay / bagus / masuk”. The most prominent words such as “gopay”, “bagus” (good), “mudah” (easy), “saldo” (balance), “transaksi” (transaction), “mantap” (excellent), and “cepat” (fast) indicate that users’ perceptions of GoPay are generally positive, particularly with regard to ease of use, transaction speed, and convenient balance access. These results suggest that GoPay has successfully created high utility value for its users, especially in terms of transaction functions and integration with other service ecosystems (such as transportation and payments). From a managerial standpoint, the dominance of these positive terms reflects a strong level of user satisfaction and trust, which is crucial for customer retention strategies and for strengthening the brand’s position in Indonesia’s digital wallet market.

Discussion

The analysis of public Google Play review datasets for the DANA and GoPay applications illustrates the dynamics of user perceptions and behaviors during the active period of 2024–2025. The concentration of data in the last two years (DANA: 345,038 reviews in 2024 and 208,462 in 2025) indicates an increase in user participation in providing feedback on digital financial services, there by suggesting that this analysis reflects the current conditions of the fintech market.

Textual Analysis and Review Patterns

Using a TF–IDF bigram approach, the analysis reveals that for the DANA application, terms such as “dana cicil”, “kode otp”, and “saldo dana” are dominant, indicating users’ focus on transaction features, security, and ease of access. In contrast, for the GoPay application, expressions such as “saldo masuk”, “pake gopay”, and “gopay pinjam” are more prominent, representing the intensity of usage in digital transactions and lending services.

Both applications also exhibit the occurrence of phrases such as “tolong perbaiki” (please fix) and “kasih bintang” (give stars), which reflect direct feedback on service quality and serve as important indicators in customer satisfaction management. This aligns with the findings of Alfa et al., (2025), who emphasize the critical role of customer satisfaction in fostering user loyalty in e-wallet services.

TF–IDF Unigram Analysis (Top 20)

The distribution of the highest TF–IDF values reinforces the previous findings. For DANA, the words “dana”, “saldo” (balance), “uang” (money), and “transaksi” (transaction) underscore the application’s primary role as a medium for digital payment and transfer that is easy, fast, and secure. For GoPay, the dominance of the terms “gopay”, “saldo”, “masuk” (credited), and “top” indicates strong brand awareness and highlights its core function as a multifunctional transaction tool.

These patterns suggest that both platforms offer a value proposition centered on convenience, speed, and security, yet with distinct strategic emphases: DANA tends to be more strongly associated with security and transaction simplicity, whereas GoPay is more prominent in terms of service integration and promotional activities. This thematic configuration is consistent with several prior studies on e-wallet adoption, which collectively highlight security, ease of use, and perceived convenience as key determinants of users’ behavioural intentions. In particular, these studies show that users are more likely to adopt and continue using digital wallets when they perceive the application as secure, easy to operate, and aligned with their everyday transactional needs, which is precisely reflected in the emphasis on security- and promotion-related themes in the present review corpus.

Cluster Analysis and User Segmentation

The Silhouette Score results indicate that DANA attains its highest internal validity at $k = 5$ (0.017), whereas GoPay reaches its optimum at $k = 2$ (0.010). Although the absolute values are relatively low which is common in high-dimensional, sparse text data the relative difference in optimal k suggests a more heterogeneous thematic structure for DANA and a more homogeneous structure for GoPay (Azarya & Budi, 2025). In practical terms, DANA users tend to be distributed across several distinct concern clusters (for example, authentication and OTP issues, transaction failures, usability, and customer support), while GoPay users are grouped into fewer, broader themes dominated by transactional convenience and integration with other services (Safitri Juanita & Cahyono, 2024). This indicates that the DANA user base is characterised by more differentiated expectations and pain points, whereas GoPay users share a more uniform set of experiences and evaluative criteria (Khairuna et al., 2025).

From a managerial perspective, the higher optimal k for DANA implies that a "one-size-fits-all" approach is likely to be less effective. DANA needs to design more granular segmentation and targeting strategies such as separating security-sensitive users from transaction-oriented users and those who prioritise customer support so that product improvements, communication styles, and promotional offers can be aligned with the dominant issues in each segment (Amalia et al., 2023). In contrast, the lower optimal k for GoPay suggests that the platform can rely more on mass strategies that leverage its integrated ecosystem (e.g., ride-hailing, food delivery, and bill payments), while still monitoring emerging subsegments that might require tailored interventions (Turnandes et al., 2025). These findings reinforce the notion that user satisfaction and loyalty depend not only on generic service quality, but also on how well providers match their value propositions to the specific configuration of themes experienced by different user groups, which is consistent with prior evidence that perceived service performance plays a critical role in fostering loyalty (Asep Koswara, 2025).

User Perception Analysis (Word Cloud)

The word cloud visualizations show that DANA is dominated by terms such as "mudah" (easy), "membantu" (helpful), "bagus" (good), "mantap" (excellent), and "aman" (secure), reflecting a positive image in terms of ease of use, reliability, and user satisfaction. For GoPay, the most salient words are "bagus" (good), "mudah" (easy), "saldo" (balance), "cepat" (fast), and "transaksi" (transaction), indicating a strong perception of ease of use and service speed. These findings suggest that both applications have successfully built a positive user experience, which is a key factor in sustaining loyalty and expanding the adoption of fintech services. This is consistent with Olivia (2022), who finds that customer satisfaction has a positive effect on the intention to continue using the service.

CONCLUSIONS AND SUGGESTIONS

This study shows that user reviews of DANA and GoPay on the Google Play Store are structured around a set of recurring themes that differ in emphasis across platforms. DANA is more strongly associated with issues of security and OTP-based authentication, ease and reliability of transactions, and customer support, reflecting a user base with diverse concerns and expectations (Apriyanto & Sitio, 2025). In contrast, GoPay reviews tend to concentrate on transactional convenience, balance accessibility, and the benefits of an integrated service ecosystem, indicating a more homogeneous configuration of user experiences (Rochmanto et al., 2024). The comparative clustering results further reveal that the optimal number of clusters is higher for DANA than for GoPay, which empirically supports the interpretation that DANA's thematic space is more fragmented and requires more nuanced segmentation, while GoPay's thematic space is more compact (Jayanti et al., 2024).

The novelty of this research lies in the comparative application of two clustering algorithms K-Means and K-Medoids to a large-scale, Indonesian-language review corpus of digital wallet services, combined with internal validity evaluation and thematic interpretation. By doing so, the study demonstrates how unsupervised text mining can produce a detailed and context-specific thematic map for Indonesia's fintech market, thereby filling the gap in prior work that has largely relied on survey-based models or single-algorithm analyses. This integrated approach strengthens the empirical basis for understanding user perceptions of digital wallets and offers a

methodological template for future Voice of Customer (VoC) research in similar settings (Trianto et al., 2023).

MANAGERIAL IMPLICATIONS

Practical recommendations for management include: (1) strengthening security infrastructure, particularly OTP-based authentication and related access mechanisms, to reduce login barriers and increase user trust; (2) prioritising user experience through intuitive interface design, clear navigation, and seamless payment processes that minimise transaction failures; (3) clarifying fee structures and promotional communication to avoid misperceptions of service value and prevent dissatisfaction related to hidden or misunderstood charges; (4) for DANA, designing more differentiated segmentation and communication strategies that address distinct clusters of user concerns for example, segments focused on security, transaction simplicity, or customer support responsiveness; and (5) for GoPay, further leveraging and expanding its integrated ecosystem of digital financial and lifestyle services to sustain perceived added value for a relatively homogeneous user base, while continuously monitoring emerging subsegments that may require targeted interventions.

THEORETICAL IMPLICATIONS AND FUTURE RESEARCH

From a theoretical standpoint, the findings support and extend existing models of digital payment adoption by showing that core constructs such as perceived security, ease of use, and promotional benefits are reflected not only in survey responses but also in naturally occurring, large-scale user reviews. The clustering-based thematic structures identified in this study can be used as an empirical foundation to refine or enrich constructs in TAM, UTAUT2, and VoC frameworks within the context of Indonesian fintech services. Future research can build on this work by broadening the data sources (for instance, incorporating reviews from additional platforms or social media, and extending the time horizon) to capture more comprehensive and long-term dynamics of user perceptions (Gosal et al., 2024). Subsequent studies may also integrate economic metrics such as average revenue per user (ARPU), retention, and churn into the analysis, enabling a more direct linkage between thematic review patterns and financial performance (Saputri & Pratama, 2021). In addition, researchers could compare a wider range of clustering and topic modelling techniques, or combine quantitative text mining with qualitative approaches, to deepen the theoretical understanding of how user experience, trust, and value perceptions evolve in digital wallet ecosystems.

REFERENCES

Alfa, R., Usman, O., & Musyaffi, A. M. (2025). Pengaruh Service Quality dan Product Quality terhadap Customer Loyalty melalui Customer Satisfaction pada Pengguna E-Commerce di Jabodetabek. *Jurnal Riset Dan Publikasi Ilmu Ekonomi*, *3*(4), 234–244. https://doi.org/10.31000/digibis.v2i1.7855

Amalia, N., Royanti, N. I., Indrayanti, I., & Ismanto, B. (2023). Analisa Sentimen Pelanggan pada Review Belanja Online Berbasis Text Mining Menggunakan Metode K-Means. *Journal of Information System Research (JOSH)*, *4*(4), 1441–1447. https://doi.org/10.47065/josh.v4i4.3781

Anwar, R. N., Gaffar, V., Disman, D., Furqon, C., & Sutisnawati, Y. (2023). Mobile Wallet Adaption Model Among Digital Immigrant Generation in Indonesia. *Journal of Eastern European and Central Asian Research*, *10*(6), 799–809. https://doi.org/10.15549/jeecar.v10i6.1499

Apriyanto, B., & Sitio, S. L. M. (2025). Penerapan K-Means dalam Menganalisis Pola Pembelian Pelanggan Pada Data Transaksi E-Commerce. *Bit-Tech*, *7*(3), 790–797. https://doi.org/10.32877/bt.v7i3.2195

Asep Koswara. (2025). Persaingan Dompet Digital di Indonesia: Analisis Tren Google dan Machine Learning pada GoPay, OVO, dan DANA . *Journal of Economic and Business*, *2*(1), 22–43.

Azarya, Y., & Budi, I. (2025). Analisis Sentimen Berbasis Aspek Aplikasi Brimo berdasarkan Ulasan Pengguna di Google Playstore. *Teh Indonesia Journal of Computer Science*, *14*(1), 1111–1125.

Azizah, N. A., & Ernawati, T. (2024). Analisis Sentimen Ulasan Pelanggan Terhadap Produk Roti Macan Artisan Di E-Commerce Tokopedia Menggunakan Metode Clustering. *Journal of Information System, Applied, Management, Accounting and Research*, *8*(3), 580. https://doi.org/10.52362/jisamar.v8i3.1576

Banutama, B., Hatta, A. J., Biyanto, F., & Kristiana, D. R. (2025). Keputusan Penggunaan E-Wallet Sebagai Alat Transaksi Digital: Sebuah Kajian Literatur 2012-2023. *Jurnal Ilmiah Akuntansi Dan Keuangan (JIAKu)*, *3*(4), 301–318. https://doi.org/10.24034/jiaku.v3i4.7025

Berlianawati, D. I., Nurabiah, N., & Ridhawati, R. (2024). Exploring The Mind of Gen Z: Deciphering E-wallet Adoption Through The Lens of TPB Theory. *Jurnal Ilmiah Akuntansi Dan Bisnis*, *19*(1), 133. https://doi.org/10.24843/jiab.2024.v19.i01.p09

Bimantara, I. M. S., & Widiartha, I. M. (2023). Optimization of K-Means Clustering Using Particle Swarm Optimization Algorithm for Grouping Traveler Reviews Data on Tripadvisor Sites. In *Jurnal Ilmiah Kursor* (Vol. 12, Issue 1, pp. 1–10). https://doi.org/10.21107/kursor.v12i01.269

Diva, M., & Anshori, M. I. (2024). Penggunaan E-Wallet Sebagai Inovasi Transaksi Digital: Literatur Review. *MULTIPLE: Journal of Global and Multidisciplinary*, *2*(6), 1991–2002. https://journal.institercom-edu.org/index.php/multiple

Gosal, Y., Pratiwi, O., Mukti, I., & Witarsyah, D. (2024). Analisis Sentimen Dan Pemodelan Topik Terhadap Aplikasi Pembelajaran Online Pada Platform Google Play (Studi Kasus: Quipper). *Jurnal Ilmiah Komputasi*, *23*(4), 469–478. https://doi.org/10.32409/jikstik.23.4.3663

Jannati, C., Paramitha, E. P., & Nur, M. J. (2024). Persepsi Pengguna Terhadap Aplikasi Tiktok (Analisis Sentimen Menggunakan Metode K-Means Dan Social

Network Analysis). *Jurnal Ilmiah Mahasiswa Komunikasi Universitas Mataram*, *2*(1).

Jatmika, S., Purboningrum, V., & Hidayat, M. T. (2023). Cashless Society Phenomenon in Jakarta City ' s use of The Digital Wallet " DANA " as A Payment Tool. *Internastional Summit on Science Technology and Humanity*, 1149–1155.

Jayanti, T. S., Budiman, B., Habibi, C., & Setiana, E. (2024). Analisis Sentimen Penggunaan Aplikasi Traveloka di Twitter Menggunakan Model Klasifikasi. *SisInfo : Jurnal Sistem Informasi Dan Informatika*, *6*(1), 1–19. https://doi.org/10.37278/sisinfo.v6i1.751

Johan, A. P., Lukviarman, N., & Putra, R. E. (2022). Continuous intention to use e-wallets in Indonesia: The impact of e-wallets features. *Innovative Marketing*, *18*(4), 74–85. https://doi.org/10.21511/im.18(4).2022.07

Khairuna, R., Nurdin, & Ar Razi. (2025). Analisis Perbandingan Algoritma K-Means Dan K-Medoids Untuk Klusterisasi Teks Ulasan Pada Aplikasi Cookpad. *Rabit : Jurnal Teknologi Dan Sistem Informasi Univrab*, *10*(2), 445–458. https://doi.org/10.36341/rabit.v10i2.6131

Naura Nadra Rizki, S., Dwita Puspa Zain, R., Maharani Putri, D., Ekonomi, F., & dan Humaniora, B. (2024). Analisis Kepercayaan Generasi Z Terhadap Dompet Digital (E-Wallet) Berdasarkan Prinsip Tata Kelola. *Proceeding Auditing and …*, 67–77. http://jurnal.iapi.or.id/index.php/prosiding/article/view/86%0Ahttp://jurnal.iapi.or.id/index.php/prosiding/article/download/86/62

Olivia, M. (2022). The Influence of Perceived Ease of Use and Perceived Usefulness on E-Wallet Continuance Intention: Intervening Role of Customer Satisfaction. *Jurnal Teknik Industri*, *24*(1), 13–22. https://doi.org/10.9744/jti.24.1.

Pratiwi, E. (2024). Pengaruh Digital Banking , Fintech Payment , Dan Fintech Lending Terhadap Kinerja Keuangan Perbankan. *Jurnal Akuntansi, Keuangan, Perpajakan Dan Tata Kelola Perusahaan (JAKPT)*, *2*(1), 51–58.

Rahmadan, S. M., & Zihni, N. N. (2025). Penerapan Metode K-Means Pada Klasifikasi Kepuasan Siswa Terhadap Media Pembelajaran Di Sekolah. *Jorunal of Science and Social Research*, *VIII*(3), 3978–3988.

Razaq, S. N. H., Pohan, A. B., Gunawan, D., & Aziz Zuama, R. (2025). Analisis Tingkat Kepuasan Pengguna Aplikasi GoPay, DANA, dan OVO Menggunakan Metode PIECES. *JAIS - Journal of Accounting Information System*, *5*(01), 11–22. https://doi.org/10.31294/jais.v5i01.7687

Rochmanto, H. B., Fitriani, F., & Anuraga, G. (2024). Analisis Perbandingan Ukuran Jarak Dalam Pengelompokan K-Means Pada Ulasan Pengguna Aplikasi Ruangguru Di Google Play. *Seminar Nasional Pendidikan Matematika*, 240–251. https://snpm.unipasby.ac.id/prosiding/index.php/snpm/article/view/257%0Ahttps://snpm.unipasby.ac.id/prosiding/index.php/snpm/article/download/257/209

Safitri Juanita, & Cahyono, R. D. (2024). K-Means Clustering With Comparison of Elbow and Silhouette Methods for Medicines Clustering Based on User Reviews. *Jurnal Teknik Informatika (Jutif)*, *5*(1), 283–289. https://doi.org/10.52436/1.jutif.2024.5.1.1349

Saputri, A. D., & Pratama, A. R. (2021). Klasifikasi Pengguna Lima Dompet Digital Terpopuler di Indonesia dengan Metode MCA. *Jurnal Sosioteknologi*, *20*(2), 138–148.

Sunarko, G. S., Wasino, & Sutrisno, T. (2023). Klasterisasi Sentimen Ulasan Pengguna Aplikasi Bca Mobile Pada Platform Google Play Store Dengan Algoritma K-

Means Clustering. *Jurnal Ilmu Komputer Dan Sistem Informasi*, *11*(1). https://doi.org/10.24912/jiksi.v11i1.24145

Trianto, R. B., Nugroho, A. S., & Supriyadi, E. (2023). Klasterisasi Menggunakan Algoritma K-Means dan Elbow pada Opini Masyarakat Tentang Kebijakan Sekolah Luring Tahun 2022. *INOVTEK Polbeng - Seri Informatika*, *8*(1), 1. https://doi.org/10.35314/isi.v8i1.2756

Turnandes, Y., Afrilli, R., & Kuning, U. L. (2025). Menerapkan Data Science Pada Dataset Review Produk di Shopee dan Tokopedia: Pengelompokan Pelanggan dan Strategi Retensi Dengan Teknik Clustering. *Journal of Science and Social Research*, *VIII*(3), 3908–3914.

Wulandari, P., & Idayanti, R. (2023). Peran Aplikasi Dompet Digital Indonesia (DANA) dalam Memudahkan Masyarakat melakukan Pembayaran Digital. *Islamic Banking and Finance*, *3*(2), 429–441. https://doi.org/10.30863/ibf.v3i2.5438