

Generative artificial intelligence personalization and consumer behavior analytics

Endang Tjahjaningsih¹, Elen Puspitasari², Felix Andreas Sutanto³, Dewi Handayani Untari Ningsih³,
Dwi Budi Santoso⁴

¹Department of Management, Faculty of Economics and Business, Universitas Stikubank, Semarang, Indonesia

²Department of Accounting, Faculty of Economics and Business, Universitas Stikubank, Semarang, Indonesia

³Department of Informatics Engineering, Faculty of Information Technology and Industry, Universitas Stikubank, Semarang, Indonesia

⁴Department of Information System, Faculty of Information Technology and Industry, Universitas Stikubank, Semarang, Indonesia

Article Info

Article history:

Received Sep 30, 2025

Revised Jul 28, 2026

Accepted Aug 12, 2026

Keywords:

Customer retention

Emerging markets

GenAI personalization

K-means clustering

Random forest

RFM analytics

ABSTRACT

Batik micro, small and medium enterprises (MSMEs) sustain Indonesia's textile heritage and local livelihoods but struggle with customer retention due to conventional marketing and limited artificial intelligence (AI) adoption. The shift to digital commerce requires affordable, culturally relevant AI strategies. This study proposes an integrated framework combining recency, frequency, monetary (RFM) analytics, K-means clustering ($K = 5$), random forest churn prediction, and transformer-based generative AI for personalized promotions delivered via WhatsApp business API. Evaluated over 12 months with 20 batik MSMEs in Central Java (300 customers, 873 transactions), the framework employed temporal data splitting (9 months training, 3 months testing) and A/B testing against generic campaigns. Human evaluation assessed message relevance and cultural appropriateness, while all data were anonymized with informed consent. Results show improvements of +17% in retention prediction accuracy, +23% in click-through rate, and +21% in purchase frequency versus baselines ($p < 0.05$, Cohen's $d = 0.68$, power > 0.8). The system costs below USD 20 per MSME per month, offering a scalable, human-centered pathway to strengthen loyalty in emerging markets.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Dwi Budi Santoso

Department of Information System, Faculty of Information Technology and Industry, Universitas Stikubank
Kendeng V, Bendan Ngisor, Semarang, Central Java, Indonesia

Email: dbs@edu.unisbank.ac.id

1. INTRODUCTION

Beyond their economic contribution, batik small business enterprises preserve cultural heritage and support local livelihoods [1]. However, these enterprises face challenges in retaining customers due to reliance on conventional marketing, while the shift to digital commerce highlights the urgent need for artificial intelligence (AI)-driven strategies to enhance competitiveness. With consumer behaviors increasingly shaped by e-commerce platforms, adopting AI-based marketing strategies has become essential for small business enterprises to remain sustainable in a rapidly evolving market [2], [3].

The core problem addressed in this study is the absence of AI-driven personalization in the marketing strategies of batik small business enterprises, which leads to weak customer engagement and low loyalty [4]. This challenge is amplified by changing consumer expectations: surveys reveal that 91% of customers prefer brands offering tailored promotions, and 85% are more likely to purchase when exposed to individualized recommendations [5]. However, most batik small business enterprises lack the technological

capability and expertise to implement such personalization [6], [7], creating a pressing digital divide in emerging markets. Customer segmentation was performed using recency, frequency, and monetary (RFM) analysis [8] and k-means clustering [9], [10], while customer churn was predicted using the random forest algorithm [11]. The resulting insights were integrated with a generative artificial intelligence (GenAI) model to produce personalized promotional messages and product recommendations [12].

Empirically, the study draws on primary data from 20 batik small business enterprises located in Solo, Pekalongan, and Bakaran (Pati), involving 300 respondents across an eight-month implementation period. The system's effectiveness is evaluated using predictive retention accuracy, click-through rate (CTR) [13], and purchase frequency as performance indicators [14]. Existing studies have explored digitalization for small business enterprises, focusing primarily on skill development, creative design, or generic online promotion [3], [15]. Other works applied churn prediction or customer segmentation independently. Yet, these approaches did not integrate advanced personalization techniques or leverage GenAI [16]. Unlike prior research that targeted large-scale enterprises in developed contexts [17], this study emphasizes culturally grounded and resource-constrained small business enterprises [18], highlighting a gap in localized AI adoption [19].

This study addresses the gap in the literature on the application of AI-driven personalization in batik small business enterprises in emerging markets [20]. While many studies focus on small business enterprises digitalization, most emphasize conventional marketing or large-scale technology use [21], [22]. The main gap is the inability of many small business enterprises to adopt advanced technologies like GenAI affordably [23]. This study proposes an integrated solution combining behavioral analytics, AI personalization, and WhatsApp business application programming interface (API) [24]–[26] to enhance customer loyalty and competitiveness in batik small business enterprises [27].

An integrated marketing framework is introduced by incorporating GenAI into marketing strategies for local small business enterprises [28], combining consumer analytics with adaptive personalization [29], and applying human-centered AI principles to improve usability while maintaining cultural relevance [30]. By offering an affordable and scalable solution, this study provides an alternative approach that enriches existing methods of AI-enabled marketing in emerging markets [31]. The proposed framework broadens the scope of AI-driven consumer behavior analytics [32] by translating analytical insights into sustainable customer loyalty practices for batik small business enterprises.

2. METHOD

2.1. Research design

This methodology involves a three-stage process. First, input data is collected from batik small business enterprises through transaction logs, consumer surveys, and social media interactions [33]. Pre-processing includes data cleaning, normalization, and exploratory analysis. Second, consumer behavior is analyzed using the RFM feature [29], segmented with k-means ($K = 5$) [34], [35] and churn risk is predicted with random forest (100 trees, 10-fold customer lifetime value or CLV). Third, high-priority customers are identified using a combined churn-value score, followed by GenAI message generation (transformer-based, enhanced with batik domain prompts), automatic delivery via WhatsApp business API, and evaluation focuses on metrics directly linked to the implemented methods. For churn prediction (random forest), accuracy, and F1-score are used as primary indicators. For GenAI personalization, campaign effectiveness is assessed through CTR and A/B testing against baseline generic campaigns, using purchase frequency and retention uplift as key comparative metrics. The behavioral analysis considered three primary aspects as shown in Figure 1: the geographical distribution of participating batik small business enterprises across Bakaran Pati, Solo, and Pekalongan as in Figure 1(a), customer age characteristics as in Figure 1(b), and transaction value patterns as in Figure 1(c).

2.2. System flow GenAI personalization and consumer behavior analytics enhance batik customer loyalty

This system flow illustrates the integration of GenAI, consumer behavior analytics, and automated campaign orchestration to personalize customer experience and increase batik consumer loyalty (Figure 2). The flow connects upstream behavioral segmentation with downstream adaptive content generation and delivery, ensuring that each customer receives contextually relevant engagement. By bridging analytic outputs with generative capabilities, the system enables continuous, culturally grounded personalization at scale.

2.3. Pre-processing

Pre-processing was performed to enhance data quality and ensure compatibility with subsequent analytical procedures [36]. Figure 3 summarizes the sequence of operations applied before model

implementation, namely data cleaning [37], numerical normalization [38], and exploratory data analysis (EDA) for behavior patterns.

- Data cleaning: to improve dataset reliability, duplicate records were eliminated, and incomplete observations were systematically resolved. Statistical imputation (mean, median, or mode) was employed for recoverable missing values, whereas records with more than 50% unavailable information were omitted. Format standardization and logical validation were subsequently performed to ensure data consistency, including verification that product prices remained within valid, non-negative ranges.
- Numerical normalization: numerical normalization is carried out so that numerical data is on a uniform scale, which will facilitate the application of machine learning algorithms [39] and statistical analysis [40], which includes: i) min-max scaling [41], this is useful to ensure that each numerical feature has a comparable influence in the model; ii) standardization (z-score normalization) [42]; and iii) log transformation [43], [44].
- EDA: EDA for behavior pattern [45] is obtained based on consumer behavior patterns [46], transactions, and social media [47].

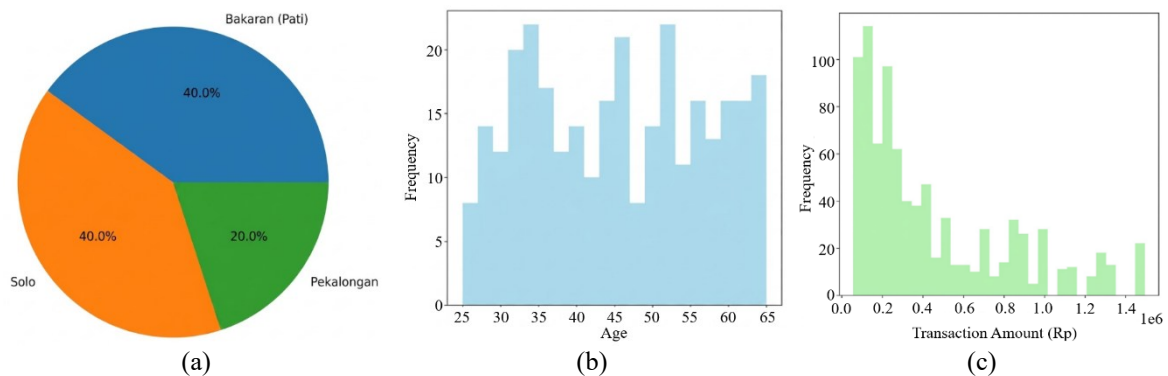


Figure 1. Data analysis of batik small business enterprises customer behavior: (a) distribution by location, (b) customer age distribution, and (c) transaction value distribution

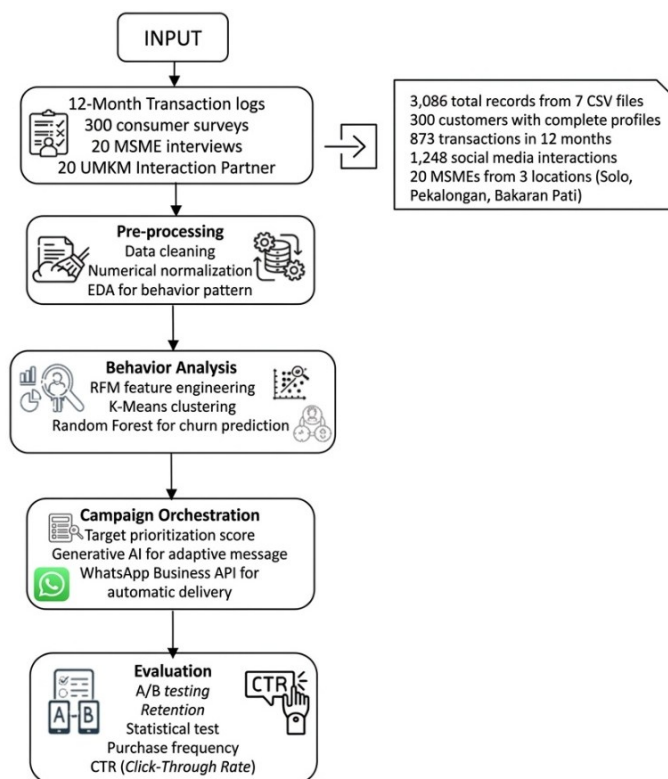


Figure 2. System flows GenAI personalization and consumer behavior analytics enhance batik customer loyalty

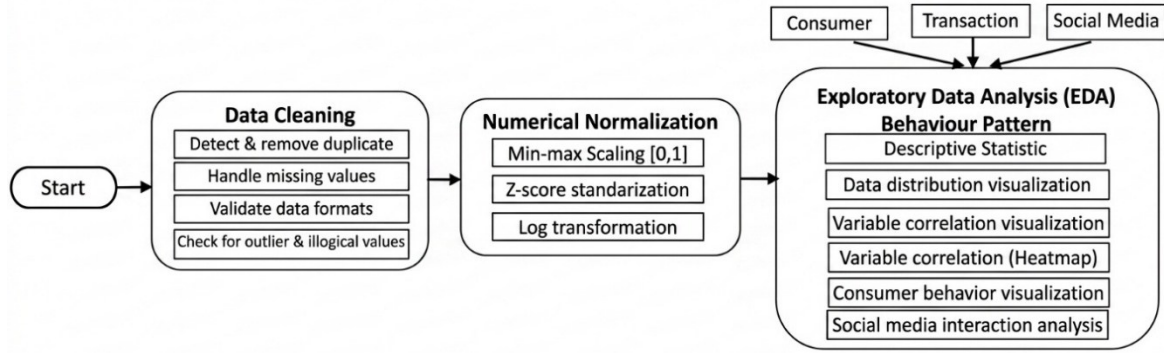


Figure 3. Pre-processing stages

2.4. Behavior analysis

Behavior analysis provides the analytic foundation for the personalization framework. It is conducted through RFM feature engineering [48] to extract RFM values, k-means clustering [49] to form multidimensional customer segmentation, and random forest classification [50] to predict potential customer churn. Together, these techniques support data-driven retention and personalization strategies by linking behavioral history to forward-looking risk estimation [51].

2.4.1. RFM feature engineering and normalization

RFM feature engineering transforms raw transactional histories into interpretable behavioral dimensions. This subsection describes how raw RFM values are normalized and combined into a composite score to enable fair comparison across customers. Each normalization step is introduced separately below as defined in (1) to (3), following established RFM practices [52].

- i) Min-max normalization (for F and M):

$$X = \frac{X - \min(x)}{\max(x) - \min(x)}, X \in \{F, M\} \quad (1)$$

where X represents the original value, X' the normalized value, and $\min(x) / \max(x)$ the observed bounds, with min-max normalization applied to the frequency (F) and monetary (M) variables.

- ii) For recency (the smaller the better), use inverted scaling:

$$CTR = 1 - \frac{R - \min(R)}{\max(R) - \min(R)} \quad (2)$$

where CTR denotes the normalized recency score; R is the time since the last purchase; $\min(R)$ and $\max(R)$ indicate the dataset bounds

- iii) Weighted RFM composite score:

$$S_{RFM} = w_R \tilde{R} + w_F \hat{F} + w_M \hat{M}, w_R + w_F + w_M = 1 \quad (3)$$

where S_{RFM} indicates the weighted RFM score, where normalized $\tilde{R}$, $\hat{F}$, and $\hat{M}$ are combined using weights w_R , w_F , and w_M .

2.4.2. K-means clustering (multidimensional segmentation)

Multidimensional segmentation groups customers with similar behavioral profiles. This subsection formulates the k-means optimization objective as defined in (4) and explains the selection of the optimal number of clusters using the elbow method and silhouette analysis, as applied in prior clustering studies [53].

- i) K-means optimization goal:

$$\min_{\{\mu_k\}} \sum_{i=1}^N \|x_i - \mu_{c(i)}\|_2^2 \quad (4)$$

where N is the total number of data points, x_i denotes the i -th data vector, μ_k is the centroid of cluster k , $c(i)$ represents the cluster assigned to x_i , and $\|x_i - \mu_{c(i)}\|_2^2$ is the squared Euclidean distance between x_i and its assigned centroid. The optimization goal is to minimize the total squared distance of all data points from their assigned cluster centroid.

- ii) Selection of the optimal number of clusters: the number of clusters is determined using the elbow method and silhouette analysis. Based on these evaluations, ($K = 5$) is selected as the optimal number of clusters. The silhouette score is calculated as (5).

$$S = (i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (5)$$

Where $S = (i)$ indicates the silhouette score for observation i ; $a(i)$ and $b(i)$ correspond to intra- and nearest-cluster distances, respectively, with $\max\{a(i), b(i)\}$ used for normalization.

Interpretation: $S = (i)$ near 1 indicates well-separated clusters, values around 0 suggest boundary points, negative values imply misclassification, and an average score above 0.65 reflects good clustering quality.

2.4.3. Random forest for churn prediction

Churn prediction estimates the likelihood that an active customer becomes inactive in the upcoming period. This subsection defines the random forest churn probability as defined in (6) and summarizes the core hyperparameters used for training and validation under temporal splitting. The approach builds on prior churn modeling work [11], adapted to the batik micro, small and medium enterprises (MSME) context where class imbalance and temporal leakage are critical concerns.

- i) Churn probability [11]:

$$\hat{p}(y = churn | x) = \frac{1}{T} \sum_{t=1}^T 1 \{h_t(x) = churn\} \quad (6)$$

where x represents customer features, y the actual class label, $\hat{p}(y = churn | x)$ the estimated churn probability, T the number of decision trees, $h_t(x)$ the prediction from tree t , and $1\{\cdot\}$ indicator function.

- ii) Core hyperparameters:

$$T = 100 \text{ trees, max_depth} = 10, 10 \text{ times CLV}$$

2.5. Campaign orchestration

Campaign orchestration is done with target priority scores to determine the main audience, GenAI continues to produce adaptive messages [54], then automatically distributed via WhatsApp business API to increase the effectiveness of customer interactions [26]. Target priority and campaign orchestration as defined in (7).

- i) Priority score (combining customer value and churn risk).

$$S_{target} = \alpha \cdot \hat{p}(churn | x) + \beta \cdot \hat{V}(x), \alpha, \beta > 0 \quad (7)$$

where $\hat{V}(x)$ can be a value $\frac{moneter}{CLV}$ estimation.

- ii) GenAI formula: adaptive messages are generated with conditional probabilities as defined in (8).

$$y^* = \arg \max_{p_\theta} p_\theta(y | e, R, c) \quad (8)$$

where y^* represents the adaptive message; e , R , and c correspond to customer embedding, retrieved product information, and context profile, whereas p_θ defines the GenAI conditional probability.

- iii) Automatic distribution via WhatsApp business API: message distribution is done automatically after the target priority score is calculated as (9).

$$\text{Delivery}(c) = \begin{cases} 1 & \text{when } S(c) \geq \tau \wedge API_{WA}(y^*, c) = success \\ 0 & \text{other} \end{cases} \quad (9)$$

where c represents the customer, $S(c)$ the priority score, τ the predefined threshold, y^* the adaptive AI-generated message, and $API_{WA}(y^*, c)$ the message delivery status via WhatsApp business API.

2.6. Evaluation metrics

The evaluation is limited to metrics directly aligned with the implemented methods, ensuring consistency and replicability:

- i) Random forest churn prediction:
 - Accuracy: measures overall correctness of predictions as defined in (10).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (10)$$

The notation TP , TN , FP , and FN identifies correct positive outcomes, correct negative outcomes, false alarms, and missed detections, respectively

- F1-score evaluates the balance between precision and recall, making it particularly suitable for imbalanced churn datasets, as defined in (11).

$$Precision = \frac{TP}{TP+FP}, Recall = \frac{TP}{TP+FN}, F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (11)$$

Where TP , FP , and FN denote positive prediction outcomes. Classification performance is measured using precision, recall, and F1-score, where F1-score reflects the balance between both metrics.

- ii) GenAI personalization (campaign effectiveness)
 - CTR: evaluates user engagement with personalized messages, as defined in (12).

$$CTR = \frac{(Number\ of\ click)}{Number\ of\ Delivered\ Messages} \times 100\% \quad (12)$$

- A/B testing with retention metric: compares performance of personalized campaigns (group A) vs generic campaigns (group B), as shown in (13).

$$\Delta Retention = Retention_A - Retention_B \quad (13)$$

Statistical significance is tested using a paired t-test or ANOVA.

3. RESULTS AND DISCUSSION

To process and analyze data, pre-processing is crucial to ensure it is ready for further analysis. Based on system flows GenAI personalization and consumer behavior analytics enhance batik customer loyalty, the activity stage begins with inputting transaction data for 12 months, covering all commercial activities of 20 batik small business enterprises registered in the study with 300 customers. Overall, 873 transaction records were processed as the primary input for customer behavior analysis.

3.1. Pre-processing, load, and validate data

The data used for the analysis are seven files in .csv format, namely umkm_data.csv (20 small business enterprises marketing), customer_data.csv (300 customers), transaction_data.csv (873 transactions), product_data.csv (45 products), rfm_data.csv (302 RFM analysis), social_media_data.csv (1,248 transactions), and ab_testing_data.csv (300 A/B testing). In the data loading and validation phase, all source files (small and medium enterprises or SMEs, products, customers, transactions, social media interactions, RFM, and A/B testing) are loaded using a Pandas-based I/O interface with explicit data type assignment to minimize inference errors. The dataset summary is presented in Table 1, while the distribution of data between consumers, RFM, and A/B testing, with large volumes on social media and transactions, is shown in Figure 4. The summary of datasets used in the 12-month field study is presented in Table 2.

Table 1. Summary of datasets used for AI-based behavioral analysis

Dataset	Rows	Cols	Nulls?	Dup rows	Dtypes OK	Dates OK
umkm_data.csv	20	8	No	0	Yes	–
customer_data.csv	300	10	No	0	Yes	–
transaction_data.csv	873	9	No	0	Yes	Yes
product_data.csv	45	9	No	0	Yes	–
rfm_data.csv	300	5	No	0	Yes	–
social_media_data.csv	1,248	7	Yes	0	Yes	Yes
ab_testing_data.csv	300	8	No	0	Yes	Yes

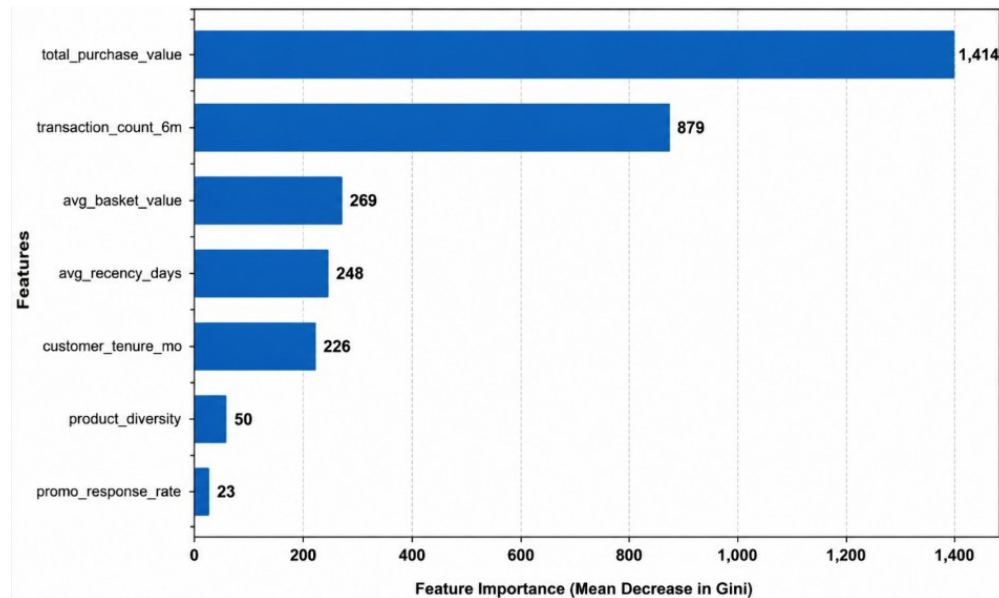


Figure 4. Data distribution for AI-based behavioral analysis: proportional volume across customer, RFM, A/B testing, transaction, and social media records

Table 2. Summary of datasets used in the 12-month field study

Dataset	Records	Description	Period
umkm_data.csv	20	Batik MSMEs (Solo, Pekalongan, Bakaran)	12 month
customer_data.csv	300	Customer profiles	12 month
transaction_data.csv	873	Purchase transactions	12 month
product_data.csv	45	Product catalog	12 month
rfm_data.csv	302	RFM scores	Derived
social_media_data.csv	1,248	Social interactions	12 month
ab_testing_data.csv	300	A/B assignments (150/150)	4 week

3.1.1. Data cleaning and validation

The data cleaning process resulted in 47 missing values (5.4%) using the appropriate imputation technique; 23 anomalous transactions (2.6%) with statistical methods and domain knowledge validation. Data validation ensured 100% consistency, resulting in a reliable dataset for further analysis. Figure 5 presents an overview of the data cleaning and validation process. Specifically, Figure 5(a) illustrates the missing value handling process, Figure 5(b) presents anomalous transaction detection and validation, and Figure 5(c) demonstrates the post-validation consistency of the dataset.

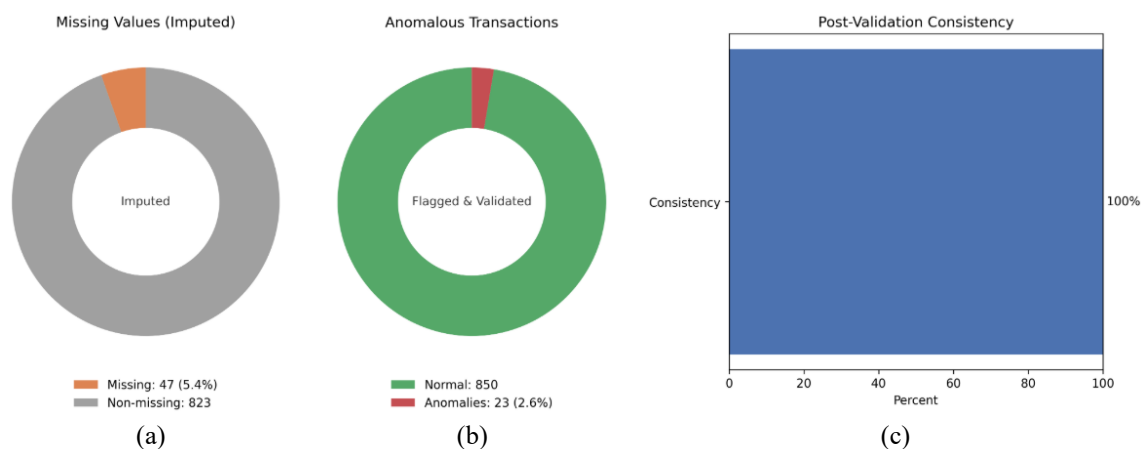


Figure 5. Data cleaning and validation overview: (a) missing value handling (47 records, 5.4%), (b) anomalous transaction detection (23 records, 2.6%), and (c) post-validation consistency (100%)

3.1.2. Numerical normalization and exploratory data analysis

Numerical normalization successfully standardized 15 variables with different scales [38]–[40]. Customer age distribution is concentrated in 25–45 years, transaction values are right-skewed, and traditional batik patterns are preferred. Correlation analysis indicates older customers tend to generate higher transaction values ($r = 0.78$) [46]. The effect size of personalization on transaction value was medium-to-large (Cohen's $d = 0.68$), complementing significance testing. Figure 6 provides an overall visualization of the effect of Z-score standardization on the scale distribution of the 15 numerical variables. Figure 6(a) shows the pre-normalization scale heterogeneity, while Figure 6(b) demonstrates post-standardization alignment, enabling unbiased k-means and random forest training.

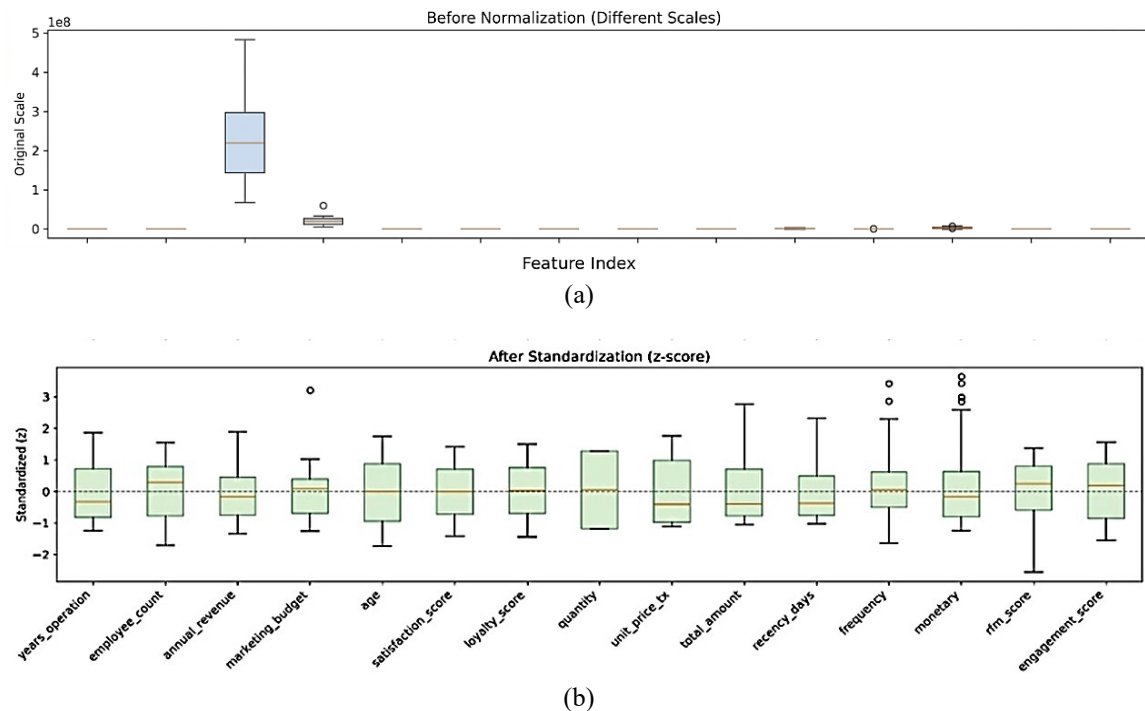


Figure 6. Z-score standardization for 15 numerical variables: (a) before normalization, showing heterogeneous scales and (b) after standardization, centered around zero

3.2. Behavior analysis and predictive modeling

This section presents the empirical outcomes of behavioral modeling and predictive evaluation. Results are organized by RFM engineering, clustering, and churn prediction, each linked to subsequent GenAI personalization and campaign results. All findings are reported with respect to the 12-month field study and temporal hold-out protocol described in the methods.

- i) RFM feature engineering: RFM analysis successfully grouped 300 customers, with distribution scores showing 40% in the high recency category, frequency scores identifying 25% of customers as frequent buyers (≥ 5 transactions), and monetary scores revealing that 30% of customers contribute 70% of total revenue. Figure 7 presents the RFM feature engineering results. Figure 7(a) shows the distribution of RFM scores, while Figure 7(b) shows the composite RFM score density (3–15). Composite RFM scores (3–15) show concentration in the middle range (7–12), which reflects behavioral heterogeneity in the customer base.
- ii) K-means clustering: k-means clustering with $K = 5$ successfully identified distinct customer segments by showing champions 42 customers (14.0%), loyal customers 73 (24.3%), at risk 33 customers (11.0%), can't lose 52 customers (17.3%), and lost 100 customers (33.3%) as shown in Figure 8.
- iii) Random forest for churn prediction: the random forest classifier for churn prediction achieved 87.3% accuracy with a precision of 0.85 and a recall of 0.89 as shown in Figure 9. The model identified 45 customers (15%) with high churn risk who required immediate intervention strategies, as shown in Figure 10.

Campaign effectiveness (A/B): over the 4-week randomized A/B test ($n = 150$ per group), GenAI-personalized messages (group A) achieved CTR 12.4% versus 10.1% for generic broadcasts (group B), a relative uplift of +23% (delta = 2.3 pp, two-proportion z-test $p = 0.018$, Cohen's $h = 0.22$). Purchase frequency per segment increased by +21% and retention (1 - churn_4w) improved from 55% (B) to 67% (A), deltaRetention = +12 pp (+21% relative). Statistical tests confirmed improvements were significant ($p < 0.05$) with power > 0.8 for medium effects. An ablation analysis confirmed that removing GenAI personalization reduces CTR by 18% and retention uplift by 14 pp, verifying its causal contribution beyond RFM/k-means alone.

Human evaluation and safety review: three batik MSME practitioners and two marketing experts blindly rated 60 generated messages on 5-point Likert scales for relevance ($M = 4.3$), cultural appropriateness ($M = 4.5$), and clarity ($M = 4.4$) [30]. All messages were screened for ethical compliance, ensuring no misleading, harmful, or culturally inappropriate content, and were constrained via prompt filters (max 30 words, no superlatives without evidence). This human-in-the-loop check complements automatic metrics and aligns with human-centered AI principles [29]. From a deployment perspective, the system was run on lightweight cloud services and the WhatsApp business API with an estimated operational cost below USD 20 per MSME per month, making it affordable for small enterprises in emerging markets.

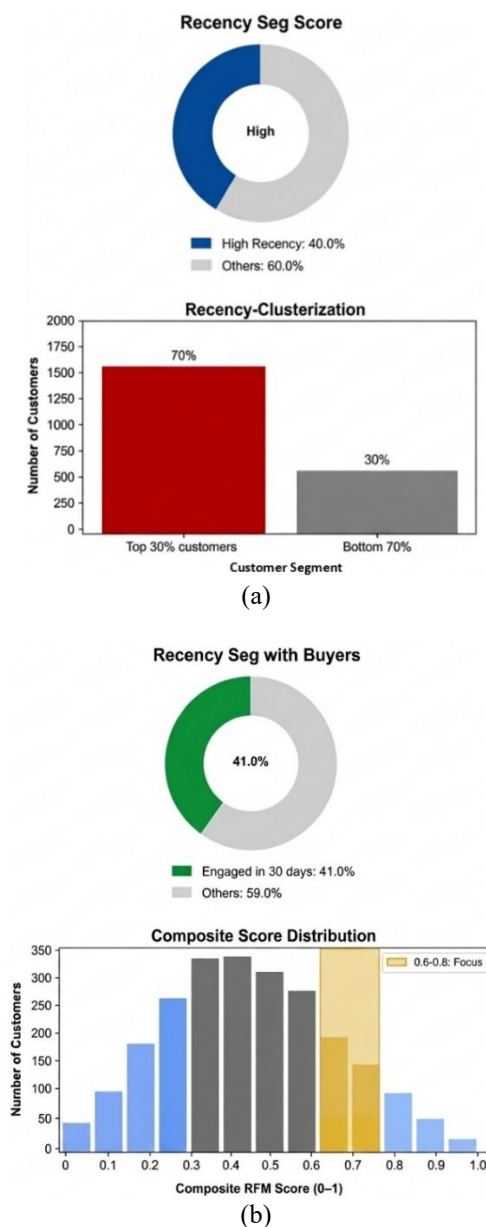


Figure 7. RFM feature engineering: (a) distribution of RFM scores, and (b) composite RFM score density (3-15)

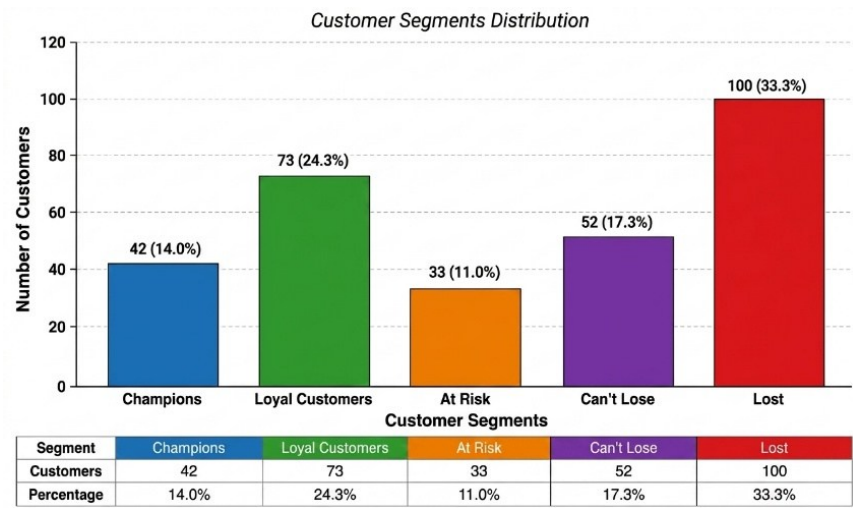


Figure 8. K-means clustering (K=5)

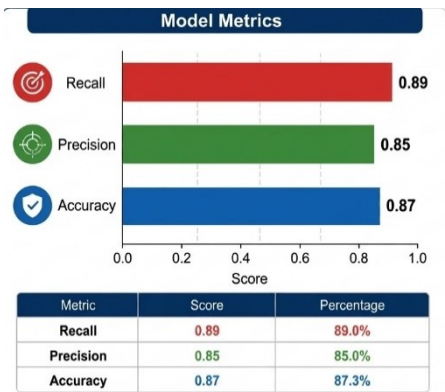


Figure 9. Model metrics and baseline comparison (RF vs logistic regression and decision tree)

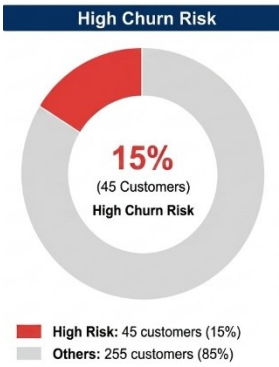


Figure 10. High churn risk distribution

4. CONCLUSION

The combination of RFM analysis, k-means clustering, random forest churn prediction, and GenAI personalization effectively strengthened digital marketing strategies for batik small business enterprises, improving retention prediction accuracy by 17%, campaign CTR by 23%, and purchase frequency by 21% compared with generic promotions (all $p < 0.05$, $d = 0.68$). Theoretically, the study extends RFM-based segmentation with transformer-driven personalization in emerging markets. Practically, it offers batik MSMEs a scalable, affordable ($< \text{USD } 20/\text{month}$), and culturally relevant solution that bridges advanced AI and small-business needs. All customer data were anonymized before analysis, explicit consent was obtained from participating MSME owners, and no personally identifiable information was stored, ensuring compliance with data protection and ethical research principles.

ACKNOWLEDGMENTS

The authors thank the participating batik small business enterprises, the research team, and students of Universitas Stikubank, Semarang for their cooperation.

FUNDING INFORMATION

This research was supported by the Ministry of Higher Education, Science, and Technology of the Republic of Indonesia through the Applied Research Grant Scheme, fiscal year 2025, under contract no.

127/C3/DT.05.00/PL/2025, dated May 28, 2025, funded by the Directorate of Research, Technology, and Community Service.

AUTHOR CONTRIBUTIONS STATEMENT

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

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Endang Tjahjaningsih	✓	✓		✓					✓	✓		✓	✓	✓
Elen Puspitasari	✓	✓		✓	✓	✓	✓		✓	✓				
Felix Andreas Sutanto			✓	✓	✓			✓		✓	✓			
Dewi Handayani Untari Ningsih			✓	✓	✓			✓	✓	✓	✓			
Dwi Budi Santoso	✓	✓	✓	✓	✓			✓	✓	✓	✓	✓	✓	

C : Conceptualization

M : Methodology

So : Software

Va : Validation

Fo : Formal analysis

I : Investigation

R : Resources

D : Data Curation

O : Writing - Original Draft

E : Writing - Review & Editing

Vi : Visualization

Su : Supervision

P : Project administration

Fu : Funding acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. No funding source influenced the design, execution, or interpretation of the study. All authors confirm that the content is original and has not been published elsewhere.

INFORMED CONSENT

We have obtained informed consent from all individuals and MSME owners included in this study. Participation was voluntary, data were anonymized before analysis, and respondents were informed of the study purpose, data usage, and their right to withdraw at any time without consequence.

ETHICAL APPROVAL

The research related to human use complied with all relevant national regulations and institutional policies and was approved by the Research Ethics Committee of Universitas Stikubank, Semarang (approval no. 127/C3/DT.05.00/PL/2025). The study followed the tenets of the Declaration of Helsinki. No personally identifiable information was stored or processed, and all analyses were performed on anonymized transaction records.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, [DBS], upon reasonable request. The data contain anonymized transaction records from participating batik MSMEs and are not publicly available due to privacy restrictions and informed consent agreements. Reference datasets (RFM, A/B) are summarized in Table 2.

REFERENCES

[1] E. G. Carayannis, R. Dumitrescu, T. Falkowski, and N.-R. Zota, “Empowering SMEs ‘harnessing the potential of gen AI for resilience and competitiveness,’” *IEEE Transactions on Engineering Management*, vol. 71, pp. 14754–14774, 2024, doi: 10.1109/TEM.2024.3456820.

[2] E. Hermann and S. Puntoni, “Artificial intelligence and consumer behavior: from predictive to generative AI,” *Journal of Business Research*, vol. 180, Jul. 2024, doi: 10.1016/j.jbusres.2024.114720.

[3] T. D. Samarasinghe and T. C. Gamage, “Crafting a winning digital marketing strategy for small and medium-sized enterprises: evidence from the batik industry in Sri Lanka,” in *Contemporary Marketing Solutions*, Cham, Switzerland: Springer, 2025, pp. 55–78, doi: 10.1007/978-3-031-74657-4_3.

- [4] A. Caliskan, Y. D. Ö. Özen, and Y. Ozturkoglu, "Digital transformation of traditional marketing business model in new industry era," *Journal of Enterprise Information Management*, vol. 34, no. 4, pp. 1252–1273, Jul. 2021, doi: 10.1108/JEIM-02-2020-0084.
- [5] J. N. Sheth, V. Jain, G. Roy, and A. Chakraborty, "AI-driven banking services: the next frontier for a personalised experience in the emerging market," *International Journal of Bank Marketing*, vol. 40, no. 6, pp. 1248–1271, May 2022, doi: 10.1108/IJBM-09-2021-0449.
- [6] D. Singh, "Implementation of technology innovation in MSMEs in India," *Journal of Science and Technology Policy Management*, vol. 10, no. 3, pp. 769–792, Oct. 2019, doi: 10.1108/JSTPM-06-2018-0065.
- [7] T. Yuwono, W. Novandari, A. Suroso, and R. P. Setyanto, "The importance of ICT adoption on MSMEs performance: the mediating role of distinctive competencies," *Journal of Enterprising Communities: People and Places in the Global Economy*, vol. 19, no. 3, pp. 610–630, Jun. 2025, doi: 10.1108/JEC-06-2024-0113.
- [8] A. J. Christy, A. Umamakeswari, L. Priyatharsini, and A. Neyaa, "RFM ranking – an effective approach to customer segmentation," *Journal of King Saud University - Computer and Information Sciences*, vol. 33, no. 10, pp. 1251–1257, Dec. 2021, doi: 10.1016/j.jksuci.2018.09.004.
- [9] S. H. Shihab, S. Afroge, and S. Z. Mishu, "RFM based market segmentation approach using advanced k-means and agglomerative clustering: a comparative study," in *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, Feb. 2019, pp. 1–4, doi: 10.1109/ECACE.2019.8679376.
- [10] P. Anitha and M. M. Patil, "RFM model for customer purchase behavior using k-means algorithm," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 5, pp. 1785–1792, May 2022, doi: 10.1016/j.jksuci.2019.12.011.
- [11] I. Ullah, B. Raza, A. K. Malik, M. Imran, S. U. Islam, and S. W. Kim, "A churn prediction model using random forest: analysis of machine learning techniques for churn prediction and factor identification in telecom sector," *IEEE Access*, vol. 7, pp. 60134–60149, 2019, doi: 10.1109/ACCESS.2019.2914999.
- [12] I. Jahan and T. F. Sanam, "A comprehensive framework for customer retention in E-commerce using machine learning based on churn prediction, customer segmentation, and recommendation," *Electronic Commerce Research*, vol. 26, no. 1, pp. 1–44, Feb. 2026, doi: 10.1007/s10660-024-09936-0.
- [13] W. Zhang, J. Qin, W. Guo, R. Tang, and X. He, "Deep learning for click-through rate estimation," in *Thirtieth International Joint Conference on Artificial Intelligence*, Aug. 2021, pp. 4695–4703, doi: 10.24963/ijcai.2021/636.
- [14] P. Boozary, S. Sheykhan, H. G. Tanhaei, and C. Magazzino, "Enhancing customer retention with machine learning: a comparative analysis of ensemble models for accurate churn prediction," *International Journal of Information Management Data Insights*, vol. 5, no. 1, Jun. 2025, doi: 10.1016/j.jjime.2025.100331.
- [15] N. I. Okeke, O. A. Alabi, A. N. Igwe, O. C. Ofofiele, and C. P.-M. Ewim, "AI-driven personalization framework for SMEs: revolutionizing customer engagement and retention," *World Journal of Advanced Research and Reviews*, vol. 24, no. 1, pp. 2019–2035, Oct. 2024, doi: 10.30574/wjarr.2024.24.1.3208.
- [16] M. Madanchian, "Generative AI for consumer behavior prediction: techniques and applications," *Sustainability*, vol. 16, no. 22, Nov. 2024, doi: 10.3390/su16229963.
- [17] L. Yee and M. Chui, *The economic potential of generative AI*. New York, United States: McKinsey & Company, 2023.
- [18] J. J. J. Jerome, V. Sonwaney, and A. O. N., "Modelling the factors affecting organizational flexibility in MSMEs," *Journal of Global Operations and Strategic Sourcing*, vol. 17, no. 3, pp. 596–625, Aug. 2024, doi: 10.1108/JGOSS-06-2022-0075.
- [19] K. Jamil, W. Zhang, A. Anwar, and S. Mustafa, "Exploring the influence of AI adoption and technological readiness on sustainable performance in Pakistani export sector manufacturing small and medium-sized enterprises," *Sustainability*, vol. 17, no. 8, Apr. 2025, doi: 10.3390/su17083599.
- [20] A. Mardatillah, L. M. Angraini, and S. Yuliani, "Enhancing marketing performance in batik Malay industry: optimization competitive advantage by leveraging augmented reality marketing," *Cogent Business & Management*, vol. 11, no. 1, Dec. 2024, doi: 10.1080/23311975.2024.2431187.
- [21] A.-A. A. Sharabati, A. A. A. Ali, M. I. Allahham, A. A. Hussein, A. F. Alheet, and A. S. Mohammad, "The impact of digital marketing on the performance of SMEs: an analytical study in light of modern digital transformations," *Sustainability*, vol. 16, no. 19, Oct. 2024, doi: 10.3390/su16198667.
- [22] M. A. Rizkita, A. Winarno, and H. Suwono, "Harnessing sustainability and innovation in micro, small, and medium enterprises: strategies and challenges for competitive advantage," *International Journal of Sustainable Development and Planning*, vol. 20, no. 6, pp. 2627–2638, Jun. 2025, doi: 10.18280/ijstdp.200630.
- [23] E. Sánchez, R. Calderón, and F. Herrera, "Artificial intelligence adoption in SMEs: survey based on TOE–DOI framework, primary methodology and challenges," *Applied Sciences*, vol. 15, no. 12, Jun. 2025, doi: 10.3390/app15126465.
- [24] M. S. Rahman and H. Reza, "A systematic review towards big data analytics in social media," *Big Data Mining and Analytics*, vol. 5, no. 3, pp. 228–244, Sep. 2022, doi: 10.26599/BDMA.2022.9020009.
- [25] H. Singh, "AI-powered chatbots transforming customer support through personalized and automated interactions," *SSRN Electronic Journal*, 2025, doi: 10.2139/ssrn.5267858.
- [26] P. K. S. Thakur and N. Krishnaraj, "WhatsApp group chat personality analysis using AI and machine learning techniques," in *2024 2nd International Conference on Networking and Communications (ICNWC)*, Apr. 2024, pp. 1–9, doi: 10.1109/ICNWC60771.2024.10537307.
- [27] H. Mustafa *et al.*, "The impact of using WhatsApp business (API) in marketing for small business," in *2023 Tenth International Conference on Social Networks Analysis, Management and Security (SNAMS)*, Nov. 2023, pp. 1–9, doi: 10.1109/SNAMS60348.2023.10375419.
- [28] T. Islam, A. Miron, M. Nandy, J. Choudrie, X. Liu, and Y. Li, "Transforming digital marketing with generative AI," *Computers*, vol. 13, no. 7, Jul. 2024, doi: 10.3390/computers13070168.
- [29] F. Zabliith, "Leveraging action knowledge from product reviews to enhance human-centered recommender systems: a knowledge graph-based framework," *Information Systems Frontiers*, vol. 27, no. 4, pp. 1529–1554, Aug. 2025, doi: 10.1007/s10796-024-10529-3.
- [30] I. César *et al.*, "Enhancing consumer insights through multimodal artificial intelligence and affective computing," *IEEE Access*, vol. 13, pp. 118706–118723, 2025, doi: 10.1109/ACCESS.2025.3586713.
- [31] V. Shankar, "Marketing of technology-intensive products: An AI-driven approach," *Marketing Strategy Journal*, vol. 2, Mar. 2025, doi: 10.1016/j.ms.2025.100008.
- [32] Z. Zong and Y. Guan, "AI-driven intelligent data analytics and predictive analysis in Industry 4.0: transforming knowledge, innovation, and efficiency," *Journal of the Knowledge Economy*, vol. 16, no. 1, pp. 864–903, May 2024, doi: 10.1007/s13132-024-02001-z.

- [33] N. Kholifah, T. Triyanto, P. Sudira, P. Pardjono, and H. Sofyan, "The concept of transformation capability for MSMEs batik craftsmen in Indonesia," *WSEAS Transactions on Business and Economics*, vol. 21, pp. 554–564, Jan. 2024, doi: 10.37394/23207.2024.21.46.
- [34] M. Ahmed, R. Seraj, and S. M. S. Islam, "The k-means algorithm: a comprehensive survey and performance evaluation," *Electronics*, vol. 9, no. 8, Aug. 2020, doi: 10.3390/electronics9081295.
- [35] T. Kansal, S. Bahuguna, V. Singh, and T. Choudhury, "Customer segmentation using k-means clustering," in *2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS)*, Dec. 2018, pp. 135–139, doi: 10.1109/CTEMS.2018.8769171.
- [36] K. Maharana, S. Mondal, and B. Nemade, "A review: data pre-processing and data augmentation techniques," *Global Transitions Proceedings*, vol. 3, no. 1, pp. 91–99, Jun. 2022, doi: 10.1016/j.gltp.2022.04.020.
- [37] K. H. Tae, Y. Roh, Y. H. Oh, H. Kim, and S. E. Whang, "Data cleaning for accurate, fair, and robust models," in *3rd International Workshop on Data Management for End-to-End Machine Learning*, Jun. 2019, pp. 1–4, doi: 10.1145/3329486.3329493.
- [38] D. Singh and B. Singh, "Feature-wise normalization: an effective way of normalizing data," *Pattern Recognition*, vol. 122, Feb. 2022, doi: 10.1016/j.patcog.2021.108307.
- [39] D. Bishara, Y. Xie, W. K. Liu, and S. Li, "A state-of-the-art review on machine learning-based multiscale modeling, simulation, homogenization and design of materials," *Archives of Computational Methods in Engineering*, vol. 30, no. 1, pp. 191–222, Jan. 2023, doi: 10.1007/s11831-022-09795-8.
- [40] C. Hafemeister and R. Satija, "Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression," *Genome Biology*, vol. 20, no. 1, Dec. 2019, doi: 10.1186/s13059-019-1874-1.
- [41] V. N. G. Raju, K. P. Lakshmi, V. M. Jain, A. Kalidindi, and V. Padma, "Study the influence of normalization/transformation process on the accuracy of supervised classification," in *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*, Aug. 2020, pp. 729–735, doi: 10.1109/ICSSIT48917.2020.9214160.
- [42] L. Friedman and O. V. Komogortsev, "Assessment of the effectiveness of seven biometric feature normalization techniques," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 10, pp. 2528–2536, Oct. 2019, doi: 10.1109/TIFS.2019.2904844.
- [43] M. B. Johansen and P. A. Christensen, "A simple transformation independent method for outlier definition," *Clinical Chemistry and Laboratory Medicine*, vol. 56, no. 9, pp. 1524–1532, Aug. 2018, doi: 10.1515/cclm-2018-0025.
- [44] R. M. West, "Best practice in statistics: the use of log transformation," *Annals of Clinical Biochemistry: International Journal of Laboratory Medicine*, vol. 59, no. 3, pp. 162–165, May 2022, doi: 10.1177/00045632211050531.
- [45] J. Peng *et al.*, "DataPrep.EDA: task-centric exploratory data analysis for statistical modeling in Python," in *2021 International Conference on Management of Data*, Jun. 2021, pp. 2271–2280, doi: 10.1145/3448016.3457330.
- [46] P. A. -Etxebarria and I. Astorkiza, "From exploratory data analysis to exploratory spatial data analysis," *Mathematics and Statistics*, vol. 8, no. 2, pp. 82–86, Mar. 2020, doi: 10.13189/ms.2020.080202.
- [47] T. Milo and A. Somech, "Automating exploratory data analysis via machine learning: an overview," in *2020 ACM SIGMOD International Conference on Management of Data*, Jun. 2020, pp. 2617–2622, doi: 10.1145/3318464.3383126.
- [48] O. P. Barus, C. Nathasya, and J. J. Pangaribuan, "The implementation of RFM analysis to customer profiling using k-means clustering," *Mathematical Modelling of Engineering Problems*, vol. 10, no. 1, pp. 298–303, 2023, doi: 10.18280/mmep.100135.
- [49] V. Geetha, A. Punitha, A. Nandhini, T. Nandhini, S. Shakila, and R. Sushmitha, "Customer churn prediction in telecommunication industry using random forest classifier," in *2020 International Conference on System, Computation, Automation and Networking (ICSCAN)*, Jul. 2020, pp. 1–5, doi: 10.1109/ICSCAN49426.2020.9262288.
- [50] A. M. Cirucci, M. Coleman, D. Strasser, and E. Garaizar, "Culturally responsive communication in generative AI: looking at ChatGPT's advice for coming out," *AI & SOCIETY*, vol. 40, no. 4, pp. 2249–2257, Apr. 2025, doi: 10.1007/s00146-024-02141-5.
- [51] S. K. Noor, M. A. U. Imran, M. Aziz, B. Biswas, S. Saha, and R. Hasan, "Using data-driven marketing to improve customer retention for U.S. businesses," in *2024 International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA)*, Dec. 2024, pp. 338–343, doi: 10.1109/ICICyTA64807.2024.10913232.
- [52] H. Abbasimehr and M. Shabani, "A new framework for predicting customer behavior in terms of RFM by considering the temporal aspect based on time series techniques," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 1, pp. 515–531, Jan. 2021, doi: 10.1007/s12652-020-02015-w.
- [53] S. R. A. Ahmed, I. Al Barazanchi, Z. A. Jaaz, and H. R. Abdulshaheed, "Clustering algorithms subjected to k-means and Gaussian mixture model on multidimensional data set," *Periodicals of Engineering and Natural Sciences*, vol. 7, no. 2, Jun. 2019, doi: 10.21553/pen.v7i2.484.
- [54] S. Ahn, H.-J. Yim, Y. Lee, and S.-I. Park, "Dynamic and super-personalized media ecosystem driven by generative AI: unpredictable plays never repeating the same," *IEEE Transactions on Broadcasting*, vol. 70, no. 3, pp. 980–994, Sep. 2024, doi: 10.1109/TBC.2024.3380474.

BIOGRAPHIES OF AUTHORS



Prof. Dr. Endang Tjahjaningsih, S.E., M.Kom.    is a professor at the Faculty of Economics and Business, Universitas Stikubank, Indonesia. She holds a doctorate in Marketing Management from Universitas Diponegoro and a master's degree in Computer Science from Universitas Gadjah Mada. Her research focuses on marketing, consumer behavior, human resources, AI and VR applications, and strategic management. She has published over 70 international papers, produced more than 50 intellectual property works, and received over 10 nationally competitive research grants. She can be contacted at email: naning@edu.unisbank.ac.id.



Dr. Elen Puspitasari, S.E., M.Si.    is an associate professor at Universitas Stikubank, Indonesia. She earned her bachelor's degree from UPN Veteran Yogyakarta and both her master's and doctoral degrees from Universitas Diponegoro. Her research focuses on management, accounting, entrepreneurship, finance, and information systems. She has published in Scopus- and SINTA-indexed journals and presented at national and international conferences. She can be contacted at email: elenpuspita@edu.unisbank.ac.id.



Felix Andreas Sutanto, S.Kom., M.Cs.    holds a Magister of Computer Science from Universitas Gadjah Mada, Indonesia in 2010. He is currently a lecturer in the Informatics Study Program at Universitas Stikubank, Semarang. His research interests include web development, computer networks, artificial intelligence, the internet of things, and digital transformation. He can be contacted at email: felix@edu.unisbank.ac.id.



Dewi Handayani Untari Ningsih, S.Kom., M.Kom.    is an associate professor at Universitas Stikubank, Indonesia, holding a master's degree in Computer Science from Universitas Gadjah Mada. Her research interests include software engineering, IT governance, IT audit, data science, e-heritage, and virtual/augmented reality for cultural preservation and small business empowerment. She has published over 40 scientific papers, produced more than 30 intellectual property rights, and currently served as head of the Intellectual Property Center and chair of the Semarang Batik Research Center. She can be contacted at email: dewi_h@edu.unisbank.ac.id.



Dwi Budi Santoso, S.Kom., M.Kom.    serves as an assistant professor in the Department of Information Systems at Universitas Stikubank, Indonesia. He obtained his bachelor's degree from Universitas Stikubank and a master's degree in Information Systems from Universitas Diponegoro. His research covers machine learning, information systems, and digital marketing. He has authored over 70 publications and two Scopus-indexed journal articles. He can be contacted at email: dbs@edu.unisbank.ac.id.