



A Metaheuristic Particle Swarm Optimization Framework to Refine K-Means-Based Customer Segmentation in E-Commerce

Mashud^{1*}, Muhammad Rizal H², Annisa Nurul Puteri³, Elly Warni⁴

¹ Department of Digital Business, Universitas Teknologi Akba Makassar, Makassar 90245, Indonesia

² Department of Informatics Engineering, Universitas Teknologi Akba Makassar, Makassar 90245, Indonesia

³ Department of Informatics and Computer Engineering, Ujung Pandang State Polytechnic, Makassar 90245, Indonesia

⁴ Department of Informatics Engineering, Hasanuddin University, Gowa 92119, Indonesia

Corresponding Author Email: mashud@akba.ac.id

Copyright: ©2026 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/mmep.130203>

ABSTRACT

Received: 5 November 2025

Revised: 26 December 2025

Accepted: 5 January 2026

Available online: 15 March 2026

Keywords:

Particle Swarm Optimization, K-Means clustering, customer segmentation, e-commerce analytics, machine learning optimization, clustering validation metrics, data-driven marketing

Customer segmentation is a critical component of data-driven personalization and retention strategies in e-commerce; however, conventional K-Means clustering can be sensitive to random centroid initialization and may converge to local optima. To address this limitation, this study proposes a metaheuristic clustering framework that integrates Particle Swarm Optimization (PSO) with K-Means to improve clustering stability and validity. The framework is evaluated on a structured e-commerce dataset consisting of 450 customer profiles with demographic and behavioral attributes, following systematic preprocessing and information-based feature selection. Experimental results show that the PSO-enhanced K-Means achieves better internal clustering validity than standard K-Means and K-Means++, with the Silhouette Coefficient increasing from 0.4940 to 0.5126 and the Davies–Bouldin Index decreasing from 0.9942 to 0.7674, indicating stronger intra-cluster cohesion and inter-cluster separation. In addition, more than 80% of fitness improvement is achieved within the first 20 optimization iterations, suggesting faster and more stable convergence. From a practical perspective, the resulting customer segments exhibit clearer behavioral differentiation, supporting more effective personalization, targeting, and customer management strategies in e-commerce. This study contributes an end-to-end clustering framework that unifies PSO-based centroid optimization with information-theoretic feature selection, aiming to improve analytical robustness and business interpretability for e-commerce customer segmentation.

1. INTRODUCTION

The digital transformation of global commerce has led to an exponential growth of customer interaction data, encompassing transactions, purchasing frequency, and satisfaction indicators [1]. Effectively analyzing this data is essential to fostering customer loyalty, improving marketing precision, and sustaining e-commerce competitiveness [2]. At the core of these efforts lies customer segmentation, a technique that classifies customers based on similar attributes and behaviors to enable personalized marketing and strategic decision-making [3].

Among segmentation methods, K-Means clustering remains widely adopted due to its computational efficiency and ease of interpretation [4]. However, its effectiveness is constrained by key limitations such as sensitivity to centroid initialization, convergence to local minima, and poor performance on heterogeneous or high-dimensional datasets [5]. These shortcomings often result in unstable and less actionable clusters, particularly in complex e-commerce environments.

To overcome such issues, metaheuristic algorithms, especially Genetic Algorithms (GA), Ant Colony

Optimization (ACO), and Particle Swarm Optimization (PSO), have been incorporated into clustering workflows. These techniques offer global optimization capabilities that guide centroid initialization and iterative refinement beyond local optima, thereby enhancing intra-cluster homogeneity and inter-cluster separation [6].

Despite their promise, conventional K-Means and even K-Means++ still rely on probabilistic heuristics for initialization, lacking adaptive search mechanisms during centroid updates [7]. This limitation underlines the need for an optimization-driven approach that ensures clustering stability, convergence, and interpretability across diverse e-commerce data landscapes.

To address this, the present study proposes a hybrid clustering framework that integrates PSO with K-Means. PSO, inspired by social behavior in nature, dynamically adjusts centroid candidates based on individual and collective learning, promoting more accurate and stable convergence. Compared to traditional methods, PSO-enhanced K-Means is expected to produce segments with improved cohesion, separation, and interpretability.

Previous works have supported the efficacy of such

integration. GA-based models, for instance, improved clustering accuracy through evolutionary adaptation [8], while PSO has been shown to reduce intra-cluster variance and accelerate convergence in high-dimensional spaces [9]. Additionally, incorporating feature selection techniques like Adaptive Gain Ratio (AGR) and Correlation-Based Feature Selection (CBFS) has been found to significantly improve both clustering robustness and interpretability [10, 11].

While optimization-based clustering has been extensively studied, its combined application with information-theoretic feature selection on structured e-commerce datasets remains limited. Most existing models prioritize computational gains over business interpretability or validation consistency [12]. Furthermore, few studies systematically evaluate performance using diverse internal validity indices (e.g., Silhouette, Davies–Bouldin Index, Calinski–Harabasz Index) alongside customer behavior metrics [13, 14].

To fill this gap, this study introduces a PSO-K-Means framework enhanced with AGR–CBFS feature selection for e-commerce customer segmentation. The novelty lies in its unified optimization of centroid positions and feature relevance within an end-to-end pipeline—bridging algorithmic performance with business-relevant insights. The framework is tested on a structured dataset combining demographic (e.g., age, city, membership type) and behavioral (e.g., total spend, items purchased, satisfaction) features. Comparative analysis of Standard K-Means, K-Means++, and PSO-K-Means across 100 iterations highlights both quantitative performance and qualitative interpretability, establishing the proposed method’s contribution to optimization-based customer segmentation.

2. LITERATURE REVIEW

2.1 K-Means and recent variants for practical segmentation

K-Means remains one of the most widely used partitional clustering algorithms because of its simplicity, scalability, and ease of interpretation in customer analytics workflows [15]. Nevertheless, its sensitivity to initialization and tendency to converge to locally optimal solutions can lead to unstable partitions, especially when applied to heterogeneous or high-dimensional customer attributes [16]. Recent studies have therefore emphasized improvements to centroid initialization and update strategies to increase robustness. For example, modified initial centroid selection schemes and enhanced K-Means variants have been reported to improve clustering consistency and reduce sensitivity to random seeding [17]. Beyond initialization, contemporary K-Means research also highlights practical adaptations for modern data settings (e.g., large-scale, high-dimensional, and business-facing segmentation), positioning K-Means as a baseline that often requires refinement through principled variants or hybrid optimization to ensure reliability in decision-making contexts [5].

2.2 Particle Swarm Optimization for centroid search and stability

PSO is a population-based metaheuristic widely adopted to address K-Means limitations by enabling global exploration of centroid candidates before local refinement [18]. In clustering,

a PSO solution typically encodes a set of centroids; particles iteratively update these candidates through social learning mechanisms to minimize intra-cluster dispersion or related objective functions. Recent meta-heuristic surveys categorize PSO-based clustering as a prominent approach for partitional clustering improvement due to its balance between exploration and exploitation and its flexibility in handling non-convex search landscapes [19]. Over the past five years, PSO-driven K-Means variants, particularly adaptive designs that adjust inertia weights and learning factors, have been reported to produce more accurate and stable customer groupings than conventional K-Means by reducing the risk of poor initialization and local minima entrapment. Collectively, these findings support PSO as an effective optimization layer that improves centroid positioning and segmentation stability while maintaining the interpretability advantages of K-Means.

2.3 Optimization-based clustering and evaluation practices

Optimization-based clustering extends beyond PSO by integrating metaheuristics with partitional clustering to improve validity, automation, and robustness—especially when the number of clusters, centroid positions, or feature subsets must be optimized jointly [20]. Recent reviews on K-Means hybridization emphasize that the greatest improvements are typically achieved when metaheuristics are aligned with well-chosen objective functions and validity measures that reflect both cohesion and separation [21]. In parallel, the selection and benchmarking of clustering validity indices have become increasingly important because different indices can favor different structural properties of the same clustering solution. Recent benchmarking work on evolutionary K-Means frameworks shows that internal validity indices (e.g., Silhouette and Calinski–Harabasz) can provide more reliable guidance for optimization-driven clustering under varied data characteristics [22]. For customer segmentation settings, internal indices and validity analysis are also recommended to ensure that improved numerical scores correspond to meaningful and separable customer groups, supporting interpretability and downstream business use [23]. More broadly, recent clustering literature in data science and management highlights that optimization-based clustering is increasingly adopted when stability and reproducibility are required for operational decision-making rather than one-off exploratory analysis [24].

3. METHODOLOGY

This study adopts a structured and reproducible methodological framework to evaluate the effectiveness of a metaheuristic-optimized clustering approach that integrates PSO with K-Means for e-commerce customer segmentation. The overall workflow follows the Cross-Industry Standard Process for Data Mining (CRISP-DM) paradigm, ensuring systematic progression from problem formulation to model evaluation [25]. Figure 1 illustrates the complete research workflow.

Figure 2 shows the customer behavior correlation matrix, illustrating the relationships among key variables in the e-commerce dataset, namely age, total spend, items bought, average rating, and days since last purchase. The colors in the heatmap represent the direction and strength of the relationships between variables, where blue indicates a

positive correlation, while red signifies a negative correlation.

Figure 3 illustrates the average total customer spending by city. This bar chart compares the average spending levels of customers across six major cities: San Francisco, New York, Los Angeles, Miami, Chicago, and Houston. From the visualization, it can be observed that San Francisco records the highest average spending at \$1,460, followed by New York with \$1,165. Meanwhile, Houston shows the lowest average spending at only \$447. This pattern indicates significant differences in purchasing power and consumer behavior across regions, which may be influenced by local economic conditions, income levels, and consumer preferences in each

city.

Figure 4 illustrates the relationship between customer age and total spending (age vs. total spend), visualized through a scatter plot. Each point on the graph represents customers categorized into three satisfaction levels: Satisfied (green), Neutral (yellow), and Unsatisfied (red). From the graph, it can be observed that customers aged around 28 to 31 years tend to have higher spending levels, particularly among those who are Satisfied, with total spending reaching over \$1,400. In contrast, customers aged above 35 years generally exhibit lower spending levels and are more likely to fall into the Neutral or Unsatisfied categories.

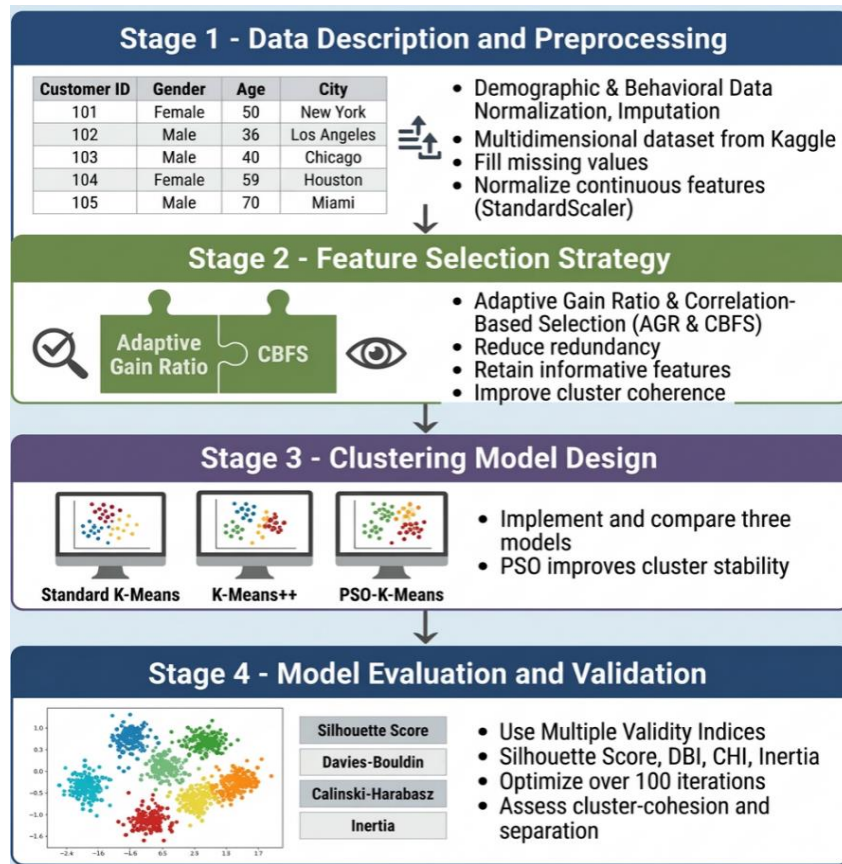


Figure 1. Research workflow

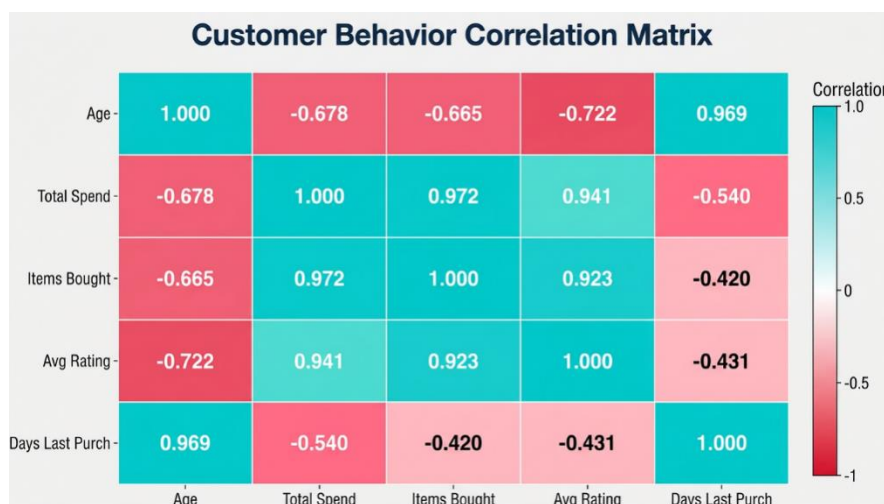


Figure 2. Customer behavior correlation matrix

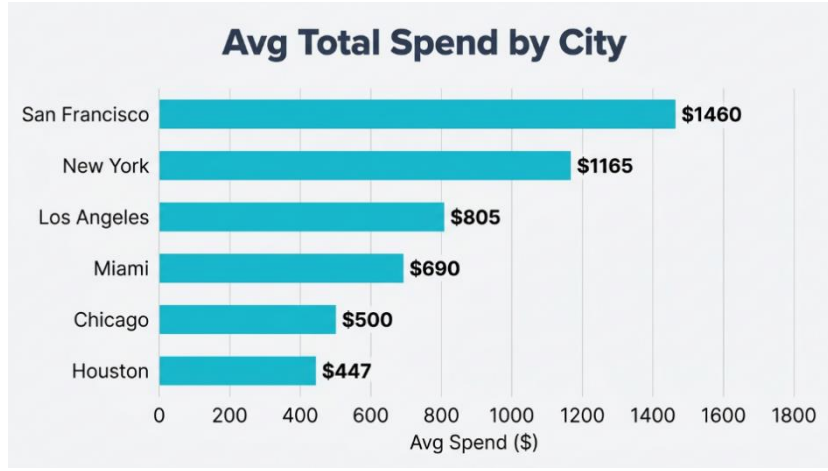


Figure 3. The average total customer spending by city

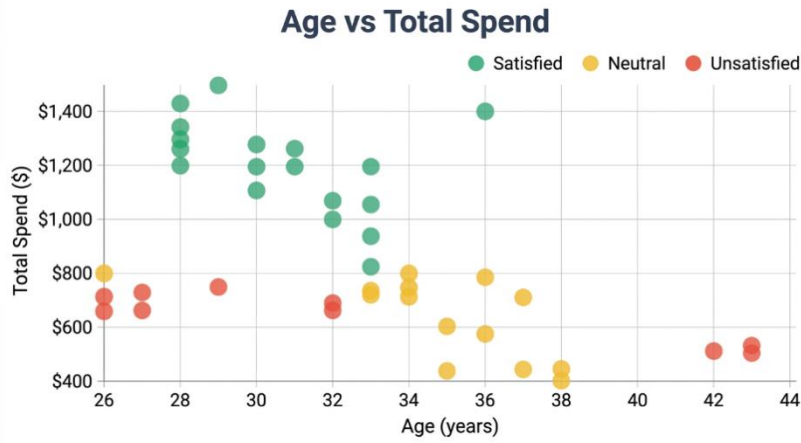


Figure 4. The relationship between customer age and total spending

3.1 Data preprocessing and feature selection

The data preprocessing stage involved handling missing values, applying normalization using StandardScaler, and implementing information-based feature selection through the AGR–CBFS approach. During the data preprocessing stage, missing values were treated using mean imputation, where each missing observation in a numerical feature x_i was replaced with the mean of its corresponding variable, expressed as:

$$x'_i = \begin{cases} x_i, & \text{if } x_i \text{ is not missing} \\ \bar{x} = \frac{1}{n} \sum_{j=1}^n x_j, & \text{if } x_i \text{ is missing} \end{cases} \quad (1)$$

Normalization was then applied using the StandardScaler method to standardize all continuous variables, ensuring that each feature contributes equally to the clustering process. The transformation follows the z-score normalization formula:

$$z_i = \frac{x_i - \mu}{\sigma} \quad (2)$$

where, μ represents the mean and σ denotes the standard deviation of the feature. For feature selection, the AGR–CBFS technique was implemented, combining AGR, which

measures the information gain of a feature relative to entropy, with CBFS that evaluates the redundancy among features. The feature relevance score R_f for each attribute was computed as:

$$R_f = \frac{IG(f)}{H(C)} - \lambda \sum_{j=1}^m |r_{f,j}| \quad (3)$$

where, $IG(f)$ is the information gain of feature f , $H(C)$ is the entropy of the class variable, $r_{f,j}$ is the Pearson correlation coefficient between feature f and other features j , and λ is a penalty factor controlling redundancy reduction. This combined preprocessing and feature selection pipeline ensures that only the most informative and non-redundant variables are retained, leading to more stable and interpretable clustering results for e-commerce customer segmentation.

3.2 Model design and optimization

In the clustering stage, three clustering algorithms: Standard K-Means, K-Means++, and PSO-enhanced K-Means—were implemented for comparative performance analysis. In this stage, each algorithm was designed to identify optimal customer clusters based on behavioral and demographic attributes. The Standard K-Means algorithm partitions the dataset into k clusters by minimizing the within-cluster sum of squared distances (WCSS), mathematically expressed as:

$$J = \sum_{i=1}^k \sum_{x_j \in C_i} \|x_j - \mu_i\|^2 \quad (4)$$

where, x_j represents a data point, μ_i denotes the centroid of cluster C_i , and J is the objective function minimized during iteration. The K-Means++ algorithm improves upon Standard K-Means by optimizing the initial centroid selection to reduce convergence time and improve clustering stability. Instead of random initialization, it selects the first centroid randomly and subsequent centroids based on a weighted probability proportional to the squared distance from the nearest existing centroid:

$$P(x) = \frac{D(x)^2}{\sum_{x' \in X} D(x')^2} \quad (5)$$

where, $D(x)$ is the distance between a data point x and the nearest chosen centroid. To further enhance clustering quality, a PSO-enhanced K-Means algorithm was implemented, integrating PSO to optimize centroid initialization and avoid local minima. In this hybrid model, each particle represents a potential centroid position, and its movement is guided by velocity and position update equations:

$$v_i^{t+1} = wv_i^t + c_1r_1(p_i^{best} - x_i^t) + c_2r_2(g^{best} - x_i^t) \quad (6)$$

$$x_i^{t+1} = x_i^t + v_i^{t+1} \quad (7)$$

where, v_i is the velocity, x_i is the position (centroid), w is the inertia weight controlling exploration, c_1 and c_2 are acceleration constants, r_1 and r_2 are random factors, p_i^{best} is the best position of particle i , and g^{best} is the global best position across all particles. The hybrid optimization process allows PSO to search globally for optimal centroids before K-Means performs local refinement, thereby increasing convergence accuracy and segmentation stability.

3.3 Model evaluation and comparative analysis

In this stage, each clustering method is evaluated using several internal validity metrics, including the Silhouette Coefficient (SC), Davies–Bouldin Index (DBI), Calinski–Harabasz Index (CHI), and Inertia (I), formulated as follows:

$$SC = \frac{b_i - a_i}{\max(a_i, b_i)} \quad (8)$$

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} \frac{s_i + s_j}{d(c_i, c_j)} \quad (9)$$

$$CHI = \frac{\text{Tr}(B_k)/(k-1)}{\text{Tr}(W_k)/(n-k)} \quad (10)$$

$$I = \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2 \quad (11)$$

4. RESULTS AND DISCUSSION

Data preprocessing was conducted through a systematic pipeline involving missing value imputation, feature selection, and normalization using StandardScaler. The ‘‘Satisfaction Level’’ attribute was imputed using the mode method to ensure categorical balance, while continuous variables (age, total spend, items purchased, average rating, and days since last purchase) were standardized to remove scale bias before clustering. This preprocessing approach follows the data quality enhancement framework proposed by Alizadeh and Bidgoli [25], ensuring consistency and comparability across features. The fitness function employed for optimization was the Sum of Squared Errors (SSE), which quantifies the total intra-cluster variance. Lower SSE values indicate more compact and coherent clusters. The PSO algorithm was employed to minimize SSE dynamically by updating cluster centroids over 100 iterations.

As shown in Table 1, the results reveal clear performance improvements for PSO-K-Means compared to traditional clustering methods. PSO-K-Means achieved the highest Silhouette Coefficient (0.5126) and the lowest Davies–Bouldin Index (0.7674), signifying improved cohesion and separation among clusters. Despite a slightly higher Inertia (521.1470), the result aligns with the metaheuristic optimization behavior, where global centroid adjustment balances compactness with better-defined boundaries. The PSO convergence pattern illustrated in Figure 5 shows a rapid fitness improvement during the early iterations, with cumulative improvement stabilizing after approximately 25 iterations. Meanwhile, Figure 6 demonstrates the progressive refinement of fitness values, indicating PSO’s ability to achieve consistent and efficient convergence.

The comparison of methods demonstrates that PSO-K-Means significantly enhances clustering validity metrics compared to both Standard K-Means and K-Means++. The improvement in the Silhouette Coefficient (+3.75%) and reduction in the Davies–Bouldin Index (–22.8%) confirm the superior optimization capacity of the PSO algorithm. The improvement is statistically consistent with findings from Kumar et al. [26], where PSO’s adaptive velocity and inertia mechanisms enabled better centroid exploration and avoided local minima commonly found in deterministic K-Means variants. Compared to K-Means++, which enhances centroid initialization through probabilistic sampling, PSO provides dynamic exploration and exploitation balance, improving the clustering outcome over multiple iterations. The marginal increase in inertia for PSO-K-Means reflects the model’s emphasis on maximizing inter-cluster separation rather than minimizing intra-cluster variance alone, a trade-off that improves interpretability and decision boundary precision.

Table 1. Comparison methods

Methods	Silhouette	Davies-Bouldin	Calinski-Harabasz	Inertia
K-Means	0.4940	0.9942	441.6521	493.5771
K-Means++	0.4940	0.9942	441.6521	493.5771
PSO-K-Means	0.5126	0.7674	409.1090	521.1470

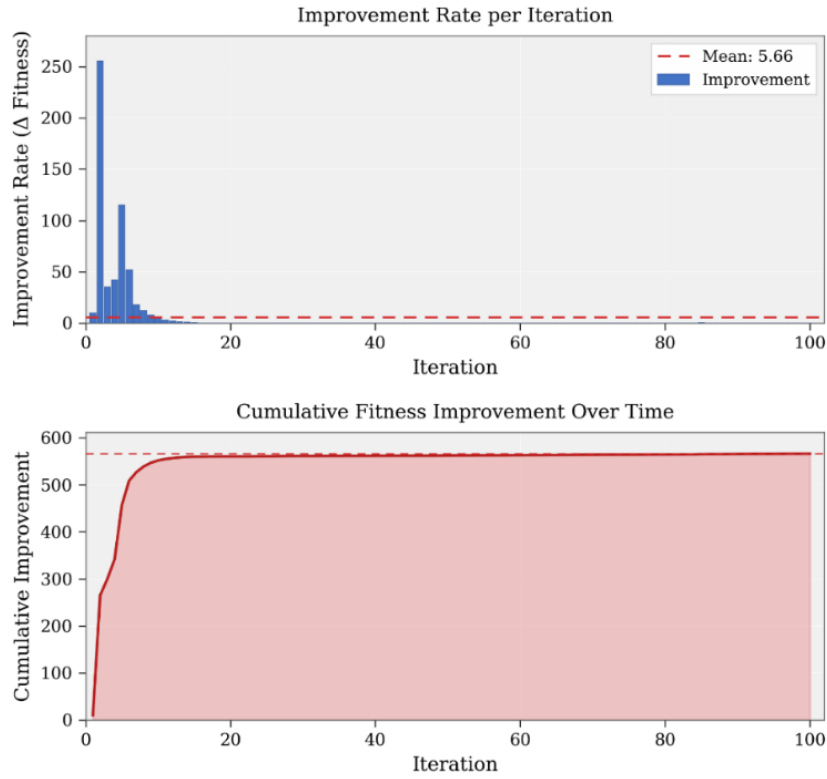


Figure 5. Particle Swarm Optimization (PSO) convergence rate analysis

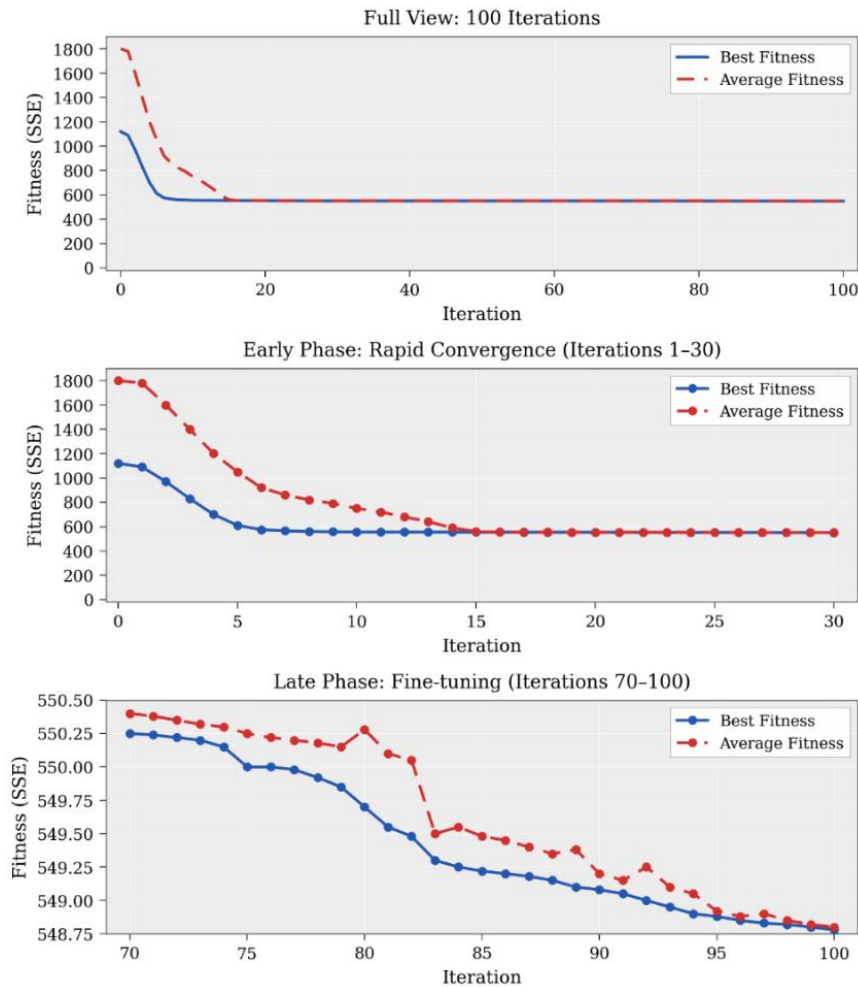


Figure 6. Convergence analysis of the Particle Swarm Optimization (PSO)-enhanced K-Means

The experimental results demonstrate that the PSO-enhanced K-Means model yields superior clustering validity compared to Standard K-Means and K-Means++, as evidenced by higher Silhouette values and substantially lower Davies–Bouldin Index scores. Beyond numerical improvement, these enhancements have direct practical implications for e-commerce customer segmentation and decision-making.

From a business perspective, improved cluster cohesion indicates that customers within the same segment exhibit more homogeneous behavioral patterns, such as similar spending levels, purchase frequency, and satisfaction tendencies. This homogeneity enables e-commerce firms to design more precise and effective personalization strategies, including tailored promotions, recommendation systems, and loyalty programs. For example, clusters characterized by high spending and frequent purchases can be prioritized for premium membership offers or exclusive incentives, while lower-engagement clusters can be targeted with reactivation or retention campaigns.

Similarly, improved inter-cluster separation implies clearer differentiation between customer segments. In operational terms, this reduces ambiguity in segment interpretation and minimizes overlap between strategic actions assigned to different customer groups. Well-separated clusters support clearer customer personas, facilitating alignment between data analytics outputs and marketing execution. This clarity is particularly important in e-commerce environments, where rapid and automated decision-making relies on stable and interpretable segmentation results [18].

The convergence behavior of the PSO-K-Means model further strengthens its business relevance. The observation that more than 80% of fitness improvement occurs within the first 20 iterations suggests that the model can achieve reliable segmentation with relatively low computational overhead. This efficiency is advantageous for real-world applications where segmentation models may need to be updated periodically as new transaction data becomes available. Faster convergence enables timely insights without sacrificing clustering quality, supporting near-real-time or iterative segmentation workflows.

Moreover, the integration of PSO-based centroid optimization with information-driven feature selection contributes to improved interpretability of the resulting clusters. By retaining only informative and non-redundant features, the segmentation outcomes align more closely with actionable business variables such as spending behavior, purchase recency, and customer satisfaction. This alignment facilitates collaboration between data analysts and business stakeholders, as segment definitions can be directly translated into marketing strategies and performance indicators.

5. CONCLUSIONS

This study proposed and evaluated a metaheuristic-optimized customer segmentation framework that integrates PSO with the K-Means clustering algorithm for e-commerce applications. The experimental results demonstrate that the PSO-enhanced K-Means consistently outperforms Standard K-Means and K-Means++ in terms of internal clustering validity. Specifically, the proposed approach achieved a higher Silhouette Coefficient (0.5126) and a substantially lower Davies–Bouldin Index (0.7674), indicating improved intra-cluster cohesion and inter-cluster separation. In addition, the

convergence analysis revealed that more than 80% of the fitness improvement was achieved within the first 20 optimization iterations, highlighting the efficiency and stability of the PSO-based optimization process.

From a practical perspective, these improvements yield more coherent and distinguishable customer segments, which are essential for effective personalization, targeted marketing, and customer retention strategies in e-commerce. By producing clusters with clearer behavioral differentiation, the proposed framework supports data-driven decision-making and enhances the interpretability of segmentation outcomes for business practitioners.

The novelty of this research lies in the unified integration of PSO-based centroid optimization and information-theoretic feature selection within an end-to-end clustering framework specifically designed for e-commerce customer segmentation. Unlike conventional approaches that treat optimization and feature selection as separate processes, this framework jointly optimizes centroid positions and feature relevance, thereby improving clustering stability, validity, and interpretability simultaneously. This integrated design represents a meaningful methodological contribution to optimization-based clustering research and extends its applicability to practical e-commerce analytics.

Despite these promising results, several limitations warrant consideration. First, the experiments were conducted on a single structured e-commerce dataset, which may limit the generalizability of the findings to other domains or datasets with different characteristics. Second, the PSO parameters were fixed throughout the optimization process; adaptive or self-tuning parameter strategies may further enhance performance. Third, the evaluation relied exclusively on internal clustering validity metrics, without incorporating external business performance indicators such as conversion rates or customer lifetime value. Future research may address these limitations by applying the proposed framework to larger and more diverse datasets, incorporating adaptive or multi-objective optimization strategies, and integrating downstream business metrics or deep representation learning to improve segmentation effectiveness further.

ACKNOWLEDGMENT

This study was funded in 2024 by the Ministry of Research, Technology, and Higher Education of the Republic of Indonesia under the Fundamental Research scheme (Grant No.: 111/E5/PG.02.00.PL/2024).

REFERENCES

- [1] Li, L., Yuan, L., Tian, J.J. (2023). Influence of online E-commerce interaction on consumer satisfaction based on big data algorithm. *Heliyon*, 9(8): e18322. <https://doi.org/10.1016/j.heliyon.2023.e18322>
- [2] Esmeli, R., Can, A.S., Awad, A., Bader-El-Den, M. (2025). Understanding customer loyalty-aware recommender systems in e-commerce: An analytical perspective. *Electronic Commerce Research*. <https://doi.org/10.1007/s10660-025-09954-6>
- [3] Alves Gomes, M., Meisen, T. (2023). A review on customer segmentation methods for personalized customer targeting in e-commerce use cases. *Information*

- Systems and e-business Management, 21(3): 527-570. <https://doi.org/10.1007/s10257-023-00640-4>
- [4] Ahmed, M., Seraj, R., Islam, S.M.S. (2020). The k-means algorithm: A comprehensive survey and performance evaluation. *Electronics*, 9(8): 1295. <https://doi.org/10.3390/electronics9081295>
- [5] Fränti, P., Sieranoja, S. (2019). How much can K-Means be improved by using better initialization and repeats?. *Pattern Recognition*, 93: 95-112. <https://doi.org/10.1016/j.patcog.2019.04.014>
- [6] Ali, A.S.H.M., Bermami, A.K., Manaa, M.E. (2025). Resource optimization of cognitive radio sensor network using hybrid metaheuristic optimization and machine learning algorithms. *Mathematical Modelling of Engineering Problems*, 12(7): 2325-2340. <https://doi.org/10.18280/mmep.120713>
- [7] Celebi, M.E., Kingravi, H.A., Vela, P.A. (2013). A comparative study of efficient initialization methods for the k-means clustering algorithm. *Expert Systems with Applications*, 40(1): 200-210. <https://doi.org/10.1016/j.eswa.2012.07.021>
- [8] Dong, Z., Jia, H., Liu, M. (2018). An adaptive multiobjective genetic algorithm with fuzzy C-means for automatic data clustering. *Mathematical Problems in Engineering*, 2018(1): 6123874. <https://doi.org/10.1155/2018/6123874>
- [9] Agbaje, M.B., Ezugwu, A.E., Els, R. (2019). Automatic data clustering using hybrid firefly particle swarm optimization algorithm. *IEEE Access*, 7: 184963-184984. <https://doi.org/10.1109/access.2019.2960925>
- [10] Amri, S., Alharis, M., Winarno, E., Putri, A.N. (2025). Average gain ratio and correlation-based feature selection for imprecise classification in algorithm C4.5. *Engineering, Technology & Applied Science Research*, 15(4): 25275-25279. <https://doi.org/10.48084/etasr.10165>
- [11] Ren, H., Hu, T.T. (2020). An adaptive feature selection algorithm for fuzzy clustering image segmentation based on embedded neighbourhood information constraints. *Sensors*, 20(13): 3722. <https://doi.org/10.3390/s20133722>
- [12] Bertsimas, D., Orfanoudaki, A., Wiberg, H. (2021). Interpretable clustering: An optimization approach. *Machine Learning*, 110(1): 89-138. <https://doi.org/10.1007/s10994-020-05896-2>
- [13] Wilbert, H.J., Hoppe, A.F., Sartori, A., Stefenon, S.F., Silva, L.A. (2023). Recency, frequency, monetary value, clustering, and internal and external indices for customer segmentation from retail data. *Algorithms*, 16(9): 396. <https://doi.org/10.3390/a16090396>
- [14] Rachwał, A., Popławska, E., Gorgol, I., Cieplak, T., Pliszczuk, D., Skowron, Ł., Rymarczyk, T. (2023). Determining the quality of a dataset in clustering terms. *Applied Sciences*, 13(5): 2942. <https://doi.org/10.3390/app13052942>
- [15] Tabianan, K., Velu, S., Ravi, V. (2022). K-Means clustering approach for intelligent customer segmentation using customer purchase behavior data. *Sustainability*, 14(12): 7243. <https://doi.org/10.3390/su14127243>
- [16] Raman, R., Kumar, V., Pillai, B.G., Rabadiya, D., Patre, S., Meenakshi, R. (2024). The impact of enhancing the K-Means algorithm through genetic algorithm optimization on high dimensional data clustering outcomes. In *2024 International Conference on Knowledge Engineering and Communication Systems (ICKECS)*, Chikkaballapur, India, pp. 1-5. <https://doi.org/10.1109/ICKECS61492.2024.10617268>
- [17] Abdunnassar, A.A., Nair, L.R. (2023). Performance analysis of Kmeans with modified initial centroid selection algorithms and developed Kmeans9+ model. *Measurement: Sensors*, 25: 100666. <https://doi.org/10.1016/j.measen.2023.100666>
- [18] Abdo, A., Abdelkader, O., Abdel-Hamid, L. (2024). SA-PSO-GK++: A new hybrid clustering approach for analyzing medical data. *IEEE Access*, 12: 12501-12516. <https://doi.org/10.1109/ACCESS.2024.3350442>
- [19] Ezugwu, A.E. (2020). Nature-inspired metaheuristic techniques for automatic clustering: A survey and performance study. *SN Applied Sciences*, 2(2): 273. <https://doi.org/10.1007/s42452-020-2073-0>
- [20] Kaur, A., Kumar, Y., Sidhu, J. (2024). Exploring meta-heuristics for partitionial clustering: Methods, metrics, datasets, and challenges. *Artificial Intelligence Review*, 57(10): 287. <https://doi.org/10.1007/s10462-024-10920-1>
- [21] Ikotun, A.M., Habyarimana, F., Ezugwu, A.E. (2025). Benchmarking validity indices for evolutionary K-means clustering performance. *Scientific Reports*, 15(1): 21842. <https://doi.org/10.1038/s41598-025-08473-6>
- [22] Sowan, B., Hong, T.P., Al-Qerem, A., Alauthman, M., Matar, N. (2023). Ensembling validation indices to estimate the optimal number of clusters. *Applied Intelligence*, 53(9): 9933-9957. <https://doi.org/10.1007/s10489-022-03939-w>
- [23] Hartanto, U.I., Buditjahjanto, I.G.P.A., Yustanti, W. (2025). Hybrid clustering and classification of at-risk customer segments in network marketing. *Journal of Information Engineering and Educational Technology (JIEET)*, 9(1): 42-50. <https://doi.org/10.26740/jieet.v9n1.p42-50>
- [24] Liu, T., Yu, H., Blair, R.H. (2022). Stability estimation for unsupervised clustering: A review. *Wiley Interdisciplinary Reviews: Computational Statistics*, 14(6): e1575. <https://doi.org/10.1002/wics.1575>
- [25] Alizadeh, H., Bidgoli, B.M. (2016). Introducing a hybrid data mining model to evaluate customer loyalty. *Engineering, Technology & Applied Science Research*, 6(6): 1235-1240. <https://doi.org/10.48084/etasr.741>
- [26] Kumar, S., Rani, R., Pippal, S.K., Agrawal, R. (2025). Customer segmentation in e-commerce: K-means vs hierarchical clustering. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, 23(1): 119-128. <http://doi.org/10.12928/telkomnika.v23i1.26384>