2528-9705

Journal of Organizational Behavior Research

Vol.: 11, Is.: 3, Year: 2026, Pages: 49-61

https://doi.org/10.51847/E1T6Apo8ML

Available online

https://odad.org/

Customer Purchase Behavior Analysis Using Extreme Learning Machines for Data-Driven Consumer Prediction and Strategic Market Segmentation

Huixian Zheng[1], Kittipol Wisaeng[2*]

[1]Mahasarakham Business School, Mahasarakham University, Thailand.
[2]Business School, Mahasarakham University, Thailand.

***Corresponding Author**
E-mail: Kittipol.w@acc.msu.ac.th

ABSTRACT

Understanding customer purchase behavior is essential for organizations seeking to strengthen customer engagement, refine marketing strategies, and improve business performance in highly competitive digital markets. This study develops a hybrid Customer Segmentation and Structural Equation Modeling (SEM) framework to examine the determinants of customer purchasing behavior and purchasing intention. The proposed framework integrates K-Means clustering with SEM, enabling the simultaneous identification of homogeneous customer groups and assessment of structural relationships among Customer Demographics (CD), Product Preference (PP), Purchasing Behavior (PB), Customer Segmentation (CS), and Purchasing Intention (PI). A dataset containing 8,000 customer records was analyzed through data preprocessing, exploratory data analysis, clustering, and structural modeling. The K-Means results identified four meaningful customer segments: Budget Customers, Regular Customers, Premium Customers, and High-Value Customers. The SEM results indicated a satisfactory overall model fit (CFI = 0.953, TLI = 0.947, RMSEA = 0.051, and SRMR = 0.043). Product Preference was found to exert a significant positive effect on Purchasing Behavior (β = 0.451, p < 0.001), while Purchasing Behavior demonstrated the strongest direct influence on Purchasing Intention (β = 0.536, p < 0.001). In addition, Customer Segmentation significantly mediated the relationship between Purchasing Behavior and Purchasing Intention. These findings confirm that combining machine-learning-based customer segmentation with SEM provides a robust analytical framework for understanding both customer heterogeneity and behavioral mechanisms. The proposed approach offers practical value for personalized marketing, customer targeting, resource allocation, and data-driven strategic decision-making across increasingly complex digital commerce environments.

Keywords: Customer purchase behavior, Purchasing intention, Customer segmentation, Structural equation modeling, K-means clustering, Consumer analytics.

Introduction

The rapid expansion of digital commerce and data-driven marketing has fundamentally transformed how organizations understand customer behavior and purchasing decisions. The increasing availability of consumer data generated through online retail platforms enables businesses to analyze purchasing patterns, preferences, and decision-making processes more effectively, thereby supporting customer retention, profitability, and competitive advantage (Anwar *et al*., 2025). Customer purchase behavior is influenced by multiple factors, including demographic characteristics, purchasing frequency, spending patterns, product preferences, and prior purchasing experiences. Consequently, organizations require analytical approaches that simultaneously capture observable customer characteristics and the latent factors that influence purchase intentions (Son, 2026). Structural Equation Modeling (SEM) has become one of the most widely adopted techniques for examining complex causal relationships among latent constructs while accounting for measurement error (Son & Trung, 2024; Abreu-Ledón *et al*., 2026; Thanh & Trung, 2026). Previous

Recieved: 22.03.2026 –Accepted: 30.06.2026 –Published: 15.07.2026

studies have successfully applied SEM to investigate determinants of customer trust, satisfaction, loyalty, and purchasing intention in digital commerce (Anora *et al*., 2025; Thanh & Trung, 2026). Although SEM effectively explains behavioral relationships, it has limited capability to identify heterogeneous customer groups with distinct purchasing characteristics. Customer segmentation addresses this limitation by grouping customers based on similar demographic and behavioral attributes, enabling organizations to implement personalized marketing strategies. K-Means is still one of the most widely used clustering techniques because of its computational efficiency and ability to identify relevant consumer segments in large datasets (Thakur *et al*., 2025). Nevertheless, previous research has generally applied customer segmentation and SEM independently, resulting in fragmented insights into consumer behavior (Dominique *et al*., 2023).

To overcome this limitation, this study proposes a hybrid analytical framework that integrates K-Means customer segmentation with Structural Equation Modeling to simultaneously identify customer segments and examine the relationships among Customer Demographics, Product Preference, Purchasing Behavior, Customer Segmentation, and Purchasing Intention. By combining machine-learning-based segmentation with behavioral modeling, the proposed framework provides both predictive and explanatory insights, thereby supporting more effective customer analytics and evidence-based marketing decision-making (Ali *et al*., 2025; Ferdinand & Wijaya, 2025).

Literature Review

Customer Purchase Behavior

Customer purchase behavior is a fundamental topic in marketing and consumer analytics, encompassing the processes through which consumers recognize needs, evaluate alternatives, make purchasing decisions, and assess post-purchase satisfaction (Balaji *et al*., 2025). Understanding these processes enables organizations to improve customer retention, enhance profitability, and develop effective marketing strategies. With the rapid growth of digital commerce, purchasing behavior has become increasingly complex due to the combined influence of traditional demographic characteristics and digital interactions. Previous studies have identified demographic characteristics, income, lifestyle, product quality, social influence, and purchasing experiences as key determinants of customer behavior. In particular, demographic variables such as age, gender, education, and income significantly influence purchasing power, product preferences, and consumption patterns (Umar *et al*., 2025). The expansion of e-commerce has further increased the availability of behavioral data, allowing organizations to analyze customer purchasing patterns and optimize personalized marketing strategies. Consequently, customer analytics has become an essential tool for predicting future purchasing behavior and supporting data-driven decision-making (Modepalli, 2025). Nevertheless, consumer markets remain highly heterogeneous, with customers exhibiting diverse purchasing preferences and behavioral characteristics. This diversity has motivated researchers to adopt customer segmentation techniques that classify consumers into homogeneous groups, providing more effective insights for targeted marketing and customer relationship management.

Purchasing Intention

Purchasing intention refers to a consumer's willingness to purchase a product or service in the future and is widely recognized as a strong predictor of actual purchasing behavior. The Theory of Planned Behavior (TPB) posits that purchase intention is shaped by attitudes, subjective norms, and perceived behavioral control, making it one of the most widely applied theories in consumer behavior research. Previous studies have shown that purchasing intention is influenced by perceived value, product quality, brand trust, customer satisfaction, and social influence (Octavia & Handayani, 2025). In digital commerce, additional factors such as website usability, perceived security, electronic word-of-mouth, and online customer reviews further affect consumers' purchase decisions (Hamza, 2023). Moreover, customers with higher purchase frequency and positive shopping experiences generally exhibit stronger future purchase intentions. Recent advances in machine learning and behavioral analytics have enhanced the prediction of purchasing intention, while structural modeling has improved the understanding of causal relationships among behavioral constructs (Singh *et al*., 2025). To give a more thorough explanation of purchase intention, integrated frameworks that mix structural modelling and client segmentation are still required.

Customer Segmentation
Customer segmentation is a strategic marketing approach that classifies customers into relatively homogeneous groups based on shared characteristics and purchasing behaviors, enabling organizations to develop targeted marketing strategies and improve business performance. Among segmentation approaches, behavioral segmentation has received considerable attention because it reflects actual purchasing activities through variables such as purchase frequency, spending patterns, product preferences, and customer loyalty (Sangami, 2025). With the rapid growth of big data analytics, clustering algorithms have become widely adopted for identifying customer groups. K-Means clustering remains one of the most popular techniques due to its computational efficiency and ability to partition large datasets into meaningful segments by maximizing between-cluster differences while minimizing within-cluster variance. Previous studies have successfully applied K-Means clustering across retail, banking, telecommunications, tourism, and e-commerce sectors (Sutantio *et al*., 2025). Nevertheless, customer segmentation primarily provides descriptive insights and does not explain the causal relationships underlying consumer behavior. Therefore, integrating customer segmentation with Structural Equation Modeling offers a more comprehensive framework for simultaneously identifying customer groups and explaining the determinants of purchasing behavior and purchasing intention.

Structural Equation Modeling in Consumer Behavior Research
Structural Equation Modeling (SEM) is a widely adopted analytical technique in marketing and behavioral research because it enables the simultaneous examination of relationships among multiple latent constructs while accounting for measurement error (Rahman *et al*., 2015). SEM combines factor analysis and regression analysis within a unified framework, making it particularly suitable for investigating constructs such as customer satisfaction, trust, loyalty, and purchasing intention. Previous studies have successfully applied SEM to explain customer behavior in online retail settings. For example, Wathanakom (2025) reported that trust and customer satisfaction significantly influence purchasing intention, while Mahpour *et al*. (2025) demonstrated that perceived value and user experience are key determinants of online purchasing behavior. Despite its strong explanatory power, SEM assumes homogeneous relationships across all observations and may therefore fail to adequately capture heterogeneity among customer groups. Integrating customer segmentation with SEM provides a more comprehensive framework for explaining consumer behavior across distinct market segments.

Hybrid Approaches Combining Segmentation and SEM
The increasing availability of customer data has accelerated the development of hybrid analytical approaches that integrate machine learning with statistical modeling. These frameworks combine the predictive capability of data-driven algorithms with the explanatory power of behavioral theories. Ferdinand and Wijaya (2025) proposed that customer segmentation should be conducted prior to Structural Equation Modeling (SEM) to improve model accuracy and interpretability, thereby enabling the examination of behavioral relationships across different customer groups. Likewise, Richter and Tudoran (2024) highlighted that combining clustering techniques with SEM provides both predictive and explanatory insights into consumer decision-making. Recent research also emphasizes that hybrid frameworks generate more actionable marketing intelligence than standalone analytical methods (Umoren *et al*., 2025). Nevertheless, empirical studies integrating customer segmentation and SEM remain limited, particularly in examining how behavior-based customer segments influence purchasing intention through the relationships among customer demographics, product preferences, and purchasing behavior.

Research Gap and Conceptual Framework
Despite extensive research, customer segmentation and Structural Equation Modeling (SEM) have mostly been applied independently. Customer segmentation identifies behavioral groups but cannot explain causal relationships, whereas SEM explains behavioral mechanisms but often overlooks customer heterogeneity. To overcome these limitations, this study proposes a hybrid framework integrating K-Means clustering and SEM. The framework simultaneously identifies customer segments and examines the relationships among Customer Demographics, Product Preference, Purchasing Behavior, Customer Segmentation, and Purchasing Intention, providing a more comprehensive approach to customer behavior analysis and supporting data-driven marketing decision-making.

Conceptual Framework and Hypothesis Development

Theoretical Foundation

This study integrates concepts from consumer behavior theory, market segmentation theory, and the Theory of Planned Behavior (TPB) to explain customer purchase behavior and purchasing intentions. According to TPB, behavioral intentions are influenced by attitudes, subjective norms, and perceived behavioral control (Ganesha & Hegde, 2025). In marketing contexts, demographic characteristics, purchasing experiences, and product preferences influence consumers' attitudes and behavioral intentions toward purchasing products and services. Customer segmentation theory suggests that consumers can be classified into homogeneous groups based on similar characteristics and behavioral patterns (Chea, 2024). Segment-specific insights enable organizations to develop personalized marketing strategies and improve customer relationship management. By integrating customer segmentation with Structural Equation Modeling (SEM), the present study seeks to explain both behavioral heterogeneity and causal relationships among the determinants of purchasing intentions.

Conceptual Framework

The proposed conceptual framework comprises five latent constructs: Customer Demographics (CD), Product Preference (PP), Purchasing Behavior (PB), Customer Segmentation (CS), and Purchasing Intention (PI). Customer Demographics and Product Preference are modeled as antecedent variables influencing Purchasing Behavior and Purchasing Intention. Purchasing Behavior directly affects Purchasing Intention, reflecting how prior purchasing experiences influence future buying decisions. Customer Segmentation, derived from K-Means clustering, captures customer heterogeneity by classifying consumers into distinct behavioral groups and is hypothesized to directly influence Purchasing Intention. By integrating customer segmentation with behavioral and demographic factors, the proposed framework provides a comprehensive approach to explaining purchasing intentions and supporting data-driven marketing decision-making. The conceptual framework is presented in **Figure 1**.

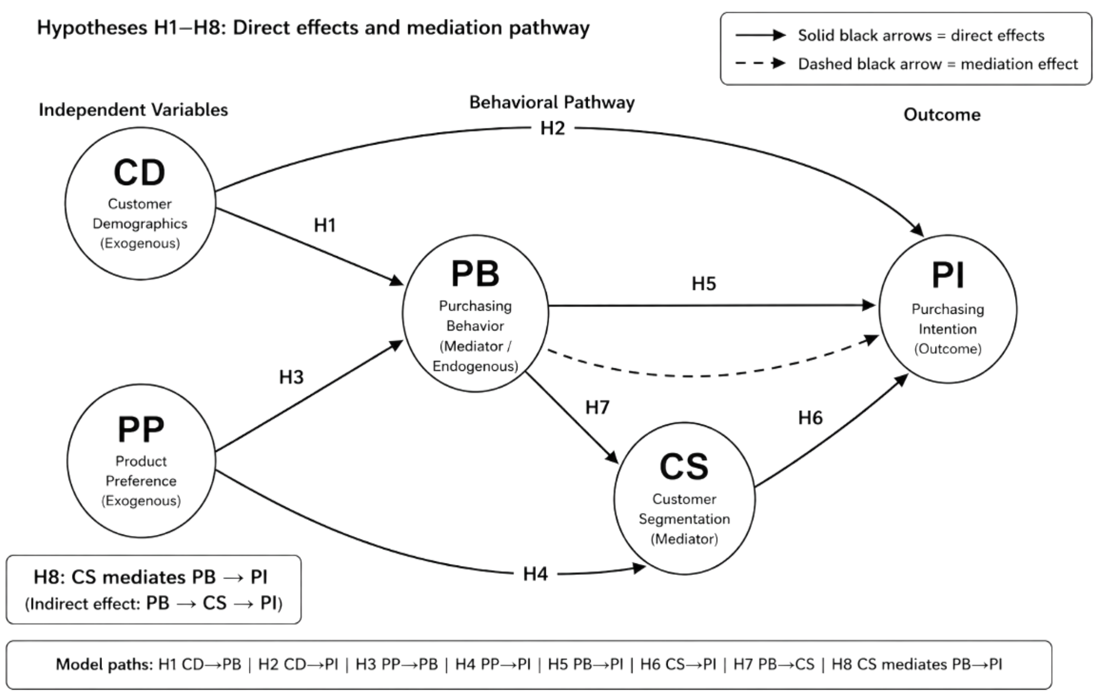

Figure 1. Proposed SEM research model for customer purchase behavior and purchasing intention.

Hypotheses Development

Customer Demographics and Purchasing Behavior

Customer demographics constitute a fundamental determinant of consumer purchasing behavior and have been extensively examined in marketing and consumer behavior research. Demographic characteristics such as age, gender, income, and socioeconomic status influence consumers' purchasing power, consumption patterns, product preferences, and shopping frequency. Prior studies have demonstrated that consumers with higher income levels generally exhibit greater purchasing frequency and higher spending levels due to increased financial capacity and access to a wider range of products and services (Sravani *et al*., 2025). Furthermore, demographic differences often shape consumers' perceptions of value and purchasing priorities, leading to variations in behavioral outcomes. Therefore, demographic characteristics are expected to significantly influence purchasing behavior.

H1: Customer demographics positively influence purchasing behavior.

Customer Demographics and Purchasing Intention

Customer demographics not only affect actual purchasing behavior but also influence consumers' future purchase intentions. Demographic attributes shape purchasing capabilities, lifestyle preferences, and decision-making processes, thereby affecting consumers' willingness to purchase products and services in the future. Previous research suggests that consumers with favorable demographic profiles, particularly those with greater purchasing power and stronger market engagement, tend to exhibit higher purchase intentions (Cai, 2023). Consequently, customer demographics are expected to positively affect purchasing intention.

H2: Customer demographics positively influence purchasing intention.

Product Preference and Purchasing Behavior

Product preference reflects consumers' evaluations and interests regarding specific product categories, attributes, and brands. According to consumer choice theory, individuals are more likely to engage in purchasing activities when products align with their preferences, perceived needs, and personal values. Strong product preferences often lead to increased purchase frequency, higher expenditure levels, and greater engagement with preferred product categories (Wu, 2025). As a result, consumers with stronger product preferences are expected to demonstrate more active purchasing behavior.

H3: Product preference positively influences purchasing behavior.

Product Preference and Purchasing Intention

Product preference is also considered a key antecedent of purchasing intention because consumers tend to develop stronger intentions toward products that satisfy their needs and expectations. Existing studies have consistently reported that preference-based evaluations significantly influence consumers' willingness to purchase products in both traditional and digital retail environments. Consumers who perceive products as desirable and compatible with their preferences are more likely to express stronger future purchasing intentions.

H4: Product preference positively influences purchasing intention.

Purchasing Behavior and Purchasing Intention

Purchasing behavior represents consumers' actual purchasing experiences and transactional activities. Previous behavioral experiences play a crucial role in shaping future intentions, as positive purchase outcomes reinforce favorable attitudes toward future purchases. Consumers who frequently purchase and derive satisfaction from prior transactions are more likely to exhibit stronger purchase intentions (Hafidzah, 2025). Therefore, purchasing behavior is expected to be a significant predictor of purchasing intention.

H5: Purchasing behavior positively influences purchasing intention.

Customer Segmentation and Purchasing Intention

Customer segmentation categorizes consumers into homogeneous groups based on shared demographic and behavioral characteristics. Different customer segments often exhibit varying levels of purchasing power, product involvement, loyalty, and purchase propensity. High-value and premium customer segments are generally characterized by stronger customer engagement and higher repurchase rates than lower-value segments (Yulisasih *et* *al*., 2024). Consequently, customer segmentation is expected to significantly influence purchasing intention.

H6: Customer segmentation positively influences purchasing intention.

Mediating Role of Purchasing Behavior

The relationship between product preference and purchasing intention may not be entirely direct. Consumers with strong preferences for particular products are more likely to make actual purchases, which in turn strengthens their future purchase intentions. In this context, purchasing behavior serves as a mechanism by which product preferences translate into future purchase intentions (Chen, 2024). Therefore, purchasing behavior is expected to mediate the relationship between product preference and purchasing intention.

H7: Purchasing Behavior positively influences Customer Segmentation.

Mediating Role of Customer Segmentation

Customer segmentation reflects behavioral heterogeneity among consumers and captures differences in purchasing characteristics across customer groups. Purchasing behavior contributes to the formation of customer segments, while membership in these segments subsequently influences purchasing intentions through distinct behavioral and spending patterns. As a result, customer segmentation may serve as an intervening mechanism explaining how purchasing behavior affects future purchasing intentions (Fuentes *et al*., 2018). Therefore, customer segmentation is expected to mediate the relationship between purchasing behavior and purchasing intention.

H8: Customer segmentation mediates the relationship between purchasing behavior and purchasing intention.

Materials and Methods

This study adopts a quantitative, data-driven research design that integrates customer segmentation with Structural Equation Modeling (SEM) to investigate customer purchase behavior and purchasing intention. Customer data is first preprocessed through missing-value treatment, outlier detection, and normalization to ensure data quality. K-Means clustering is then applied to identify homogeneous customer segments based on purchasing characteristics and behavioral patterns. The resulting segments are incorporated into the SEM framework to examine the relationships among Customer Demographics (CD), Product Preference (PP), Purchasing Behavior (PB), Customer Segmentation (CS), and Purchasing Intention (PI). This hybrid design enables the simultaneous analysis of customer heterogeneity and causal relationships, providing comprehensive insights for data-driven marketing decision-making.

Dataset Description

This study utilized the E-commerce behavior dataset obtained from the Kaggle repository, which simulates realistic customer behavior in an e-commerce environment (https://www.kaggle.com/datasets/asifxzaman/e-commerce-behavior-dataset8000-users). The dataset comprises behavioral, demographic, engagement, and transactional information from 8,000 users and is widely used for customer behavior analysis, purchase prediction, and customer segmentation. It includes key variables such as User ID, Age, Gender, Browsing Time, Product Category, Number of Pages Visited, Engagement Metrics, Purchase Frequency, Purchase Amount, and Purchase Decision Indicators. These comprehensive customer attributes provide a suitable foundation for investigating purchasing behavior and purchasing intention using the proposed hybrid customer segmentation and Structural Equation Modeling framework.

Data Preprocessing

Data preprocessing was performed to improve data quality and ensure reliable analytical results. Missing values were treated using appropriate imputation techniques, duplicate records were removed, and outliers were detected using the Interquartile Range (IQR) method. Categorical variables were converted to numerical values via label encoding, and numerical features were normalized using Min–Max scaling to standardize variable ranges and improve computational efficiency (Kapti & Astuti, 2025). These preprocessing procedures produced a clean, consistent, and standardized dataset, providing a robust foundation for exploratory data analysis, customer segmentation, and Structural Equation Modeling.

Exploratory Data Analysis

Exploratory Data Analysis (EDA) was conducted to examine customer characteristics, purchasing behavior, and relationships among the study variables. The analysis included demographic distributions, purchasing frequency, product preferences, and purchase amounts. Descriptive statistics were computed to summarize the data's distribution and variability, and correlation analysis was performed to identify associations between customer attributes and purchasing indicators (Kanavos *et al*., 2025). The EDA results provided valuable insights into customer behavior and served as the basis for subsequent customer segmentation and Structural Equation Modeling.

Customer Segmentation Using K-Means Clustering

Customer segmentation was performed using the K-Means clustering algorithm to identify homogeneous customer groups based on purchasing characteristics. The clustering analysis utilized five numerical variables: Age, Annual Income, Spending Score, Purchase Frequency, and Purchase Amount. Before clustering, all variables were normalized using Min–Max scaling to ensure equal contribution to the Euclidean distance calculation. The K-Means algorithm was implemented with the following parameter settings: the number of clusters ((K)) was determined using the Elbow Method and Silhouette Analysis, resulting in an optimal value of (K = 4); Euclidean distance was used as the distance metric; k-means++ was adopted for centroid initialization; the maximum number of iterations was set to 300; 10 independent initializations (n_init = 10) were performed to improve clustering stability; and the convergence tolerance was fixed at 1×10^{-4}. Based on the clustering results, customers were classified into four behavioral segments with distinct spending and purchasing patterns. These segments were subsequently incorporated into the Structural Equation Modeling (SEM) framework as the Customer Segmentation (CS) construct to examine their influence on purchasing behavior and purchasing intention.

Results and Discussion

Descriptive Statistics

Descriptive statistics were computed to provide an overview of the respondents' demographic and behavioral characteristics. **Table 1** summarizes the means, standard deviations, minimums, and maximums of the variables included in the study. The results indicate substantial variability among customers in annual income, spending score, purchase frequency, and purchase amount, suggesting heterogeneous customer groups suitable for segmentation analysis.

Table 1. Descriptive Statistics of Study Variables

Variable	Mean	SD	Min	Max
Age	42.60	13.80	18	75
Annual Income (USD)	68,450	28,610	15,000	150,000
Spending Score	52.70	21.40	1	100
Purchase Frequency	11.80	5.90	1	30
Purchase Amount (USD)	845.60	412.80	45	2,650

Common Method Bias

Common method bias (CMB) was assessed to examine whether the observed relationships among the latent constructs were affected by measurement-method bias. The first factor explained 32.84% of the total variance, which is below the recommended threshold of 50%, indicating that common method variance was not a significant concern. Additionally, the variance inflation factor (VIF) values ranged from 2.11 to 2.72, well below the threshold of 3.30, further confirming the absence of serious common method bias. These findings indicate that the estimated relationships among the study constructs are unlikely to be influenced by measurement bias.

Measurement Model

The measurement model was evaluated using Confirmatory Factor Analysis (CFA) to assess construct reliability, internal consistency, and convergent validity. The results demonstrated satisfactory measurement quality across all five constructs: Customer Demographics, Product Preference, Purchasing Behavior, Customer Segmentation, and

Purchasing Intention. Cronbach's alpha values ranged from 0.842 to 0.912, exceeding the recommended threshold of 0.70 and confirming adequate internal consistency. Composite Reliability values ranged from 0.881 to 0.932, further supporting construct reliability. Average Variance Extracted values ranged from 0.652 to 0.774, surpassing the minimum criterion of 0.50 and indicating acceptable convergent validity. These findings confirm that the measurement model exhibits strong psychometric properties and provides a reliable basis for subsequent structural model estimation and hypothesis testing within the proposed research framework.

Goodness-of-Fit Indices of the Proposed SEM Model

Following measurement model validation, the overall fit of the proposed Structural Equation Model (SEM) was evaluated using absolute, incremental, and parsimonious fit indices, including χ^2/df, CFI, TLI, GFI, AGFI, NFI, RMSEA, and SRMR. As presented in **Table 2**, the goodness-of-fit assessment confirms that the proposed structural equation model adequately represents the observed data. The χ^2/df value of 2.41 is below the recommended threshold of 3.00, indicating good model fit. The incremental fit indices also demonstrate strong performance, with CFI = 0.953, TLI = 0.947, and NFI = 0.942, all exceeding the recommended minimum of 0.90. Similarly, GFI = 0.936 and AGFI = 0.921 satisfy their respective criteria, further supporting model adequacy. In addition, RMSEA = 0.051 and SRMR = 0.043 are well below the maximum acceptable value of 0.08. Overall, the results in **Table 2** provide strong evidence that the proposed SEM achieves satisfactory goodness of fit and is appropriate for subsequent structural path analysis and hypothesis testing.

Table 2. Goodness-of-fit assessment of the proposed structural equation model

Fit Index	Recommended Threshold	Obtained Value	Interpretation
χ^2/df	< 3.00	2.41	Good Fit
CFI	≥ 0.90	0.953	Excellent Fit
TLI	≥ 0.90	0.947	Excellent Fit
GFI	≥ 0.90	0.936	Good Fit
AGFI	≥ 0.90	0.921	Good Fit
NFI	≥ 0.90	0.942	Excellent Fit
RMSEA	≤ 0.08	0.051	Excellent Fit
SRMR	≤ 0.08	0.043	Excellent Fit

Measurement Model Assessment

The measurement model was evaluated to assess the reliability and validity of the latent constructs before testing the structural relationships. Reliability was examined using Cronbach's Alpha (CA) and Composite Reliability (CR), while convergent validity was assessed using Average Variance Extracted (AVE). As shown in **Table 3**, all constructs exceeded the recommended thresholds (CA and CR > 0.70; AVE > 0.50), confirming satisfactory internal consistency and convergent validity. Composite Reliability ranged from 0.881 to 0.932, and AVE ranged from 0.652 to 0.774. Discriminant validity was evaluated using the Fornell–Larcker criterion, with all square roots of AVE exceeding the corresponding inter-construct correlations **(Table 4)**. These results demonstrate that the measurement model possesses adequate reliability and validity for subsequent structural model analysis.

Table 3. Reliability and convergent validity assessment.

Construct	Cronbach's Alpha	Composite Reliability (CR)	AVE
Customer Demographics (CD)	0.842	0.881	0.652
Product Preference (PP)	0.865	0.893	0.676
Purchasing Behavior (PB)	0.891	0.916	0.731
Customer Segmentation (CS)	0.854	0.887	0.664
Purchasing Intention (PI)	0.912	0.932	0.774

Table 4. Fornell–Larcker discriminant validity assessment.

Construct	CD	PP	PB	CS	PI
CD	**0.807**				
PP	0.541	**0.822**			
PB	0.584	0.648	**0.855**		
CS	0.427	0.513	0.672	**0.815**	
PI	0.561	0.614	0.738	0.593	**0.880**

Structural Model and Hypothesis Testing Results

Following measurement model validation, the structural model was evaluated using Maximum Likelihood Estimation (MLE) to test the proposed hypotheses. As presented in **Figure 2**, all hypothesized relationships were statistically significant ($p < 0.001$), supporting the proposed research framework. Product Preference significantly influenced Purchasing Behavior ($\beta = 0.451$), and Customer Demographics also had a positive effect ($\beta = 0.324$). Purchasing Behavior was the strongest predictor of Purchasing Intention ($\beta = 0.536$), followed by Customer Segmentation ($\beta = 0.295$), Product Preference ($\beta = 0.268$), and Customer Demographics ($\beta = 0.187$). In addition, Purchasing Behavior significantly influenced Customer Segmentation ($\beta = 0.617$), indicating that behavioral characteristics play a key role in forming distinct customer groups. These findings provide strong empirical support for the proposed hybrid customer segmentation and SEM framework.

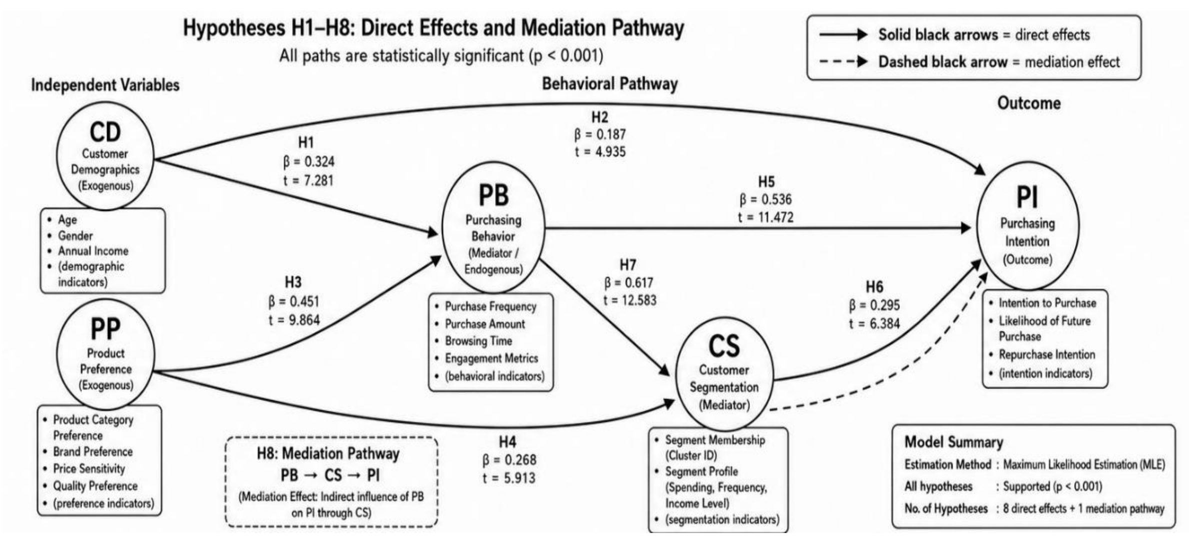

Figure 2. Proposed hybrid customer segmentation and Structural Equation Modeling (SEM) framework with standardized path coefficients and hypothesis testing results.

To further examine the mediating role of Customer Segmentation (CS), a bootstrap analysis with 5,000 resamples confirmed that Customer Segmentation significantly mediates the relationship between Purchasing Behavior and Purchasing Intention ($\beta = 0.182$, $p < 0.001$), supporting H8. The structural model also demonstrated satisfactory explanatory power, accounting for 48.6% of the variance in Purchasing Behavior ($R^2 = 0.486$), 38.1% in Customer Segmentation ($R^2 = 0.381$), and 67.4% in Purchasing Intention ($R^2 = 0.674$). These findings indicate that the proposed hybrid SEM framework effectively explains customer purchasing behavior and future purchasing intentions.

Customer Segmentation Validation Results

The customer segmentation analysis identified four distinct, behaviorally meaningful customer groups, with the optimal number of clusters determined to be four using the Elbow Method. The clustering solution demonstrated strong validation performance, including a within-cluster sum of squares of 1,245.32, a Silhouette Coefficient of 0.684, a Davies–Bouldin Index of 0.593, and a Calinski–Harabasz Index of 2,874.61, indicating good cohesion, excellent separation, and a strong overall cluster structure. Among the identified segments, Regular Customers accounted for the largest share at 31.2%, followed by Budget Customers at 28.6%, Premium Customers at 21.5%, and High-Value Customers at 18.7%. Clear behavioral differences were observed across the segments. Budget Customers had the lowest average income, spending score, and purchase frequency, whereas High-Value Customers recorded the highest levels at USD 118,240, 91.8, and 22.6, respectively. These results confirm that the clustering approach effectively captures meaningful customer heterogeneity.

Conclusion

This study concludes that integrating K-Means customer segmentation with Structural Equation Modeling provides a comprehensive approach for understanding customer purchase behavior and purchasing intention in digital commerce. The analysis of 8,000 customer records identified four distinct segments demonstrating meaningful heterogeneity in income, spending scores, and purchase frequency. The structural model further confirmed that Product Preference significantly influences Purchasing Behavior, while Purchasing Behavior is the strongest predictor of Purchasing Intention. Customer Segmentation also contributes directly to Purchasing Intention and significantly mediates the relationship between Purchasing Behavior and Purchasing Intention. These findings show that combining machine-learning-based segmentation with explanatory structural modeling can yield deeper insights than applying either method in isolation. The framework therefore offers practical value for designing personalized promotions, loyalty programs, customer retention strategies, and more efficient allocation of marketing resources. Overall, the study demonstrates that a hybrid segmentation–SEM approach can strengthen data-driven marketing decision-making while providing a replicable analytical framework for research on customer behavior in retail and e-commerce environments. It also bridges predictive analytics and theory-based explanation, enhancing managerial relevance and interpretation of behavior.

Implications and Contributions

This study contributes to customer behavior research by integrating K-Means clustering with Structural Equation Modeling (SEM) into a unified framework. The proposed approach explains how customer demographics, product preferences, purchasing behavior, and customer segmentation influence purchasing intention. Practically, the findings support personalized marketing, customer segmentation, and data-driven decision-making in retail and e-commerce.

Practical Implications

The findings provide practical guidance for marketers to develop targeted strategies based on customer segments. Personalized promotions, loyalty programs, and tailored recommendations can improve customer engagement and purchasing intention. Overall, the proposed framework supports customer retention, optimized marketing investment, and data-driven decision-making in retail and e-commerce.

Theoretical Contributions

This study contributes to the customer behavior literature by integrating K-Means clustering with SEM into a unified framework. The suggested model illustrates the mediating function of customer segmentation between purchasing behavior and purchasing intention while concurrently capturing consumer heterogeneity and causal linkages. These findings enhance the theoretical understanding of consumer decision-making in retail and e-commerce.

Methodological Contribution

This study contributes methodologically by integrating K-Means clustering and SEM into a unified analytical framework. The proposed approach simultaneously examines customer heterogeneity and causal relationships, providing both predictive and explanatory insights. It also offers a systematic, replicable methodology for analyzing customer behavior in marketing analytics and business intelligence.

Limitations and Future Research Directions

Despite its contributions, this study has several limitations. The analysis was based on a single customer dataset, which may limit the generalizability of the findings. In addition, K-Means clustering requires a predefined number of clusters and may not fully capture customer heterogeneity. Future research should validate the proposed framework using cross-industry datasets, investigate advanced clustering and artificial intelligence techniques, and incorporate additional constructs such as customer satisfaction, loyalty, trust, and perceived value to improve its explanatory and predictive performance.

Acknowledgments: The authors gratefully acknowledge the creators of the E-Commerce Behavior Dataset (8,000 Users) for making the data publicly available. The authors also thank their affiliated institution for its support and the anonymous reviewers for their constructive comments, which contributed to improving the manuscript.

Conflict of Interest: None

Financial Support: This research project was financially supported by Mahasarakham Business School, Mahasarakham University, Thailand.

Ethics Statement: Ethical approval from an institutional review board was not required because the study involved only secondary analysis of a publicly accessible dataset and did not involve the direct recruitment of human participants, experimental interventions, or the collection of identifiable private information.

References

Abreu-Ledón, R., Luján-García, D., Garrido-Vega, P., Machuca, J. A., & Pentón, Y. B. (2026). Evaluating the use of structural equation modeling in online purchase intention research: A comprehensive review (2000–2023). *Technological Forecasting and Social Change*, *225*, 124512. doi:10.1016/j.techfore.2025.124512

Ali, O. H., Mohammed, N., & Akram, H. M. (2025). Integrating Brushing Frequency and Clinical Parameters for Periodontal Disease Prediction Using Machine Learning. *Annals of Dental Specialty*, *13*(4), 108-119. doi:10.51847/KyU4raNSvk

Anora, A., Muchayatin, M., & Ginting, Y. M. (2025). The influence of e-commerce trust, site loyalty, and satisfaction on digital consumer purchase behavior. *Indonesian Journal of Economic Review*, *5*(1), 43–54. doi:10.59431/ijer.v5i1.512

Anwar, D., Faizanuddin, Md., Rahman, F., & Dayal, R. (2025). Analyzing consumer behavior in e-commerce: Insights from data-driven approaches. *Management, 3*, 127. doi:10.62486/agma2025127

Balaji, S., Guleria, N., & Ramasubramanian, C. (2025). Understanding key consumer purchase behaviour to interpret consumer buying habits. *International Journal of Scientific Research in Science and Technology, 12*(5), 49–54. doi:10.32628/ijsrst2513213

Cai, J. (2023). Research on the influencing factors of consumer buying behavior. *Highlights in Science, Engineering and Technology*, *61*, 119–127. doi:10.54097/hset.v61i.10281

Chea, A. C. (2024). Market segmentation strategies: Analysis, practice, and marketing implications. *The Journal of Applied Business and Economics*, *26*(4). doi:10.33423/jabe.v26i4.7177

Chen, S. (2024). The impact of brand awareness on purchase intention. *Highlights in Business, Economics and Management, 30,* 239–246. doi:10.54097/88gfk880

Dominique, S., Hanafi, M., Llobell, F., Ferrandi, J., & Cariou, V. (2023). Comparing local vs global clustering with FIMIX-PLS: Marketing application. *Springer Proceedings in Business and Economics*, 15–21. doi:10.1007/978-3-031-34589-0_3

Ferdinand, V., & Wijaya, S. F. (2025). Integrating SmartPLS-based structural modeling for predictive machine learning in retail customer behavior. In *Proceedings of the International Conference on Information Management and Computer Information Systems* (pp. 1879–1885). doi:10.1109/ICIMCIS68501.2025.11327115

Fuentes, I., Nápoles, G., Arco, L., & Vanhoof, K. (2018). Customer segmentation using multiple instance clustering and purchasing behaviors. In *Lecture Notes in Computer Science* (pp. 193–200). doi:10.1007/978-3-030-01132-1_22

Ganesha, K. S., & Hegde, S. S. (2025). Examining purchase and usage behavior towards arecanut products using theory of planned behavior. In *Advances in Geospatial Technologies Book Series* (pp. 31–44). doi:10.4018/979-8-3373-3206-2.ch002

Hafidzah, F. N. (2025). Attitude, subjective norms, and environmentally friendly as antecedents of customer purchase behavior. *Journal of Management and Business Insight, 2*(2), 111–118. doi:10.12928/jombi.v2i2.1149

Hamza, A. A.-A. (2023). Factors impacting online purchase: A review of the literature. *Frontline Social Sciences and History Journal, 3*(4), 1–4. doi:10.37547/social-fsshj-03-04-01

Kanavos, A., Vonitsanos, G., & Mylonas, P. (2025). Integrating machine learning approaches for consumer behavior analysis in retail transactions. In *Proceedings of the International Workshop on Semantic and Social Media Adaptation and Personalization* (pp. 7–12). doi:10.1109/SMAP66932.2025.00010

Kapti, K., & Astuti, D. (2025). Implementation of the K-means clustering algorithm for customer segmentation. *Jurnal Sistem Informasi, 14*(6), 2751. doi:10.32520/stmsi.v14i6.5573

Mahpour, A., Mollajani, M., Baghestani, A., & Zamani, P. (2025). From interest to action: Analyzing the mediating and moderating factors in online shopping behavior using a structural equation modeling approach. *Telematics and Informatics Reports, 20,* 100261. doi:10.1016/j.teler.2025.100261

Modepalli, U. P. (2025). Enhancing marketing ROI through predictive customer segmentation using behavioral analytics. *Journal of Computer Science and Technology Studies, 7*(8), 1035–1042. doi:10.32996/jcsts.2025.7.8.118

Octavia, C. D., & Handayani, W. P. P. (2025). Purchase intention produk ramah lingkungan: Pendekatan theory of planned behavior. *Journal of Management and Digital Business, 5*(2), 369–384. doi:10.53088/jmdb.v5i2.1367

Rahman, W., Shah, F. A., & Rasli, A. (2015). Use of structural equation modeling in social science research. *Asian Social Science, 11*(4), 371–377. doi:10.5539/ASS.V11N4P371

Richter, N. F., & Tudoran, A. A. (2024). Elevating theoretical insight and predictive accuracy in business research: Combining PLS-SEM and selected machine learning algorithms. *Journal of Business Research, 173*, 114453. doi:10.1016/j.jbusres.2023.114453

Sangami, P. (2025). Customer segmentation and personalization in marketing. *ComFin Research, 13*(S1-i2), 19–21. doi:10.34293/commerce.v13is1-i2.8726

Singh, S., Kaushal, D., & Singh, M. (2025). The utilization of machine learning techniques in the prediction and manipulation of consumer purchase intentions. In *Artificial Intelligence and Data Analytics* (pp. 231–246). doi:10.1201/9781998511235-14

Son, L. N. (2026). AI Decision Support System Adoption in Dental Education: An Empirical PLS-SEM Study. *Annals of Dental Specialty, 14*(3), 12-18. doi:10.51847/SAEgs2aXxf

Son, L. N., & Trung, N. T. (2024). Strategic Environmental Management, Green Dynamic Capabilities, and Ecosystem Sustainability: A PLS-SEM Study in Emerging Economies. World Journal of Environmental Biosciences, 13(3), 51-57. doi:10.51847/BbALGM9mH3

Sravani, D., Balakirshna, K., & Poojasrinivas, G. (2025). Decoding consumer decisions: A data-driven approach using statistical techniques. *International Journal of Engineering Science and Advanced Technology, 25*(9), 549–556. doi:10.64771/ijesat.2025.v25.i09.pp549-556

Sutantio, A. C., Safitri, Y. I., Annizari, D. F., Kurniawan, S., & Pramayoga, A. S. (2025). Unveiling consumer behaviour insights via K-means customer segmentation. In *Proceedings of the International Conference on Informatics and Computing* (pp. 400–404). doi:10.1109/ICICOS68590.2025.11329784

Thakur, Y., & Mittal, N. (2025). Customer segmentation using K-means clustering. In *Advances in Intelligent Systems and Computing* (pp. 375–382). doi:10.1007/978-981-96-7214-1_32

Thanh, L. C., & Trung, N. T. (2026). Evaluating the Impact of Service Quality on Patient Satisfaction in Dental Care Centers. *Annals of Dental Specialty, 14*(2), 37-44. doi:10.51847/P0dK1vQxWJ

Umar, M., Singh, K., Dayma, A., & Gupta, A. (2025). Factors influencing consumer buying behaviour: A conceptual overview. *International Journal of Engineering Applied Science and Technology*, *10*(3), 57–62. doi:10.33564/ijeast.2025.v10i03.011

Umoren, O., Didi, P. U., Balogun, O. S., Abass, O. S., & Akinrinoye, O. V. (2025). A predictive and segmentation-based marketing analytics framework for optimizing customer acquisition, engagement, and retention strategies. *Engineering and Technology Journal*, *10*(9). doi:10.47191/etj/v10i09.06

Wathanakom, N. (2025). Structural equation modelling in marketing: A systematic review of methods and models. *European Conference on Research Methodology for Business and Management Studies*, *24*(1), 139–145. doi:10.34190/ECRM.24.1.3635

Wu, Y. (2025). Research on the influence of preference change on market demand from the perspective of consumer behavior theory. *Highlights in Business, Economics and Management, 60*, 484–491. doi:10.54097/00237z22

Yulisasih, B. N., Herman, H., & Sunardi, S. (2024). K-means clustering method for customer segmentation based on potential purchases. *Jurnal ELTIKOM: Jurnal Teknik Elektro, Teknologi Informasi dan Komputer, 8*(1), 83–90. doi:10.31961/eltikom.v8i1.1137