Journal of Mathematics and Scientific Computing with Applications

E-ISSN : 2580-5754; P-ISSN : 2580-569X
Volume 7, Number 1, 2026
DOI: 10.53806/jmscowa.v7i1.1582
Page: 207–220

MARKET SEGMENTATION OF STUDENT E-WALLET USERS USING K-MEANS CLUSTERING AND RANDOM FOREST

Diah Putri Kartikasari[1], **Ilka Zufria**[2]

[1,2]Department of Computer Science, State Islamic University of North Sumatra, Medan, 20353, Indonesia

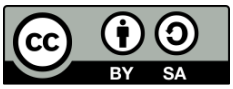
Article Info

Article history:
Received : May 23, 2026
Revised : June 25, 2026
Accepted : July 30, 2026

Keywords:
E-wallet;
K-Means;
Market Segmentation;
Random Forest;
Students.

ABSTRACT

The rapid growth of e-wallet use among students creates a need for computational workflows that explore variation in survey-based behavior and perception. This study aimed to segment student e-wallet users and identify features associated with the resulting clusters. Data were collected through an online questionnaire, and 569 valid responses were analyzed using 13 behavioral and perception features. K-Means produced two exploratory clusters: High-Perception Promo-Responsive Users (86.3%) and Lower-Perception Selective Users (13.7%). Because internal indices disagreed and the Silhouette score of 0.568 indicated only moderate separation, $k = 2$ was retained as an interpretability-oriented judgment; Silhouette and Davies-Bouldin supported $k = 2$, whereas Calinski-Harabasz and the elbow pattern supported k = 3. Random Forest was applied as a post-clustering interpretability tool to examine label reproducibility and identify influential features. Ease of use, trust, promotion and incentives, security, and risk perception were dominant, but the self-reported data require cautious interpretation.



Corresponding Author:
Diah Putri Kartikasari,
Computer Science,
State Islamic University of North Sumatra
Email: dputrikss@gmail.com

1. INTRODUCTION

The development of the digital economy in Indonesia has significantly changed the way people conduct financial transactions, particularly through the increasing adoption of digital payment systems. E-wallet services have become one of the most widely used forms of digital payment because they provide convenience, speed, accessibility, and integration with various online and offline transaction activities. In Indonesia, the growth of electronic money and QRIS transactions shows that digital payment services are no longer merely supporting tools, but have become an important part of daily economic activities [1]. This development is also supported by the expansion of

application-based digital services, which has encouraged wider e-wallet use among productive-age users, including university students [2].

Students represent an important user group in the e-wallet market because their daily activities are closely related to digital technology and mobile-based services. E-wallets are commonly used by students for purchasing food, using transportation services, shopping online, and paying for digital services. However, student users cannot be viewed as a homogeneous group. Differences in economic background, lifestyle, transaction frequency, promotion preferences, perceived security, trust, and risk perception may create different patterns of e-wallet usage. Some students may use e-wallets intensively because of convenience and promotional benefits, while others may be more selective because of concerns about security, service reliability, or transaction risk [3], [4].

Previous studies on e-wallets and digital payment services have generally focused on technology adoption, intention to use, user acceptance, satisfaction, and engagement. Several studies found that perceived usefulness, ease of use, trust, security, and lifestyle are important factors in explaining e-wallet adoption and user behavior [5], [6]. Other studies applied machine learning methods to analyze digital wallet acceptance and predict digital payment adoption [7], [8]. Although these studies provide important insights into digital payment behavior, most of them remain directed toward adoption prediction or acceptance analysis, rather than demonstrating a complete computational workflow for survey-based user segmentation.

Recent studies have also demonstrated the relevance of segmentation and interpretable machine learning in digital consumer analysis. Jaiswal et al [9]. profiled fintech-led mobile payment users and showed that digital payment users can be grouped into distinct market profiles. Joung and Kim [10] highlighted the value of interpretable machine learning for explaining customer segment characteristics. Meanwhile, Ch'ng [11] applied a hybrid machine learning approach to analyze e-wallet adoption among higher education students. However, these studies differ from the present study in computational emphasis. They do not specifically demonstrate a survey-based segmentation workflow that combines clustering, sensitivity analysis, initialization stability testing, and post-clustering feature interpretation for student e-wallet users.

Therefore, this study frames student e-wallet segmentation as an exploratory computational workflow rather than as a final validated market classification. The workflow combines K-Means clustering, internal cluster evaluation, scaling sensitivity analysis, initialization stability testing, and Random Forest-based post-clustering interpretation. This design allows the study not only to form user segments, but also to examine whether the selected cluster structure is robust across preprocessing and initialization choices and to identify which features contribute most to the resulting labels.

The computational contribution of this study lies in demonstrating a reproducible workflow for clustering survey-based e-wallet user data while explicitly discussing methodological limitations. These limitations include the use of Euclidean distance on ordinal-encoded behavioral variables, the possibility of common-method variance in self-reported perception items, and the exploratory nature of the resulting cluster labels. Thus, the study contributes a methodological workflow for interpreting survey-based segmentation results cautiously, rather than claiming externally validated market segments.

This study was guided by three research questions: (RQ1) What student e-wallet user segments can be identified using behavioral and perception-based features? (RQ2) How robust is the selected cluster solution across scaling and initialization settings? and (RQ3) Which features are most strongly associated with the resulting cluster labels? Based on these questions, this study aimed to apply and evaluate a survey-based segmentation workflow using K-Means clustering, robustness checks, and Random Forest-based post-clustering interpretation.

2. RESEARCH METHOD

This study used a quantitative survey design combined with data mining and machine learning analysis. The quantitative approach was selected because the data were collected in a structured form through a questionnaire and analyzed numerically to identify segmentation patterns among student e-wallet users. The study followed a post-clustering interpretation framework in which K-Means was used as an unsupervised learning algorithm to form user segments, while Random Forest was used after clustering to examine the feature patterns associated with the K-Means segment labels and to identify the main distinguishing features between segments. The general research procedure consisted of planning, data collection, data preprocessing, instrument testing, feature construction, clustering, cluster evaluation, post-clustering interpretation, and result analysis. The overall computational workflow of the study is shown in Figure 1.

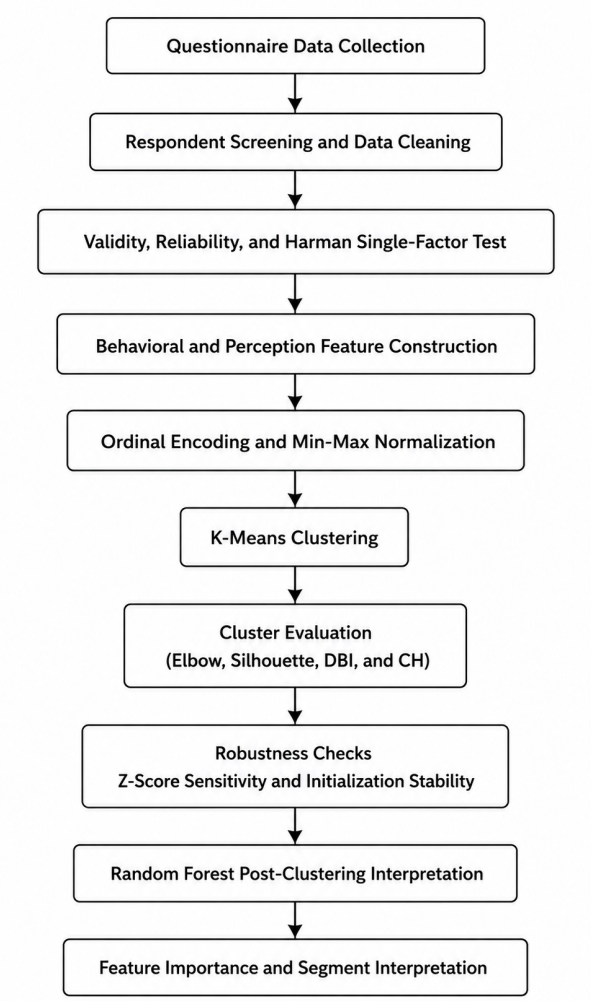


Figure 1. Computational Workflow of the Study

The data used in this study were primary data collected directly from student e-wallet users through an online questionnaire. The population consisted of active university students who used e-wallet services. The sampling technique used was purposive sampling with a quota-oriented approach because the study targeted respondents with specific characteristics relevant to the research objective. Respondents were included when they met three criteria: they were active university students, they used at least one e-wallet service, and they had conducted transactions using an e-wallet within the last three months. Screening questions were provided at the beginning of the questionnaire to ensure that only eligible respondents entered the research dataset.

The questionnaire collected three main groups of variables: demographic variables, e-wallet usage behavior variables, and user perception variables. Demographic variables were used to describe the basic profile of respondents. Behavioral variables described actual usage patterns, including frequency of use, number of active e-wallets, duration of use, transaction amount, monthly e-wallet spending, influence of promotions, and frequency of switching e-wallets because of promotions. User perception variables were measured using a five-point Likert scale and consisted of six constructs: perceived usefulness, ease of use, security, trust, promotion and incentives, and risk perception. Each perception construct consisted of four item indicators.

The questionnaire items for the perception constructs were developed by adapting dimensions commonly used in e-wallet and digital payment studies, including perceived usefulness, ease of use, security, trust, promotion and incentives, and risk perception. The item wording was adjusted to fit the context of student e-wallet users. The complete questionnaire items for the perception constructs are provided in Appendix B to support instrument transparency and reproducibility. Because the data were collected using a self-reported questionnaire, the responses may be affected by reporting bias and common-method bias. This limitation was considered when interpreting the segmentation results.

Data collection was conducted online using a digital form. The initial dataset contained 571 responses. After screening, 569 responses were retained as valid data because they met the respondent criteria, while 2 responses were excluded because they did not meet the research requirements. The minimum sample size had been set at 385 valid respondents based on a large-population proportion formula with a 95% confidence level, a 5% margin

of error, and an assumed proportion of 0.5. Therefore, the final valid dataset exceeded the minimum sample requirement.

Before the perception items were used as modeling features, validity and reliability tests were conducted. The validity test was performed using corrected item-total correlation to examine whether each item was suitable for representing its corresponding construct. Items were considered valid when their correlation values exceeded the minimum validity criterion. Reliability was examined using Cronbach's Alpha to measure the internal consistency of each construct. After the items met the validity and reliability criteria, each perception construct score was calculated as the mean value of its four indicators. Thus, 24 perception items were transformed into six perception features.

In addition, Harman's single-factor test was conducted as an indicative diagnostic for common-method bias because all perception items were collected using the same questionnaire instrument. The test was performed using principal component analysis without rotation on the 24 perception items. The first unrotated factor was examined to determine whether a single general factor accounted for a dominant proportion of the total variance.

Data preprocessing was conducted to transform raw questionnaire data into a clean and structured dataset suitable for clustering and post-clustering analysis. The preprocessing stages included respondent eligibility screening, removal of identity columns, category text cleaning, missing value and duplicate checking, outlier handling, ordinal encoding, construction of perception scores, and feature normalization. Identity-related columns such as timestamp and email address were removed because they were not relevant to the modeling process and were related to respondent privacy. Each respondent was then assigned an anonymous respondent identifier. A detected outlier in the number of active e-wallets was corrected by referring to the multi-select e-wallet response. Ordinal categorical variables were converted into numerical values according to their ordered levels.

Feature selection was guided by both theoretical relevance and clustering suitability. The seven behavioral features represented observable e-wallet usage patterns, while the six perception features represented construct-level evaluations of e-wallet services. Demographic and nominal variables, such as gender, main e-wallet, and dominant transaction type, were not included in the K-Means distance calculation because their numerical codes did not represent meaningful ordinal distances. They were retained only for descriptive profiling. The perception constructs were represented using mean scores to reduce item-level redundancy and to ensure that each construct contributed as a single conceptual feature in the clustering model.

The behavioral and perception features were combined in the same model space after numerical transformation and normalization. Min-Max normalization was selected because the dataset contained encoded ordinal variables and Likert-based construct means with bounded but different numerical ranges. Scaling all features into a comparable 0-1 range reduced the risk that variables with larger numerical values would dominate the Euclidean distance calculation. After normalization, all 13 features were treated with equal weight in the K-Means process, and no additional feature weighting was applied. This decision was made to allow the cluster structure to emerge from the combined behavioral and perception patterns without imposing prior assumptions about feature importance.

To examine whether the clustering result was sensitive to the selected scaling method, an additional robustness check was conducted using z-score standardization. The resulting z-score-based cluster labels were compared with the main Min-Max-based labels using the Adjusted Rand Index. This analysis was used only as a sensitivity check, while the main clustering model retained Min-Max normalization because the behavioral ordinal variables and Likert-based construct means were bounded numerical features that could be transformed into a comparable 0-1 range.

The ordinal behavioral variables were encoded according to their ordered response levels and treated as approximate interval inputs in the K-Means distance calculation. This assumption was used to preserve the ordered nature of the categories and to allow the behavioral variables to be combined with the perception construct means in a single numerical feature space. However, the use of Euclidean distance on ordinal-encoded variables remains a methodological limitation. This encoding choice may also have contributed to the ambiguity observed between the $k = 2$ and $k = 3$ solutions, since a distance metric better suited to mixed ordinal and continuous data could alter the cluster assignments and, consequently, the relative support provided by the internal validation indices reported in Table 5. Future studies may compare this approach with clustering methods designed for mixed data types, such as Gower-distance-based clustering or k-prototypes.

The modeling process used 13 features consisting of seven behavioral features and six perception features. These features are shown in Table 1.

Table 1: Features Used in the Modeling Process

No.	Feature Name	Feature Code	Type	Description
1	Frequency of use	`frequency_of_use_encoded`	Behavior	Frequency of e-wallet use
2	Number of active e-wallets	`active_ewallet_count_clean`	Behavior	Number of active e-wallets after data cleaning
3	Duration of use	`duration_of_use_encoded`	Behavior	Duration of e-wallet use
4	Transaction amount	`transaction_amount_encoded`	Behavior	Typical e-wallet transaction amount
5	Monthly e-wallet spending	`monthly_ewallet_spending_encoded`	Behavior	Monthly spending through e-wallets
6	Influence of promotions	`promotion_influence_encoded`	Behavior	Influence of promotions on e-wallet use
7	Promotion-based switching frequency	`promotion_switching_frequency_encoded`	Behavior	Frequency of switching e-wallets because of promotions
8	Perceived usefulness	`perceived_usefulness_mean`	Perception	Mean score of perceived usefulness
9	Ease of use	`ease_of_use_mean`	Perception	Mean score of ease of use
10	Security	`security_mean`	Perception	Mean score of perceived security
11	Trust	`trust_mean`	Perception	Mean score of trust
12	Promotion and incentives	`promotion_incentives_mean`	Perception	Mean score of promotion and incentives
13	Risk perception	`risk_perception_mean`	Perception	Mean score of risk perception

The Min-Max normalization formula used to scale the modeling features is shown in Equation (1).

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{1}$$

After preprocessing, two main datasets were produced. The first dataset was the preprocessed dataset, which contained cleaned values and was used for segment interpretation. The second dataset was the scaled feature dataset, which contained normalized values and was used as the main input for the K-Means clustering process. K-Means was applied to group student e-wallet users based on the similarity of behavioral and perception features. This clustering procedure follows the general data mining principle in which observations are grouped according to similarity patterns in the feature space [12], and is consistent with previous K-Means-based grouping studies [13]. In this study, the algorithm assigned each respondent to the nearest cluster centroid using Euclidean distance, as shown in Equation (2).

$$d(x_i, \mu_j) = \sqrt{\sum_{l=1}^{p} (x_{il} - \mu_{jl})^2} \tag{2}$$

where (x_i) represented the data point of respondent (i), (μ_j) represented the centroid of cluster (j), (x_{il}) represented the value of respondent (i) on feature l, (μ_{jl}) represented the centroid value of cluster (j) on feature l, and (p) represented the number of features used in clustering [14]. The K-Means procedure consisted of determining the number of clusters, initializing centroids, assigning each respondent to the nearest centroid, updating centroid values, and repeating the process until cluster assignments became stable.

The number of clusters was evaluated using the Elbow method and Silhouette score. The Elbow method evaluated cluster compactness using Within-Cluster Sum of Squares, as shown in Equation (3).

$$\mathrm{WCSS}(k) = \sum_{j=1}^{k} \sum_{x_i \in C_j} \|x_i - \mu_j\|^2 \tag{3}$$

Silhouette score was used to assess both cluster compactness and separation, as shown in Equation (4).

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i),\, b(i)\}} \tag{4}$$

where ($a(i)$) represented the average distance between data point (i) and other members of the same cluster, while ($b(i)$) represented the average distance between data point (i) and members of the nearest other cluster. Additional cluster evaluation metrics, including Davies-Bouldin Index and Calinski-Harabasz Score, were also used to support the assessment of clustering quality. The final number of clusters was selected by prioritizing Silhouette score, Davies-Bouldin Index, and interpretability of the resulting segment profiles. Because the internal validation metrics did not necessarily indicate the same optimal number of clusters, an alternative $k = 3$ solution was also inspected to support the final selection of the main clustering solution. The Calinski-Harabasz Score was considered as a supporting metric, but it was not used as the sole decision criterion because the selected cluster solution also needed to provide a clear and interpretable exploratory cluster structure.

To reduce dependence on a single clustering initialization, cluster stability was examined using two initialization settings. First, K-Means was repeated 50 times using k-means++ initialization with $n\_init = 10$, following the main model configuration. Second, a stricter sensitivity check was conducted using random initialization with $n\_init = 1$ across the same 50 random-state values. The stability evaluation considered the consistency of internal validation metrics and the Adjusted Rand Index against the main Min-Max-based K-Means labels.

After the K-Means labels were obtained, Random Forest was applied as a post-clustering interpretability model. In this stage, the cluster label generated by K-Means was used as the target variable, while the same 13 behavioral and perception features were used as input variables. Because the model reproduced labels generated from the same feature space, its classification performance was interpreted as label reproducibility rather than independent external validation. The main role of Random Forest was to identify the features that contributed most to differentiating the resulting segments through feature-importance analysis.

The Random Forest model was implemented with 300 trees, the Gini impurity criterion, no maximum depth restriction, $class\_weight$ = "balanced", $max\_features$ = "sqrt", $min\_samples\_split = 2$, $min\_samples\_leaf = 1$, and $random\_state = 42$. The balanced class-weight setting was used because the K-Means labels produced an imbalanced segment distribution. These parameters were used to reproduce the K-Means-generated labels and to obtain feature-importance values for post-clustering interpretation.

The dataset was divided into training and testing data using an 80:20 ratio. From 569 valid respondents, 455 data points were used for training and 114 data points were used for testing. The split was conducted using a stratified approach so that the proportion of each cluster was preserved in both training and testing data. Model performance was evaluated using confusion matrix, accuracy, precision, recall, macro F1-score, weighted F1-score, and per-class classification metrics. In addition, Stratified K-Fold Cross-Validation was conducted to reduce dependence on a single train-test split and to provide a more stable evaluation of the Random Forest model under imbalanced segment distribution.

The data analysis was implemented using Python in Google Colaboratory. The main Python libraries used in the analysis included Pandas and NumPy for data processing, Scikit-learn for preprocessing, K-Means clustering, Random Forest classification, model evaluation, and cross-validation, as well as Matplotlib and Seaborn for data visualization. These tools were selected because they support reproducible data analysis and are commonly used in machine learning-based research.

Ethical considerations were applied because the study involved human respondents. Participation was based on respondent willingness, and the questionnaire included screening questions to ensure that participants met the research criteria. Personal identity columns, including timestamp and email address, were removed from the analytical dataset. The dataset was anonymized using respondent identifiers, and the data were used only for research analysis. These procedures were applied to protect respondent privacy and reduce the risk of exposing personal information during data processing.

3. RESULT AND ANALYSIS

The analysis was conducted using valid questionnaire data collected from student e-wallet users. From the initial 571 responses, 569 responses met the respondent criteria and were used in the analysis, while 2 responses were excluded because they did not meet the research requirements. The final dataset exceeded the minimum sample requirement of 385 respondents, indicating that the available data were sufficient for the segmentation process.

Table 2: Summary of Research Dataset

Description	Number
Initial responses	571
Invalid responses	2
Valid responses	569
Dataset used for analysis	569

Table 2 shows that most collected responses were eligible for further analysis. The exclusion of invalid responses ensured that the segmentation process was based only on respondents who were active university students, used at least one e-wallet service, and had conducted transactions using an e-wallet within the last three months. This screening process was important because the quality of the segment formation depended on the relevance and consistency of respondent characteristics.

Before the perception variables were used in the modeling process, the questionnaire items were tested for validity and reliability. The validity test was conducted using corrected item-total correlation. The questionnaire contained 24 perception items grouped into six constructs: perceived usefulness, ease of use, security, trust, promotion and incentives, and risk perception. The results showed that all items had corrected item-total correlation values above the minimum validity criterion, indicating that each item was suitable for representing its corresponding construct.

Table 3: Validity Test Summary of User Perception Items

Construct	Item Code	Corrected Item-Total Correlation Range	Criterion	Result
Perceived usefulness	Q22–Q25	0.833–0.850	> 0.30	Valid
Ease of use	Q26–Q29	0.812–0.837	> 0.30	Valid
Security	Q30–Q33	0.801–0.845	> 0.30	Valid
Trust	Q34–Q37	0.810–0.851	> 0.30	Valid
Promotion and incentives	Q38–Q41	0.822–0.852	> 0.30	Valid
Risk perception	Q42–Q45	0.824–0.854	> 0.30	Valid

As shown in Table 3, all perception constructs met the validity criterion. This result indicates that the items used in the questionnaire were able to represent their intended constructs. The highest item-total correlation values were found in the promotion and incentives construct and the risk perception construct, while the lowest values were still far above the minimum criterion. Therefore, all 24 perception items were retained for the construction of perception scores.

The complete item-level validity results, including corrected item-total correlation values and Cronbach's Alpha if item deleted for all 24 perception items, are provided in Appendix A.

Reliability testing was then conducted using Cronbach's Alpha to evaluate the internal consistency of each construct. The results are shown in Table 4.

Table 4: Reliability Test of User Perception Constructs

Construct	Number of Items	Cronbach's Alpha	Interpretation
Perceived usefulness	4	0.932	Very high reliability
Ease of use	4	0.922	Very high reliability
Security	4	0.922	Very high reliability
Trust	4	0.929	Very high reliability
Promotion and incentives	4	0.932	Very high reliability
Risk perception	4	0.931	Very high reliability

Table 4 shows that all constructs had Cronbach's Alpha values above 0.90, indicating very high internal consistency. These results support the use of construct mean scores in the modeling process. Each perception construct was then represented by the mean score of its four item indicators, so that the original 24 perception items were transformed into six perception features without losing their conceptual meaning.

Harman's single-factor test was also conducted as an indicative diagnostic for common-method bias. The first unrotated factor explained 73.5% of the total variance across the 24 perception items, exceeding the commonly used 50% threshold. This result suggests that common-method bias may be a concern because the perception items were collected using the same self-reported questionnaire instrument. Therefore, the perception-based findings in this study should be interpreted cautiously, and this issue is acknowledged as a methodological limitation. The detailed variance explained by each component is provided in Appendix C.

K-Means clustering was applied to 13 modeling features consisting of seven e-wallet usage behavior features and six user perception features. The number of clusters was evaluated by testing candidate values of k from 2 to 10 using WCSS/Inertia, Silhouette score, Davies-Bouldin Index, and Calinski-Harabasz Score. The Elbow method and Silhouette score visualization are shown in Figure 2 and Figure 3.

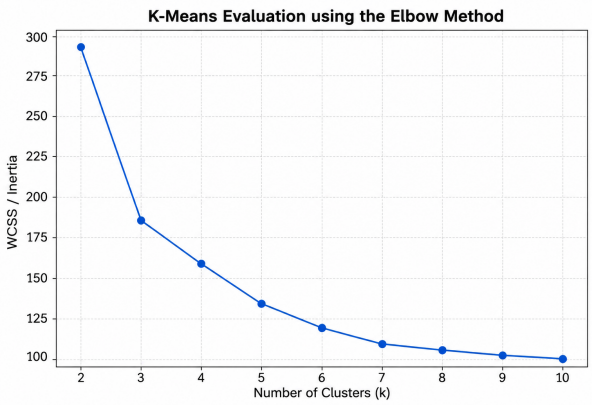


Figure 2. Elbow Method Graph

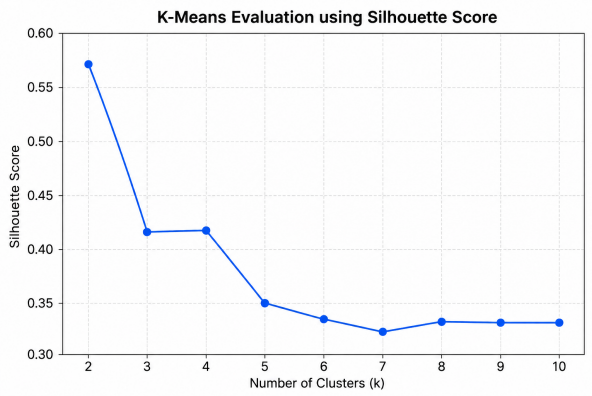


Figure 3. Silhouette Score Graph Based on the Number of Clusters

The detailed evaluation results for each candidate number of clusters are presented in Table 5.

Table 5: K-Means Evaluation Across Candidate Numbers of Clusters

k	WCSS/Inertia	Silhouette Score	Davies-Bouldin Index	Calinski-Harabasz Score
2	293.836	0.568	0.674	491.614
3	185.074	0.413	0.913	555.884
4	158.388	0.414	1.157	463.992
5	132.775	0.351	1.152	441.590
6	118.522	0.335	1.102	408.595
7	107.727	0.329	1.186	383.334
8	102.409	0.336	1.340	349.181
9	97.520	0.334	1.155	323.791
10	94.340	0.334	1.142	299.078

Table 5 shows that the clustering evaluation metrics did not fully agree on a single optimal value of k. The WCSS/Inertia decreased as k increased, and the elbow pattern suggested that improvement became more gradual after approximately $k = 3$. The Calinski-Harabasz Score also reached its highest value at $k = 3$. However, $k = 2$ produced the highest Silhouette score of 0.568 and the lowest Davies-Bouldin Index of 0.674, indicating better

separation and compactness according to these two internal validation indices. Therefore, the final selection of $k = 2$ was based on Silhouette score, Davies-Bouldin Index, and interpretability, while acknowledging that $k = 3$ was also supported by the Calinski-Harabasz Score and the elbow pattern.

To examine this alternative, the $k = 3$ solution was inspected separately. The $k = 3$ model produced three clusters with 201 respondents (35.3%), 76 respondents (13.4%), and 292 respondents (51.3%). The second cluster in the $k = 3$ solution closely resembled the lower-perception group, while the first and third clusters mainly separated high-perception respondents into lower-usage and higher-usage subgroups. Although this provided a more detailed partition, the lower Silhouette score of 0.413 and higher Davies-Bouldin Index of 0.913 indicated weaker separation than the $k = 2$ solution. For this reason, $k = 2$ was retained as the main clustering solution because it provided a broader, more stable, and more interpretable exploratory structure. The detailed inspection of the $k = 3$ solution is provided in Appendix D.

Table 6: Scaling Sensitivity Analysis for K-Means with $k = 2$

Scaling Method	**Silhouette Score**	**Davies-Bouldin Index**	**Calinski-Harabasz Score**	**ARI vs. Min-Max Labels**
Min-Max	0.568	0.674	491.614	1.000
Z-score	0.574	0.652	506.206	0.981

Table 6 shows that the two-cluster solution was highly similar across the two scaling approaches. The z-score model produced a Silhouette score of 0.574, a Davies-Bouldin Index of 0.652, and a Calinski-Harabasz Score of 506.206, while the Min-Max model produced a Silhouette score of 0.568, a Davies-Bouldin Index of 0.674, and a Calinski-Harabasz Score of 491.614. The Adjusted Rand Index between the Min-Max and z-score clustering labels was 0.981, indicating that the selected two-cluster structure was not strongly dependent on the use of Min-Max normalization.

To further examine the robustness of the selected two-cluster solution, cluster stability was evaluated under two initialization settings. The first setting followed the main model configuration using k-means++ initialization with $n\_init = 10$. The second setting used random initialization with $n\_init = 1$ as a stricter sensitivity check. The results are shown in Table 7.

Table 7: Cluster Stability Analysis under Different Initialization Settings

Initialization	**Runs**	n_{init}	**Silhouette Mean**	**DBI Mean ± Std**	**CH Mean ± Std**	**ARI Mean ± Std**	**ARI Range ± Std**
k-means++	50	10	0.568 ± 0.000	0.674 ± 0.000	491.614 ± 0.000	1.000 ± 0.000	1.000-1.000
random	50	1	0.494 ± 0.127	0.903 ± 0.391	406.721 ± 144.673	0.738 ± 0.447	−0.035-1.000

Table 7 shows that the main K-Means configuration using k-means++ initialization with $n\_init = 10$ produced identical validation metrics across 50 random-state values, with ARI = 1.000 against the main labels. This indicates that the selected two-cluster solution was stable under the optimized initialization setting used in the main analysis. In contrast, the stricter sensitivity check using random initialization with $n\_init = 1$ produced more variable results, with mean ARI = 0.738 and ARI values ranging from -0.035 to 1.000. These findings indicate that single random initialization could converge to less consistent local solutions, while the optimized k-means++ configuration provided a more stable clustering result. Therefore, k-means++ with multiple initializations was retained for the main analysis. It should be emphasized that this stability evidence reflects the computational reproducibility of the partition under different initialization settings, not confirmation that the partition corresponds to genuine, real-world behavioral heterogeneity among users.

The final K-Means model divided the respondents into two user segments. The distribution and interpretation of the segments are presented in Table 8.

Table 8: Distribution and Interpretation of E-Wallet User Segments

Cluster	**Segment Name**	**Number of Respondents**	**Percentage**	**Main Characteristics**
Cluster 1	High-Perception Promo-Responsive Users	491	86.3%	Higher perception scores across service-related constructs, stronger promotion responsiveness, and higher risk awareness.
Cluster 2	Lower-Perception Selective Users	78	13.7%	Lower perception scores and lower promotion responsiveness; some usage indicators were slightly higher, but behavioral differences were small.
Total	-	569	100.0%	-

Cluster 1 was interpreted as High-Perception Promo-Responsive Users because this group showed higher scores on ease of use, trust, security, promotion and incentives, perceived usefulness, and risk perception, as well as stronger responsiveness to promotional incentives.

Cluster 2 was interpreted as Lower-Perception Selective Users. This group showed lower scores across all perception constructs and lower promotional responsiveness than Cluster 1. Although several behavioral indicators, such as frequency of use, monthly e-wallet spending, number of active e-wallets, transaction amount, and duration of use, were slightly higher in Cluster 2, these differences should not be overinterpreted because the behavioral features later showed very low importance in the Random Forest model. Therefore, Cluster 2 is interpreted mainly as a lower-perception and lower-promotion-response pattern rather than as a strongly behaviorally distinct intensive user segment.

The unequal distribution between Cluster 1 and Cluster 2 provides an important interpretation for the exploratory segmentation result. The dominance of Cluster 1 suggests that most student respondents reported generally positive evaluations of e-wallet services and were receptive to promotional incentives, while also showing stronger awareness of potential transaction and security risks. However, the smaller size of Cluster 2 should not be interpreted as lower importance. Although this segment represented only 13.7% of respondents, it captured a lower-perception and lower-promotion-response pattern that may require further investigation using additional data sources.

The cluster imbalance may also be influenced by the characteristics of the sample and the use of self-reported survey data. Since respondents were active student e-wallet users, the dataset may naturally contain more users with positive attitudes toward digital payment services. In addition, students who are less interested in or less satisfied with e-wallet services may have been less represented in the survey. Therefore, the dominance of Cluster 1 should be interpreted as a pattern observed in this dataset rather than as a definitive representation of all student e-wallet users.

After the K-Means segments were formed, Random Forest was used to examine how consistently the K-Means labels could be reproduced from the same 13 behavioral and perception features and to identify which features contributed most to segment differentiation.

The confusion matrix of the Random Forest model is shown in Figure 4, and the detailed classification performance is presented in Table 9. These results are interpreted as label reproducibility within the selected feature representation.

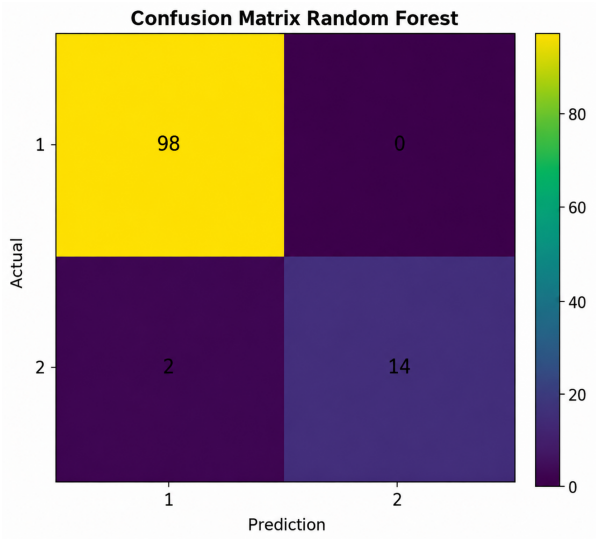


Figure 4. Random Forest Confusion Matrix

Table 9: Random Forest Classification Report

Class	Precision	Recall	F1-score	Support
Cluster 1	0.980	1.000	0.990	98
Cluster 2	1.000	0.875	0.933	16
Accuracy	-	-	0.983	114
Macro average	0.990	0.938	0.962	114
Weighted average	0.983	0.983	0.982	114

Table 9 shows that the Random Forest model reproduced most K-Means labels correctly in the testing data. The model achieved an accuracy of 0.983 and a macro F1-score of 0.962. It is important to emphasize that because the Random Forest target labels were generated from the same 13 features used as predictors, these figures cannot, even in principle, exceed the information already contained in the K-Means clustering itself; they should therefore be read strictly as a measure of label reproducibility rather than as independent validation of the segmentation. The per-class results provide useful information about label reproducibility. Cluster 1 achieved a recall of 1.000, while Cluster 2 achieved a recall of 0.875. This indicates that the model identified all test samples from the majority segment correctly, but two samples from the smaller segment were classified as Cluster 1. Because the segment distribution was imbalanced, the macro average was important for evaluating both clusters more fairly.

To reduce dependence on a single train-test split, Stratified K-Fold Cross-Validation was also applied. The stratified procedure preserved the proportion of both clusters in each fold, which was important because the segment distribution was imbalanced. The cross-validation results are shown in Table 10.

Table 10: Stratified K-Fold Cross-Validation Results of Random Forest

Metric	Mean	Standard Deviation
Accuracy	0.997	0.007
F1-score Macro	0.992	0.016

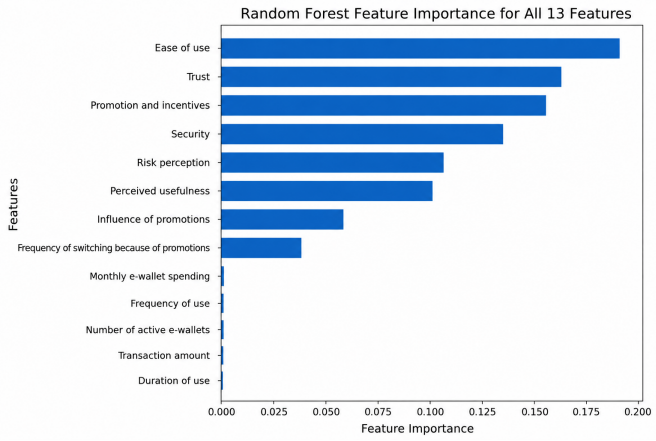


Figure 5. Feature Importance of Random Forest

Table 11: Full Feature Importance and Cluster Mean Direction

No.	Feature	Type	Importance	Cluster 1 Mean	Cluster 2 Mean	Higher Mean in Cluster
1	Ease of use	Perception	0.193	4.428	1.660	Cluster 1
2	Trust	Perception	0.167	4.344	1.676	Cluster 1
3	Promotion and incentives	Perception	0.162	4.362	1.651	Cluster 1
4	Security	Perception	0.144	4.342	1.753	Cluster 1
5	Risk perception	Perception	0.115	4.390	1.660	Cluster 1
6	Perceived usefulness	Perception	0.111	4.376	1.747	Cluster 1
7	Influence of promotions	Behavior	0.060	4.287	1.538	Cluster 1
8	Frequency of switching because of promotions	Behavior	0.044	4.246	1.500	Cluster 1
9	Monthly e-wallet spending	Behavior	0.001	3.580	3.897	Cluster 2
10	Frequency of use	Behavior	0.001	3.625	3.910	Cluster 2
11	Number of active e-wallets	Behavior	0.001	2.953	3.115	Cluster 2
12	Transaction amount	Behavior	0.000	2.874	2.897	Cluster 2
13	Duration of use	Behavior	0.000	3.843	4.013	Cluster 2

Note: Feature importance values are rounded to three decimals; values shown as 0.000 represent very small non-zero values.

Table 12: Aggregated Feature Importance by Feature Type

Feature Type	Total Importance	Percentage
Perception	0.892	89.2%
Behavior	0.108	10.8%

Tables 11 and 12 present the complete feature-importance results for all 13 modeling features and the aggregated importance by feature type. The importance values summed to 1.000, confirming that the table represents the full Random Forest feature-importance output. The six perception features occupied the top six positions, with ease of use, trust, promotion and incentives, security, risk perception, and perceived usefulness contributing the largest importance values. In total, perception features accounted for 89.2% of the total importance, while behavioral features accounted for 10.8%. This result supports the interpretation that the K-Means segment labels were mainly reproduced through perception-related differences rather than transaction behavior alone. Because the seven behavioral features individually contributed near-zero importance, this indicates that the 'Selective Users' label for Cluster 2 is primarily a lower-perception profile rather than a distinctly selective behavioral pattern. The behavioral component of the label is therefore treated as a descriptive convenience rather than an empirically distinguishing characteristic, and this perception-driven reading is adopted as the primary interpretation of the segment throughout the remainder of this section.

The cluster mean direction shows that Cluster 1 had higher mean scores on positive service perception features, promotion-related behavior features, and risk perception. Because the risk perception items were phrased as concerns about system errors, account misuse, personal data misuse, security risks, and hesitation in high-value transactions, a higher risk perception mean indicates stronger risk awareness rather than lower perceived risk. This indicates that High-Perception Promo-Responsive Users evaluated e-wallet services positively and were more responsive to promotional factors, while still recognizing possible transaction and security risks. In contrast, Cluster 2 had slightly higher means on monthly e-wallet spending, frequency of use, number of active e-wallets, transaction amount, and duration of use, although these behavioral features had very low importance values in the Random Forest model. Therefore, Cluster 2 should be interpreted mainly as a lower-perception selective group rather than as a strongly behaviorally distinct intensive segment, while Cluster 1 can be interpreted as a larger group with stronger service perception, promotional responsiveness, and risk awareness.

A rival explanation for the two-cluster structure is that the clustering may partly reflect response-style bias or acquiescence tendency rather than substantively distinct e-wallet market segments. The perception means were consistently higher in Cluster 1 and consistently lower in Cluster 2 across all six perception constructs, while Harman's single-factor test showed that the first unrotated factor explained 73.5% of the total variance. Together with the dominance of perception features in the Random Forest importance results, which accounted for 89.2% of the total importance, this pattern suggests that common-method variance may have contributed to the separation between clusters. Therefore, the two segments should be interpreted as exploratory survey-based patterns rather than externally validated market segments. Replication using behavioral logs, transaction records, or other non-self-reported data is needed before the segments can be claimed as validated market segments.

These findings are consistent with previous fintech and digital payment studies showing that users of digital financial services are heterogeneous and may differ in their motivations, perceptions, and behavioral responses. Jaiswal et al. [9] showed that mobile payment users could be grouped into different fintech-led market profiles, while the present study provides a more specific segmentation of student e-wallet users in the Indonesian context. The result also aligns with Joung and Kim [10], who emphasized the importance of interpretable machine learning in explaining customer segment characteristics. In addition, this study extends the hybrid machine learning perspective used by Ch'ng [11] by focusing not on e-wallet adoption prediction, but on market segmentation and feature-importance-based interpretation of student user segments.

Another possible explanation is that student respondents may have relatively similar levels of digital familiarity and basic e-wallet usage experience. Under this condition, basic usage behavior alone may provide limited differentiation between user groups, while differences may appear more clearly in how students evaluate the service. However, this explanation should be treated as secondary and speculative because the study did not directly measure digital familiarity as a separate variable.

Overall, the integration of K-Means and Random Forest produced complementary findings. K-Means formed two exploratory survey-based clusters from behavioral and perception features, while Random Forest supported post-clustering interpretation by identifying the most important features associated with the cluster labels. From a practical perspective, the findings may offer preliminary guidance for understanding how student respondents differed in their reported evaluations of e-wallet services. However, these implications should be treated cautiously because the cluster structure may partly reflect response-style bias and common-method variance. Promotional strategies may be relevant for High-Perception Promo-Responsive Users, but the interpretation should not be treated as a validated market prescription. For Lower-Perception Selective Users, practical implications should re-

main exploratory because the segment was defined mainly by lower perception scores rather than strong behavioral distinctiveness.

This study has several limitations. First, the respondents were limited to student e-wallet users, so the findings may not be directly generalizable to other demographic groups. Second, the data were collected using self-reported questionnaire responses, which may be affected by response bias, acquiescence tendency, and common-method variance. Harman's single-factor test indicated that the first unrotated factor explained 73.5% of the total variance, suggesting that common-method variance is a substantial concern in the perception items. Therefore, the resulting clusters should be interpreted as exploratory survey-based patterns rather than validated market segments. Third, the study used a cross-sectional design, so it could not capture changes in e-wallet behavior over time. Fourth, the ordinal behavioral variables were treated as approximate interval inputs in the K-Means distance calculation, which may not fully represent the true distance between ordinal response levels. Finally, the two-cluster solution provided an interpretable but broad segmentation structure. Future studies should replicate the workflow using more diverse respondents, behavioral logs, transaction records, or other non-self-reported data, and may compare additional clustering algorithms or mixed-data clustering methods.

4. CONCLUSION

This study applied an exploratory computational workflow to segment student e-wallet users using 13 behavioral and perception features. The K-Means results produced two exploratory clusters at k = 2: High-Perception Promo-Responsive Users, consisting of 491 respondents (86.3%), and Lower-Perception Selective Users, consisting of 78 respondents (13.7%). The selected $k = 2$ solution was supported by the highest Silhouette score and the lowest Davies-Bouldin Index, although the alternative $k = 3$ solution was also inspected because it was supported by the Calinski-Harabasz Score and the elbow pattern. Because these internal indices did not converge on a single optimal solution and the Silhouette score of 0.568 indicated only moderate cluster separation, the selection of $k = 2$ should be understood as a judgment call that prioritized interpretability over an unambiguous statistical optimum, rather than a solution uniquely dictated by the validation metrics.

Random Forest was used as a post-clustering interpretability tool to identify features associated with the K-Means cluster labels. The feature-importance results showed that ease of use, trust, promotion and incentives, security, risk perception, and perceived usefulness were the most important distinguishing features. Aggregated feature importance showed that perception features contributed 89.2% of the total importance, while behavioral features contributed 10.8%. These results suggest that the exploratory clusters were shaped more strongly by perception-related differences than by transaction behavior alone.

The findings should be interpreted cautiously. The consistently high perception scores in Cluster 1 and consistently low perception scores in Cluster 2, together with the Harman's single-factor result, suggest that response-style bias and common-method variance may have contributed to the two-cluster structure. Therefore, the clusters should be treated as exploratory survey-based patterns rather than externally validated market segments. Future studies are recommended to replicate the workflow using more diverse respondent groups, behavioral logs, transaction records, or other non-self-reported data, and to compare clustering methods designed for mixed data types.

REFERENCES

[1] Bank Indonesia. (2024, Mar.) Transaksi ekonomi dan keuangan digital (tkm) maret 2024. [Online]. Available: https://www.bi.go.id/id/publikasi/laporan/Pages/TKM-Maret-2024.aspx

[2] Kementerian Komunikasi dan Digital Republik Indonesia. (2024) Menkomdigi paparkan visi ekonomi digital dalam orasi ilmiah di universitas. [Online]. Available: https://bpsdm.komdigi.go.id/berita-menkomdigi-paparkan-visi-ekonomi-digital-dalam-orasi-ilmiah-di-universitas--46-2169

[3] R. N. P. B. Puspaningrum and A. D. R. Atahau, "Penggunaan e-wallet dalam transaksi e-commerce: Analisis unified theory of acceptance and use of technology (utaut)," *Jurnal Ekonomi Pendidikan dan Kewirausahaan*, vol. 11, no. 2, pp. 191–208, 2023.

[4] S. Saliyeh, Y. Firayanti, and E. S. Saputra, "Pengaruh pembayaran non-tunai (e-wallet) dan gaya hidup terhadap perilaku konsumtif mahasiswa universitas nahdlatul ulama kalimantan barat angkatan 2020," *AKSIOMA: Jurnal Sains Ekonomi dan Edukasi*, vol. 1, no. 9, pp. 800–819, 2024.

[5] P. S. Isnaini and S. Jamilah, "Analysis of the adoption of digital wallet technology in indonesian society," *BASKARA: Journal of Business and Entrepreneurship*, vol. 7, no. 2, pp. 254–269, 2025.

[6] R. Nurcahyo *et al.*, "Enhancing digital wallet services based on electronic word of mouth using social media reviews," *SAGE Open*, vol. 15, no. 3, pp. 1–15, 2025.

[7] S. K. Abbas, M. Hussain, and Y. N. Rimal, "Machine learning-based analysis of technology acceptance in fintech: A behavioral study using digital wallet data," *SN Computer Science*, vol. 6, pp. 1–12, 2025, article number 674.

[8] J. W. Mamani Lopez, A. V. Morales Gonzales, and P. P. Chambi Condori, "Predictive models based on machine learning to analyze the adoption of digital payments in latin america and the caribbean," *International Journal of Data and Network Science*, vol. 9, no. 3, pp. 411–418, 2025.

[9] D. Jaiswal, A. Mohan, and A. K. Deshmukh, "Cash rich to cashless market: Segmentation and profiling of fintech-led mobile payment users," *Technological Forecasting and Social Change*, vol. 193, p. 122627, 2023.

[10] J. Joung and H. Kim, "Interpretable machine learning-based approach for customer segmentation for new product development from online product reviews," *International Journal of Information Management*, vol. 70, p. 102641, 2023.

[11] C. K. Ch'ng, "Hybrid machine learning approach for predicting e-wallet adoption among higher education students in malaysia," *Journal of Information and Communication Technology*, vol. 23, no. 2, pp. 177–210, 2024.

[12] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 4th ed. Cambridge, MA, USA: Morgan Kaufmann, 2022.

[13] I. S. Tinendung and I. Zufria, "Pengelompokan status stunting pada anak menggunakan metode k-means clustering," *Jurnal Media Informatika Budidarma*, vol. 7, no. 4, pp. 2014–2023, 2023.

[14] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning: With Applications in R*, 2nd ed. New York, NY, USA: Springer, 2021.