e-ISSN 2549-7472
Edumatic: Jurnal Pendidikan Informatika
Volume 10, Issue 1, **160-169.**
https://e-journal.hamzanwadi.ac.id/index.php/edumatic
DOI: 10.29408/edumatic.v10i1.34163

Behaviorally Interpretable Transactional Features for Customer Segmentation Using K-Means in Grocery Retail

Rendy Muhammad Aprizal [1] , Oktalia Kumala Sari [1,*] , Arif Bramantoro [2]

[1] Universitas Budi Luhur, Indonesia
[2] Universiti Teknologi Brunei, Brunei Darussalam
[*] Corresponding author: Oktalia Kumala Sari, Universitas Budi Luhur, Indonesia
✉ 2311601187@student.budiluhur.ac.id

Received: 21 February 2026 | Revised: 4 March 2026 | Accepted: 8 April 2026 | Published: 21 April 2026

Abstract
Customer segmentation based on transactional data is widely used to understand purchasing behavior in retail. However, many existing studies tend to emphasize algorithm performance, with limited discussion on how transactional variables represent actual customer behavior. This study adopts a quantitative approach using transactional sales data from a grocery retail store (Toko Solo Latri), consisting of 10,000 item-level records collected during June 2025. The analysis follows the CRISP-DM framework, covering data understanding, preparation, modeling, and evaluation stages. Customer behavior is represented through several aggregated variables, including transaction frequency, total items purchased, and product diversity. The K-Means clustering algorithm is applied to group customers into meaningful segments. The number of clusters is determined using the Elbow Method and further evaluated using Silhouette analysis. The results reveal three distinct customer segments with different levels of purchase intensity and product diversity. The Silhouette Score of 0.464 indicates a moderate clustering structure. In addition, one-way ANOVA shows significant differences across the observed variables, with large effect sizes (η^2 ranging from 0.736 to 0.822). These findings suggest that constructing behavior-based transactional features can improve the interpretability of customer segmentation results.

Keywords: behavioral segmentation; customer segmentation; k-means clustering; transactional feature construction; unsupervised learning

To cite this article: Aprizal, R. M., Sari, O. K., & Bramantoro, A. (2026). Behaviorally Interpretable Transactional Features for Customer Segmentation Using K-Means in Grocery Retail. *Edumatic: Jurnal Pendidikan Informatika*, *10*(1), 160–169. https://doi.org/10.29408/edumatic.v10i1.34163

INTRODUCTION

Retail businesses generate large volumes of transactional data from daily purchasing activities. These records provide an empirical basis for analyzing customer purchasing patterns and identifying behavioral differences among customers. In grocery retail, high-frequency purchasing and recurring consumption cycles generate detailed behavioral traces that allow firms to observe patterns of engagement, intensity, and product choice over time (Sari et al., 2024). Recent research highlights that transactional data provide a robust empirical foundation for customer segmentation and behavioral modeling in retail environments (Stylianou & Pantelidou, 2025; Tang, 2025). Transactional datasets enable retailers to uncover purchasing patterns and support more informed marketing and decision-making processes in retail environments (Firdausi et al., 2025; Riswanto et al., 2024). The growing availability of large-scale consumer data has also encouraged organizations to integrate data analytics and artificial intelligence technologies into marketing activities in order to better understand customer behavior and improve decision-making processes (Dianti et al., 2024).

Customer segmentation based on transactional data is commonly structured through the Recency, Frequency, and Monetary framework. Empirical studies show that combining RFM attributes with clustering techniques such as K-Means can generate operationally useful customer groups (Anitha & Patil, 2022; Ashraf et al., 2025). K-Means remains one of the most widely applied clustering methods due to its computational efficiency and interpretability in purchase behavior datasets (Tabianan & Velu, 2022). Several extensions of the RFM model have attempted to improve behavioral representation. Mahfuza et al. (2022) proposed LRFM variants to capture extended purchase characteristics, while Smaili and Hachimi (2023) introduced a diversity dimension to represent variation in customers' product baskets. Wang et al. (2024) further emphasize the importance of incorporating multidimensional transactional characteristics in segmentation models. Previous studies also demonstrate that RFM-based analysis combined with clustering and data mining techniques can effectively analyze purchasing behavior and support recommendation strategies in retail systems (Hsiang et al., 2023).

Although clustering techniques have been widely applied in customer segmentation, many studies emphasize algorithm performance while giving limited attention to how transactional variables represent underlying customer behavior (Turkmen, 2022; Ufeli et al., 2025). Many studies focus on improving algorithm accuracy or validation indices without clearly articulating how selected transactional variables represent underlying behavioral constructs. Consequently, transactional aggregates are frequently treated as numerical inputs rather than as conceptually grounded behavioral indicators, limiting interpretability in customer segmentation studies (Dhandayudam & Krishnamurthi, 2013).

Transactional measures can be interpreted as observable manifestations of latent consumption tendencies. Total purchase quantity reflects behavioral intensity, indicating the depth of engagement and consumption commitment within a defined period. Product diversity represents exploratory engagement and variety-seeking behavior, which has been widely discussed as a stable driver of brand switching and category exploration (Plank & Koll, 2026; Zhang, 2022). Insights derived from customer purchasing behavior can support the development of targeted marketing strategies and improve customer relationship management in retail environments (Lin et al., 2021). Recent studies also highlight that the integration of data analytics and machine learning techniques allows organizations to analyze large-scale consumer data and extract meaningful behavioral insights. These approaches enable retailers to identify hidden consumption patterns and improve decision-making processes based on empirical data analysis (Lee et al., 2024).

Although these behavioral constructs are well established in consumer behavior literature, their systematic integration into clustering-based transactional segmentation remains relatively limited. In many empirical applications, variables such as purchase quantity and product diversity are incorporated into clustering models without a clear theoretical explanation of how these variables represent underlying behavioral mechanisms. As a result, clustering results may be statistically valid but remain difficult to interpret from a behavioral perspective.

The grocery retail sector provides a particularly suitable context for investigating transactional purchasing behavior. Grocery retail is characterized by frequent purchases, recurring consumption cycles, and substantial variability in customers' purchasing patterns. These characteristics generate dense transactional datasets that capture repeated interactions between customers and retailers, making grocery retail a relevant domain for behavioral analysis based on transaction records.

This study analyzes transactional purchasing data to examine how purchase quantity and product diversity can represent behavioral indicators for customer segmentation in grocery retail. Specifically, total purchase quantity is interpreted as an indicator of purchasing intensity, while product diversity represents exploratory consumption behavior. These behavioral indicators are evaluated using the K-Means clustering algorithm to identify meaningful customer segments. This study contributes by interpreting transactional aggregates as behavioral indicators, demonstrating their integration within a K-Means framework, and providing empirical evidence that such features enhance the interpretability of clustering-based customer segmentation.

METHOD

This study uses transactional sales data obtained from a grocery retail store in Ciledug, Indonesia. The dataset consists of 10,000 item-level transaction records collected from 1 to 30 June 2025. Each record includes transaction ID, customer ID, product code, product name, and transaction date. Data cleaning was performed to remove incomplete and duplicate records. The analysis follows the CRISP-DM workflow, including business understanding, data understanding, data preparation, modeling, evaluation, and interpretation. The research objective is to identify customer segments based on observable purchasing behavior in grocery retail.

Although the data are recorded at the transaction level, the unit of analysis is the customer. All transactions associated with the same customer ID were aggregated to construct a single behavioral profile per customer. This ensures that clustering results represent customer-level behavior rather than isolated transaction patterns. Five transactional indicators were derived at the customer level: total_transactions, total_items, unique_products, avg_items_per_transaction, and purchase_days. These variables represent visit frequency, cumulative purchase intensity, product breadth, basket depth, and temporal engagement. All derived indicators were included in the clustering process. However, total purchase quantity and product diversity serve as the primary dimensions for behavioral interpretation, as they directly reflect engagement intensity and exploratory breadth. Similar behavioral aggregation approaches have been reported in recent retail segmentation research (Suh, 2025).

Both features were standardized using z-score normalization because K-Means relies on Euclidean distance and is sensitive to scale differences. Customer segmentation was conducted using the K-Means algorithm, which is widely applied in retail analytics due to its efficiency and interpretability (Guney et al., 2024). The algorithm was executed with multiple random initializations with n_init set to 50 to reduce sensitivity to centroid initialization. The solution with the lowest within cluster sum of squares was retained. Cluster stability was examined through repeated runs on resampled subsets of the data.

The optimal number of clusters was determined using the Elbow Method and supported by the Silhouette Score. One-way ANOVA was conducted to assess differences in behavioral features across clusters. In this study, ANOVA is used as a descriptive assessment of between cluster separation rather than as a causal test. Effect size was calculated using eta squared. The dataset is derived from a single retail store over a one-month period; therefore, findings reflect segmentation within this specific context and should be generalized cautiously.

RESULTS AND DISCUSSION

Results

The dataset contains 10,000 item-level transaction records collected during June 2025. After aggregating transactions by customer ID, 1,547 customer-level profiles were obtained for clustering analysis. Each record includes transaction_id, customer_id, product_code, product_name, and transaction date. Table 1 presents a sample of the transactional dataset. The table illustrates how each purchase transaction is recorded at the item level, where a single transaction may contain multiple purchased products associated with different customers and product codes. This structure reflects typical retail transaction data in which individual items are stored as separate records before being aggregated for behavioral analysis.

Table 1. Sample of transaction data

transaction_id	customer_id	product_code	product_name	date
TRX2025000001	CUST000001	2102	*Gula Pasir 1kg*	01/06/2025
TRX2025000035	CUST000002	2114	*Teh Celup 25pcs*	01/06/2025
TRX2025000035	CUST000003	2113	*Susu Kental Manis*	01/06/2025
TRX2025000035	CUST000004	2112	*Gula Pasir 1kg*	01/06/2025
TRX2025000035	CUST000005	2107	*Minyak Goreng 1L*	01/06/2025

The original dataset records purchase activity at the item level, where each row represents an individual product transaction. However, since customer segmentation is conducted at the customer level, it is necessary to transform the data into a more suitable representation. Therefore, all transactions associated with the same customer_id were aggregated to construct a unified behavioral profile for each customer. This aggregation process enables the extraction of meaningful behavioral indicators, such as total purchase quantity and product diversity, which reflect purchasing intensity and exploratory consumption patterns. By converting item-level data into customer-level features, the dataset becomes more appropriate for clustering analysis, allowing the identification of distinct customer segments based on their overall purchasing behavior.

This aggregation produced 1,547 customer-level profiles summarizing purchasing activity during the observation period. Five transactional indicators were derived for each customer: *total_transactions*, *total_items*, *unique_products*, *avg_items_per_transaction*, and *purchase_days*. The *total_transactions* variable represents the number of purchase visits made by a customer, while *total_items* indicates the overall quantity of products purchased. The *unique_products* variable reflects product diversity, capturing the variety of items purchased across transactions. The *avg_items_per_transaction* measures the average number of items bought in each visit, providing an indication of purchase intensity per transaction. Finally, *purchase_days* represents the number of distinct days on which purchases occurred, reflecting the temporal frequency of customer shopping behavior during the observation period.

Table 2. Aggregated transactional features

customer_id	total_transactions	total_items	unique_products	avg_items_per_transaction	purchase_days
CUST0001	8	45	15	5.62	7
CUST0002	4	22	13	5.50	4
CUST0003	1	10	8	10.00	1

The aggregated dataset provides a structured representation of customer purchasing behavior. The observed variation across total_transactions, total_items, and unique_products indicates meaningful heterogeneity in engagement frequency, purchase intensity, and product diversity. This behavioral dispersion supports the application of clustering techniques to identify structured customer segments.

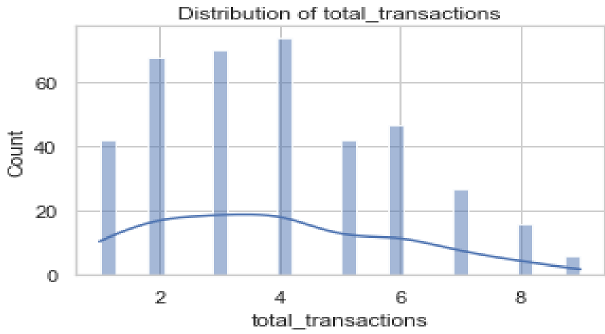

Figure 1. Distribution of total transactions

Figure 1 shows the distribution of *total_transactions* among customers during the observation period. Most customers make between two and five purchase visits, indicating a moderate shopping frequency. The highest concentration appears around three to four transactions. A smaller number of customers record higher visit frequencies, reaching up to eight or nine transactions. This distribution suggests variation in customer shopping activity, where most customers shop occasionally while a few visit the store more frequently.

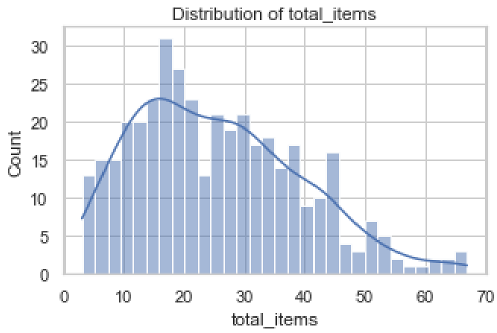

Figure 2. Distribution of total items

The distribution of *total_items*, as illustrated in Figure 2, indicates that most customers purchase between 10 and 35 items during the observation period, reflecting a moderate level of purchasing activity. The highest concentration is observed within this middle range. Meanwhile, a smaller proportion of customers purchase more than 50 items, suggesting the presence of higher-intensity buyers. Overall, this pattern highlights variability in purchasing intensity, where the majority demonstrate moderate behavior while a minority exhibits significantly higher consumption levels.

The distribution of *unique_products*, shown in Figure 3, reveals that most customers purchase between 8 and 14 distinct products, although some display lower levels of product diversity. This variation supports the interpretation of product breadth as a behavioral dimension associated with exploratory tendencies. Collectively, these distributions demonstrate sufficient variability and dispersion, justifying the application of clustering analysis.

Illustrates the Elbow Method in Figure 4 used to determine the optimal number of clusters. The inertia value decreases sharply as the number of clusters increases from $k = 1$ to $k = 3$. After this point, the reduction becomes more gradual, indicating diminishing improvement in cluster compactness. The visible change in slope forms an elbow at $k = 3$. This suggests that three clusters provide a reasonable balance between model simplicity and within-cluster variance, and therefore were selected for the clustering analysis.

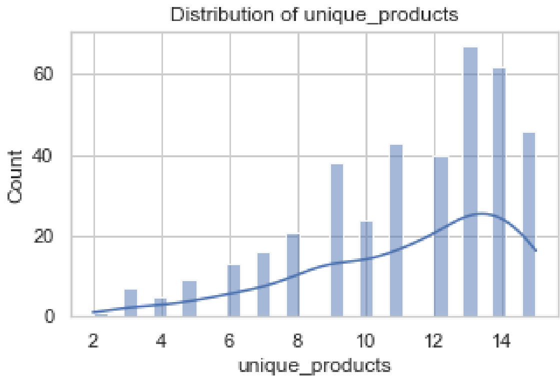

Figure 3. Distribution of unique products

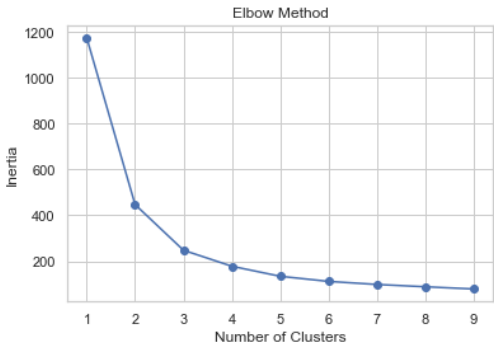

Figure 4. Elbow Method

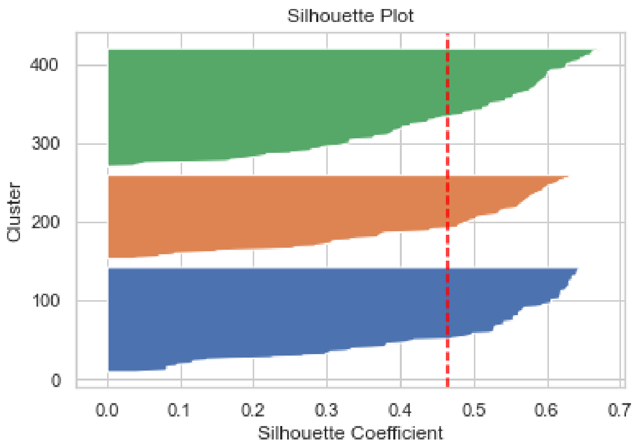

Figure 5. Silhouette analysis

The Silhouette analysis is presented in Figure 5, with an overall Silhouette Score of 0.464, indicating a moderate level of separation among clusters. The majority of observations exhibit positive silhouette coefficients, suggesting that most data points are appropriately assigned to their respective clusters. However, a limited number of observations are situated near zero values, indicating the presence of overlap between clusters and reflecting gradual transitions in behavioral patterns rather than sharply defined boundaries.

The allocation of customers across clusters is presented in Figure 6. A closer inspection shows that Cluster 2 constitutes 38.52% of the total customers, followed by Cluster 0 with

34.18% and Cluster 1 with 27.30%. Such a distribution reflects a relatively even segmentation, indicating that no single cluster disproportionately dominates the dataset and suggesting the presence of meaningful behavioral diversity among customers. To further evaluate the robustness of the clustering results, the model was executed multiple times using different random initializations. The resulting Silhouette Scores remained highly consistent, with only minimal variation observed, thereby confirming the stability and reliability of the three-cluster solution irrespective of centroid initialization.

cluster	count	percentage
0	134	34.18
1	107	27.30
2	151	38.52

Figure 6. Customer distribution across clusters

One-way ANOVA was conducted to quantify the mean differences across clusters. The results show that all behavioral variables exhibit statistically significant differences among clusters ($p < 0.001$). Effect size analysis using eta squared indicates substantial between-cluster variance, with values of 0.822 for *total_transactions*, 0.810 for *total_items*, and 0.736 for *unique_products*. These relatively high values suggest that cluster membership explains a large proportion of variance in the observed behavioral indicators. In other words, customers in different clusters demonstrate clearly distinguishable purchasing behaviors in terms of visit frequency, purchase quantity, and product diversity. However, since clustering was performed using the same variables, the ANOVA results are interpreted as an indication of differentiation magnitude rather than as an independent validation of the clustering model.

cluster	total_transactions	total_items	unique_products
0	1.910448	11.537313	7.753731
1	6.607477	43.289720	14.046729
2	3.867550	25.311258	12.370861

Figure 7. Cluster mean behavioral profiles

The mean behavioral characteristics associated with each cluster are illustrated in Figure 7, revealing clear differentiation across the variables *total_transactions*, *total_items*, and *unique_products*. Cluster 1 demonstrates the highest level of purchasing intensity and product diversity, with mean values of 6.60 transactions, 43.29 total items, and 14.05 unique products, thereby representing a segment of highly engaged customers with extensive purchasing patterns. In contrast, Cluster 0 records the lowest values across all indicators, with averages of 1.91 transactions, 11.53 total items, and 7.75 unique products, reflecting a group characterized by low engagement and limited purchasing breadth. Cluster 2 occupies an intermediate position, indicating a moderate level of engagement and transactional activity. The progressive increase in these behavioral indicators across clusters reinforces the interpretability of the segmentation structure and is consistent with the intended behavioral conceptualization of transactional features. Furthermore, the ordered pattern observed suggests the presence of a systematic gradient in engagement intensity, rather than an arbitrary numerical segmentation.

A two-dimensional projection of customer segments derived through Principal Component Analysis (PCA) is presented in Figure 8, aiming to elucidate the structural separation of clusters within a reduced-dimensional representation. It is noteworthy that the clustering procedure was performed using standardized behavioral features without the application of dimensionality reduction techniques. The projection reveals three discernible groups distributed along the first principal component, although a certain degree of overlap is evident between adjacent clusters. Overall, the observed pattern indicates a stable and consistent segmentation structure, primarily driven by variations in purchasing intensity and product diversity.

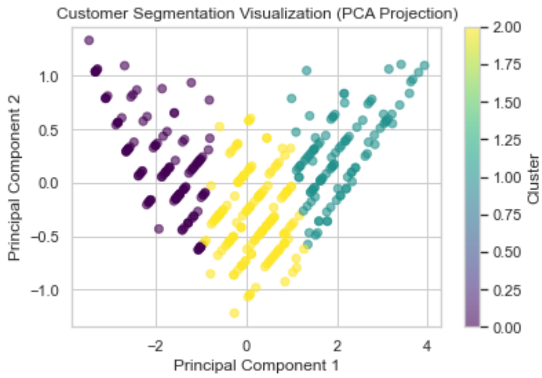

Figure 8. PCA projection of customer segments

Discussion

The results indicate that customer segmentation derived from transactional data forms a coherent behavioral structure characterized by a continuum of engagement intensity. The differentiation into high-, moderate-, and low-engagement segments reflects systematic variation in purchasing intensity and product diversity, supporting the interpretation of transactional aggregates as observable proxies of latent consumption behavior (Dhandayudam & Krishnamurthi, 2013; Stylianou & Pantelidou, 2025). Consistent with prior studies, purchase quantity and diversity emerge as key dimensions in distinguishing customer behavior (Smaili & Hachimi, 2023; Zhang, 2022; Plank & Koll, 2026). However, unlike these studies, which primarily treat such variables as input features for improving clustering accuracy, the present findings demonstrate their explicit role as behaviorally grounded constructs that enhance interpretability.

A notable limitation in the existing literature is the prevailing emphasis on algorithmic optimization, whereby enhancements in clustering performance are frequently prioritized at the expense of conceptual interpretability (Tabianan & Velu, 2022; Turkmen, 2022; Ufeli et al., 2025). Although prior studies often report strong clustering validity, they tend to offer limited insight into how transactional variables correspond to underlying behavioral mechanisms. The present findings demonstrate that aligning transactional features with theoretically grounded behavioral constructs leads to clusters with greater semantic interpretability and analytical coherence. This evidence further reinforces recent research highlighting the critical role of feature representation in advancing customer analytics (Wang et al., 2024; Lee et al., 2024).

In contrast to conventional RFM-based segmentation approaches, which primarily focus on recency, frequency, and monetary value (Anitha & Patil, 2022; Ashraf et al., 2025), this study demonstrates that the inclusion of product diversity adds a distinct dimension of behavioral differentiation. Existing RFM-oriented studies tend to inadequately capture exploratory consumption behavior, whereas the present findings indicate that product diversity effectively reflects variations in product selection breadth, particularly within grocery retail settings characterized by frequent transactions and multi-category purchasing patterns (Stylianou & Pantelidou, 2025; Tang, 2025).

The moderate cluster separation (Silhouette Score = 0.464) and the overlap observed in the PCA projection (Figure 8 on page 9) are consistent with recent findings that customer behavior in transactional datasets tends to form continuous rather than sharply bounded segments (Stylianou & Pantelidou, 2025). This contrasts with earlier studies that implicitly assume well-separated clusters, suggesting that moderate separation should be interpreted as a realistic representation of behavioral complexity rather than a methodological limitation.

This study contributes by demonstrating that segmentation quality is not solely dependent on algorithmic sophistication, as commonly emphasized in prior research, but is critically influenced by the behavioral interpretability of input features. This provides a conceptual extension to existing customer segmentation literature by integrating behavioral theory with transactional data modeling, resulting in more interpretable and actionable segmentation outcomes.

CONCLUSION

This study demonstrates that customer segmentation based on transactional data can produce a coherent and interpretable behavioral structure when feature construction is grounded in theoretically meaningful constructs. Purchasing intensity and product diversity are shown to function as robust behavioral indicators, enabling the identification of customer segments that reflect a continuum of engagement rather than rigid categorical boundaries. The findings contribute to the customer analytics literature by emphasizing that the interpretability and analytical value of clustering outcomes are primarily determined by the conceptual alignment of input features with underlying behavioral mechanisms, thereby extending prior research that predominantly prioritizes algorithmic performance. In practical terms, the resulting segments provide actionable insights for targeted marketing and customer relationship management strategies. Nevertheless, the generalizability of the findings is limited by the use of a single-store dataset and a short observation period, suggesting that future research should incorporate longitudinal data, additional behavioral dimensions, and multi-context datasets to further strengthen the robustness of the proposed approach.

REFERENCES

Anitha, P., & Patil, M. M. (2022). RFM model for customer purchase behavior using K-Means algorithm. *Journal of King Saud University - Computer and Information Sciences*, *34*(5), 1785–1792. https://doi.org/10.1016/j.jksuci.2019.12.011

Ashraf, A., Rayed, C. A., Awad, N. A., & Sabry, H. M. (2025). ScienceDirect A Framework for Customer Segmentation to Improve Marketing Strategies Using Machine Learning. *Procedia Computer Science*, *260*, 616–625. https://doi.org/10.1016/j.procs.2025.03.240

Dhandayudam, P., & Krishnamurthi, I. (2013). Customer Behavior Analysis Using Rough Set Approach. *Journal of Theoretical and Applied Electronic Commerce Research*, *8*(2), 21–33. https://doi.org/10.4067/S0718-18762013000200003

Dianti, A. R., Mualifah, T., & Dirgantara, I. M. B. (2024). Artificial Intelligence for Marketing: Systematic Literature Review. *Research Horizon*, *04*(06). https://doi.org//10.54518/rh.4.6.2024.393

Firdausi, A. A., Hartanti, D., & Sari, A. A. (2025). A Recommendation System Using the Content-Based Filtering Method for Batrisyia Herbal Face Care Products. *Journal Of Information Systems And Computer Engineering*, *10*(2), 232–237. https://doi.org/10.51876/simtek.v10i2.1569

Guney, S., Peker, S., & Turhan, C. (2024). A Combined Approach for Customer Profiling in Video on Demand Services using Clustering and Association Rule Mining. *IEEE Access*, 1–11. https://doi.org/10.1109/ACCESS.2020.2992064

Hsiang, A., Chen, L., & Gunawan, S. (2023). applied sciences Enhancing Retail Transactions :

A Data - Driven Recommendation Using Modified RFM Analysis and Association Rules Mining. *Applied Sciences*, *13*(18). https://doi.org/10.3390/app131810057

Lee, Y. U., Chung, S. H., & Park, J. Y. (2024). Online Review Analysis from a Customer Behavior Observation Perspective for Product Development. *Sustainability Article*, *16*(9), 1–18. https://doi.org/10.3390/su16093550

Lin, R., Chuang, W., Chuang, C., & Chang, W. (2021). Applied Big Data Analysis to Build Customer Product Recommendation Model. *Sustainability*, *13*(9). https://doi.org/doi.org/10.3390/su13094985

Mahfuza, R., Islam, N., & Emon, A. F. (2022). *LRFMV : An efficient customer segmentation model for superstores*. 1–29. https://doi.org/10.1371/journal.pone.0279262

Plank, A., & Koll, O. (2026). Who seeks variety ? Profiles and behaviors of consumers with varying brand choice concentration. *Journal of Retailing*, *xxxx*, 1–24. https://doi.org/10.1016/j.jretai.2026.01.004

Riswanto, A. L., Ha, S., Lee, S., & Kwon, M. (2024). Online Reviews Meet Visual Attention : A Study on Consumer Patterns in Advertising , Analyzing Customer Satisfaction , Visual Engagement , and Purchase Intention. *Journal of Theoretical and Applied Electronic Commerce Research*, *19*(4), 3102–3122. https://doi.org/10.3390/jtaer19040150

Sari, A. A., Pramono, P., Saputra, I. T., & Prakoso, A. D. (2024). Optimalisasi Proses Digitalisasi UMKM melalui Aplikasi Marketplace berbasis Design Thinking. *Edumatic: Jurnal Pendidikan Informatika*, *8*(2), 535–544. https://doi.org/10.29408/edumatic.v8i2.27702

Smaili, M. Y., & Hachimi, H. (2023). New RFM-D classification model for improving customer analysis and response prediction. *Ain Shams Engineering Journal*, *14*(12), 102254. https://doi.org/10.1016/j.asej.2023.102254

Stylianou, T., & Pantelidou, A. (2025). A machine learning approach to consumer behavior in supermarket analytics. *Decision Analytics Journal*, *16*(January), 100600. https://doi.org/10.1016/j.dajour.2025.100600

Suh, Y. (2025). Discovering customer segments through interaction behaviors for home appliance business. In *Journal of Big Data*. Springer International Publishing. https://doi.org/10.1186/s40537-025-01111-y

Tabianan, K., & Velu, S. (2022). K-Means Clustering Approach for Intelligent Customer Segmentation Using Customer Purchase Behavior Data. *Sustainability (Switzerland)*, *14*(12), 1–15. https://doi.org/10.3390/su14127243

Tang, J. (2025). Unlocking Retail Insights : Predictive Modeling and Customer Segmentation Through Data Analytics. *Journal of Theoretical and Applied Electronic Commerce Research*, *20*(2), 1–20. https://doi.org/10.3390/jtaer20020059

Turkmen, B. (2022). Customer segmentation with machine learning for online retail industry. *The European Journal of Social and Behavioural Sciences*, *31*(2). https://doi.org/10.15405/ejsbs.316

Ufeli, C. P., Sattar, M. U., & Hasan, R. (2025). Enhancing Customer Segmentation Through Factor Analysis of Mixed Data (FAMD)-Based Approach Using K-Means and Hierarchical Clustering Algorithms. *Information*, *16*(6), 1–25. https://doi.org/10.3390/info16060441

Wang, S., Sun, L., & Yu, Y. (2024). A dynamic customer segmentation approach by combining LRFMS and multivariate time series clustering. *Scientific Reports*, 1–18. https://doi.org/10.1038/s41598-024-68621-2

Zhang, Y. (2022). Variety-Seeking Behavior in Consumption : A Literature Review and Future Research Directions. *Frontiers in Psychology*, *18*(6). https://doi.org/10.3389/fpsyg.2022.874444