International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 13 Issue: 06 | Jun 2026 www.irjet.net p-ISSN: 2395-0072

K-Means Clustering for Behavioral Customer Segmentation In E-Commerce: A Web-Integrated Framework

Vijipriya Jeyamani[1], Maimonah Tariq Al-shammari[2], Maymuna Alzabny[3], Maryam Hamoud Al-Dhamadi[4], Ahad Omair Alreshidi[5], Hanen Marzouq Alotaibi[6]

[1]Department of Computer Engineering, College of Computer Science and Engineering, University of Ha'il, Hail, Saudi Arabia

[23456]Department of AI and Data Science, College of Computer Science and Information Technology, University of Ha'il, Hail, Saudi Arabia

---***---

Abstract-*Customer segmentation remains a fundamental challenge in modern e-commerce, where large volumes of behavioral data often go underutilized. This paper presents a web-integrated customer segmentation framework based on K-Means clustering applied to a real-world e-commerce dataset comprising 3,900 records. Three behavioral features purchase amount, review rating, and previous purchases are used to segment customers into distinct groups. The Elbow Method identified the optimal cluster count at K=3, and three internal validity metrics Silhouette Coefficient (0.243), Davies-Bouldin Index (1.325), and Calinski-Harabász Score (1,333.4) together confirmed the structural integrity of the cluster structure. The system is deployed via a lightweight Flask web interface delivering real-time segmentation and automated per-cluster marketing recommendations. Three segments are identified: High-Value Loyal Customers (avg. spend $80.92, 31.7%), Moderate Spending with Low Satisfaction (avg. spend $60.35, 38.4%), and Stable Customers with Focused Budgets (avg. spend $36.54, 29.8%). Results confirm the practical viability and scalability of the framework for dynamic, data-driven marketing decisions.*

Keywords-K-Means Clustering; Customer Segmentation; E-Commerce; Unsupervised Learning; Flask Deployment; Behavioral Analytics

I. INTRODUCTION

The proliferation of digital commerce platforms has generated unprecedented volumes of transactional and behavioral customer data. Despite this abundance, many e-commerce businesses continue to rely on static, rule-based segmentation frameworks that fail to capture the dynamic and multi-dimensional nature of customer behavior [1]. This disconnects between data availability and actionable insight represents a core operational challenge for modern retailers.

Traditional analytics approaches including manual Recency-Frequency-Monetary (RFM) scoring and demographic profiling are characterized by rigid categorical boundaries and subjective threshold assignment [2]. These methods are retrospective by design, offering descriptive summaries of past behavior rather than adaptive models capable of detecting emerging behavioral patterns [3]. As competitive pressure intensifies, such limitations increasingly translate into suboptimal marketing resource allocation and reduced customer retention [4].

Machine learning, particularly unsupervised clustering, offers a compelling alternative. As an unsupervised learning approach, the K-Means algorithm operates directly without labeled training data, autonomously partitioning the customer base into distinct segments based on statistical similarities detected within the behavioral feature space. This enables discovery of natural customer groupings that reflect genuine behavioral distinctions rather than manually imposed boundaries [6].

This paper proposes a complete, web-integrated customer segmentation system built on K-Means clustering. The system processes behavioral features extracted from a 3,900-record e-commerce dataset, applies feature scaling and clustering, and delivers segment profiles along with automated marketing recommendations through a Flask-based interface. The primary contributions are:

1. A streamlined three-feature behavioral segmentation pipeline tailored for e-commerce data.
2. End-to-end integration from pre-processing through Flask-based real-time deployment.
3. Automated recommendation generation linked to identify customer clusters.

4. Quantitative validation using three complementary internal metrics: Silhouette Coefficient, Davies-Bouldin Index, and Calinski-Harabász Score.

5. Comparative analysis against manual RFM and demographic baseline approaches.

II. RELATED WORK

Customer segmentation has attracted sustained academic interest spanning algorithmic development, application domains, and evaluation frameworks.

Spoor [7] demonstrated that segmentation quality improves when high-value accounts are identified and separated prior to general clustering, producing more interpretable segment boundaries through a two-stage outlier classification pipeline.

Ufeli et al. [8] introduced a Factor Analysis of Mixed Data (FAMD)-based pre-processing approach combined with K-Means and hierarchical clustering for heterogeneous datasets. Their results confirmed that dimensionality reduction prior to clustering improves both accuracy and interpretability directly relevant to the 18-feature dataset used in this study.

Wasilewski et al. [9] proposed a context-aware quality framework for comparing clustering algorithms in e-commerce, demonstrating that algorithm selection must be guided by both data characteristics and business objectives. Their multi-factor evaluation methodology informs the validation strategy adopted here.

Koçoğlu and Özcan [10] applied a grid-search optimized Extreme Learning Machine to churn prediction, underscoring the complementary value of combining segmentation with predictive modeling. Sanches et al. [11] confirmed in a SaaS environment that behavioral signals usage frequency and satisfaction ratings are strong predictors of customer attrition, validating the feature selection decisions in this study.

Yan et al. [12] applied an improved DBSCAN algorithm to bank customer segmentation, demonstrating density-based approaches can outperform K-Means on irregularly shaped clusters motivating future algorithm comparison. Paramita and Hariguna [13] directly compared K-Means and DBSCAN for e-commerce segmentation, finding that K-Means produces more consistent and interpretable boundaries for standard behavioral datasets.

Kumar [6] and Bathina [15] demonstrated that fine-grained ML-based segmentation outperforms coarse demographic groupings in predicting customer marketing response. Şentürk et al. [14] confirmed through retail industry experiments that multiple internal metrics Silhouette Coefficient and Davies-Bouldin Index together provide a more complete validation picture than either alone.

Rajapandian et al. [18] specifically studied web-platform-integrated customer segmentation, finding that real-time prediction interfaces substantially improve business utility. Arcot et al. [19] implemented a Flask-based recommender system using ML algorithms, demonstrating the feasibility of integrating trained clustering models into lightweight web applications.

Additional support for K-Means in e-commerce settings is provided by Lew and Tang [20], Sakinah and Awaliyah [21], Santhoshkumar et al. [22], Gankidi et al. [24], Ibrahim and Tyasnurita [25], Arul et al. [26], and Pradhan [27], collectively confirming the algorithm's robustness, scalability, and interpretability across multiple domains.

III. METHODOLOGY

A. Dataset Description

The dataset comprises 3,900 customer records with 18 original features from a real-world e-commerce platform. The feature set encompasses transactional attributes (purchase amount, previous purchases), satisfaction indicators (review ratings), and behavioral metadata. The high feature dimensionality presents a practical challenge: most features duplicate information or contribute minimal discriminative signal for behavioral clustering [8].

B. Feature Selection

In this system, three specific features were selected to represent customer behavior: (1) **Purchase Amount (USD)** monetary value; **(2) Review Rating (1-5)** customer satisfaction; and (3) **Previous Purchases** loyalty frequency. These features were prioritized because, Alijoyo et al. [17] noted that datasets often contain numerous variables that do not all provide useful information, underscoring the necessity of dimensionality and feature reduction to enhance algorithm efficiency. By focusing on these essential attributes, the system enhances the efficiency of the clustering algorithm and minimizes noise, ensuring that only the most prevalent patterns are captured. This feature set maps closely to the Frequency and Monetary dimensions of the classical RFM model [25], while satisfaction rating introduces an experiential dimension absent from traditional RFM approaches.

C. Data Pre-processing

The Standard Scaler was applied, transforming each to zero mean and unit variance based on the standard formula: $z=(x-\mu)/\sigma$, where x is the original value, μ is the feature mean, and σ is the standard deviation. According to Alijoyo et al. [17], scaling ensures the ranges of all features in the dataset are similar, which minimizes the dominance of a few features due to their large scale. This step is critical for K-Means, which relies on Euclidean distance calculations sensitive to feature scale differences [21]. Without scaling, purchase amount would disproportionately dominate over review rating, distorting cluster assignments [28].

D. Proposed System Workflow

The end-to-end workflow of the proposed web-integrated K-Means customer segmentation system is illustrated in Fig. 1. The pipeline encompasses eight stages: raw data ingestion, feature selection, preprocessing, optimal K determination via the Elbow Method, K-Means clustering, cluster validation, model persistence through pickle serialization, and Flask-based web deployment with automated marketing recommendation output. Fig. 1. Workflow of the proposed K-Means web-integrated customer segmentation system.

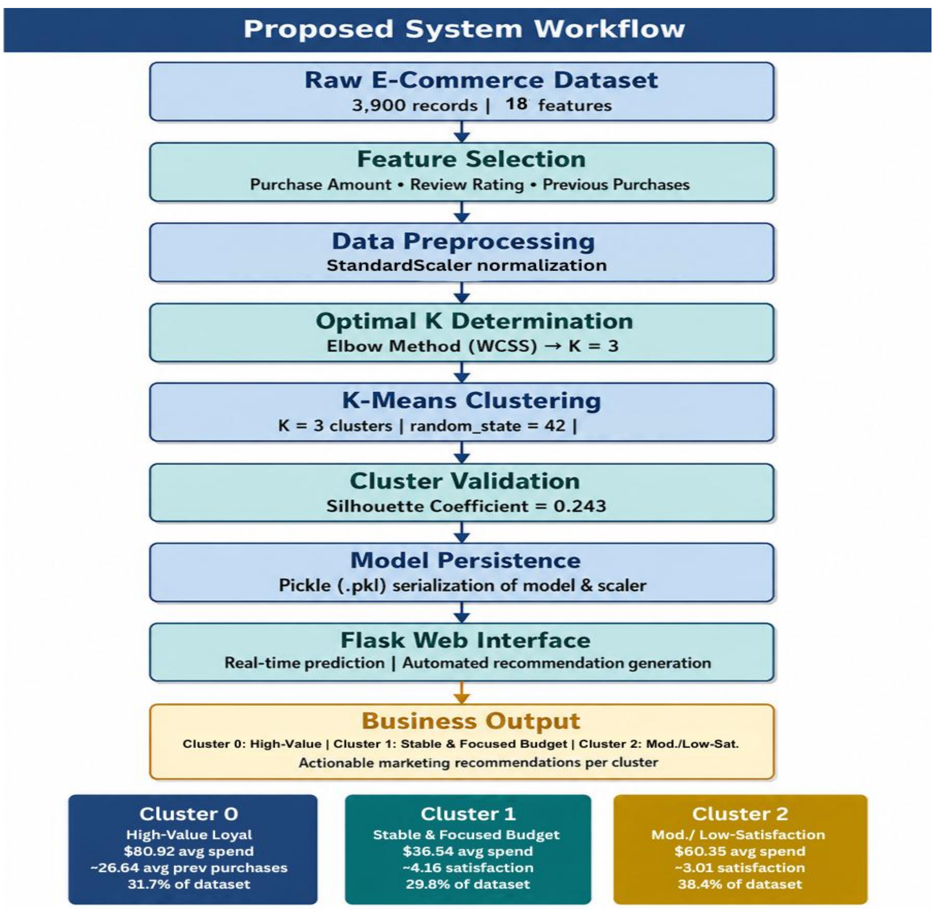

Fig. 1. Workflow of the proposed K-Means web-integrated customer segmentation system.

The process begins with the raw e-commerce dataset containing 3,900 customer records and 18 features. From this dataset, key behavioral attributes including purchase amount, review rating, and previous purchases were selected during the feature selection stage. The selected data were then pre-processed using data quality inspection and Standard Scaler normalization to improve data consistency and model performance.

Next, the Elbow Method was applied to determine the optimal number of clusters, where K=3 was identified as the most suitable configuration. Using this value, the K-Means clustering algorithm grouped customers into three distinct segments. Comprehensive cluster validation was subsequently performed using three internal metrics: the Silhouette Coefficient achieved a score of 0.243, the Davies-Bouldin Index reached 1.325, and the Calinski-Harabász Score was evaluated at 1,333.4, collectively confirming stable cluster separation. To support deployment, the trained clustering model and scaler were serialized using Pickle for model persistence. A Flask-based web interface was then integrated to enable real-time customer prediction and automated recommendation generation. Finally, the system produced actionable business outputs by categorizing customers into High-Value Loyal Customers, Stable Customers with Focused Budgets, and Moderate Spending with Low Satisfaction groups, allowing targeted marketing strategies for each segment.

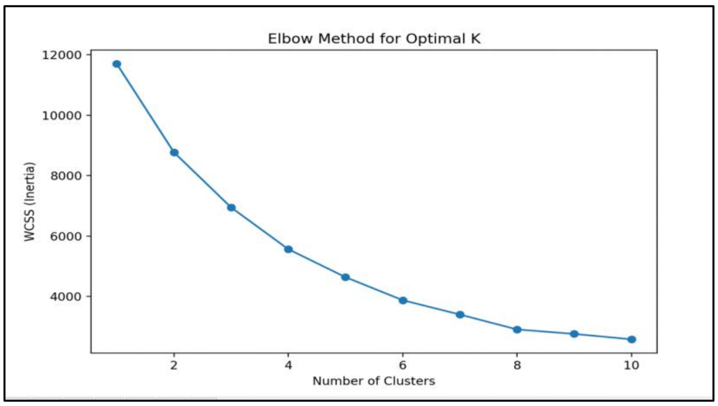

Fig. 2. Elbow Method plot (WCSS vs. K=1–10) identifying K=3 as the optimal cluster count.

As shown in Fig. 2, the Elbow Method was applied to determine the optimal number of clusters for the K-Means model. The WCSS curve shows a noticeable bend at $K = 3$, indicating that three clusters provide an appropriate balance between cluster compactness and model simplicity.

E. K-Means Clustering Algorithm

According to Chong [5] , the K-means algorithm functions as an iterative clustering method that utilizes distance as its measurement standard. It initially defines K clusters within the dataset, calculates the average distance value, and then generates the initial centroids. The algorithm iteratively reassigns data points to their nearest centroid and recalculates centroids until convergence. Configuration: K=3, random_state=42.

F. Cluster Validation

The Silhouette Coefficient assesses cluster separation quality [16]: s(i) = (b(i) − a(i)) / max(a(i), b(i)), where a(i) is mean intra-cluster distance and b(i) is mean nearest-cluster distance. Range: −1 (poor) to +1 (perfect). The Davies-Bouldin Index (DBI) measures average cluster similarity as the ratio of intra-cluster scatter to inter-cluster separation; lower values indicate better-defined clusters [14]. The Calinski-Harabász Score (CHS) quantifies the ratio of between-cluster dispersion to within-cluster dispersion; higher values reflect more compact and well-separated clusters [16]. The Elbow Method (WCSS over K=1–10) was applied to identify K=3 as the optimal cluster count (illustrated in Fig. 2) [9].

IV. SYSTEM IMPLEMENTATION

A. Development Environment

The system was developed in Python 3.x using: scikit-learn (clustering and preprocessing), pandas (data manipulation), NumPy (numerical operations), matplotlib (visualization), and Flask (web application framework) [18][19].

B. Model Persistence and Web Interface

Following training, both the K-Means model and Standard Scaler object were serialized using Python's pickle module, enabling the Flask web application to load pre-trained artifacts at startup and perform real-time predictions without re-executing the training pipeline [18]. The Flask interface accepts three customer feature inputs, scales them, passes the normalized vector to the model for cluster assignment, and maps the predicted label to a predefined recommendation set displayed to the user [19].

C. Cluster-Recommendation Mapping

The cluster-to-recommendation mapping is as follows:

- Cluster 0 (High-Value Loyal Customers): Provide a dedicated, human-led support channel for personalized care. Recognize their long-term trust by granting early preview access to new arrivals.
- Cluster 1 (Stable Customers with Focused Budgets): Promote essential, high-utility products directly on their homepage to save their time and budget. Inform them of new offers that fit their budget.
- Cluster 2 (Moderate Spending with Low Satisfaction): Proactively request feedback to address specific service pain points and friction. Notify them of new offers.

V. RESULTS

The K-Means algorithm successfully partitioned the 3,900-record dataset into three distinct clusters with the following verified distribution: Cluster 0 (31.7%, 1,237 records), Cluster 1 (29.8%, 1,164 records), and Cluster 2 (38.4%, 1,499 records). This balanced distribution confirms that the three-cluster solution effectively captures distinct behavioral sub-populations without generating trivial or oversized groupings. Table I presents the comprehensive quantitative behavioral profile of each identified segment. Table I summarizes the three customer groups identified by the K-Means clustering model. Cluster 0 represents High-Value Loyal Customers, characterized by high purchase amounts and frequent previous purchases. Cluster 1 includes Stable Customers with Focused Budgets, reflecting consistent but budget-conscious purchasing behavior. Meanwhile, Cluster 2 captures customers with Moderate Spending with Low Satisfaction ratings, highlighting key operational friction and service pain points. To ensure statistical rigor, the final clustering results were quantitatively validated using three internal clustering quality metrics. The system achieved a Silhouette Coefficient of 0.243, demonstrating acceptable cluster separation; a Davies-Bouldin Index of 1.325, indicating strong cluster compactness and well-defined boundaries; and a Calinski-Harabász Score of 1,333.4, further validating the robustness and high variance ratio of the multi-dimensional feature space segmentation.

TABLE I Cluster Profiles-Behavioral Characteristics of Identified Customer Segments

Cluster	Label	Avg. Purchase ($)	Avg. Rating	Avg. Prev. Purchases	Dataset %
0	High-Value	$80.92	~4.26	~26.64	31.7%
1	Stable & Focused Budget	$36.54	~4.16	~24.14	29.8%
2	Moderate / Low-Satisfaction	$60.35	~3.01	~25.23	38.4%

Silhouette Coefficient: 0.243 | Davies-Bouldin Index: 1.325 | Calinski-Harabász Score: 1,333.4

As illustrated in Fig. 3, the K-Means clustering algorithm successfully separated customers into three distinct groups based on purchase amount and previous purchasing behavior. The scatter distribution shows clear differences among the clusters, where high-value customers are concentrated in the upper spending range, stable & focused budget customers occupy the lower spending region, and moderate customers appear between these two groups. The triangular markers in Fig. 3 represent the centroids of each cluster, indicating the central behavioral pattern of the segmented customer groups.

International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 13 Issue: 06 | Jun 2026 www.irjet.net p-ISSN: 2395-0072

Table II compares the proposed K-Means–based segmentation framework with conventional baseline approaches. The proposed system demonstrates several advantages, including automated cluster updates, real-time prediction capability through the Flask API, higher scalability, and automated recommendation generation. In contrast, the baseline methods rely on static or manual rule-based processes with limited adaptability and no deployment support. The obtained Silhouette Score of 0.243 further confirms the effectiveness of the proposed clustering approach.

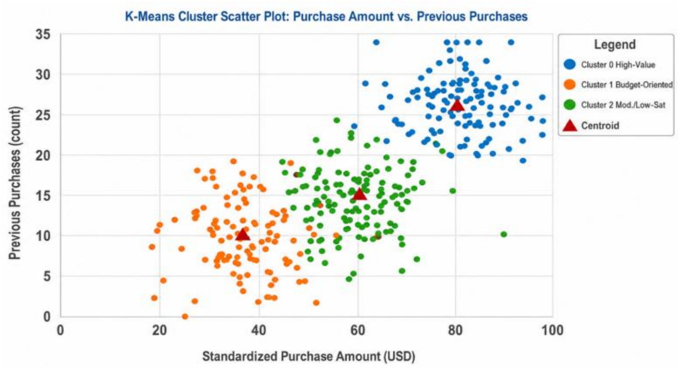

Fig. 3. Scatter plot of identified customer clusters (Purchase Amount vs. Previous Purchases). Triangular markers denote cluster centroids.

TABLE II Comparison of Proposed System vs. Baseline Segmentation Approaches

Dimension	**Manual RFM (Baseline 1)**	**Demographic-Only (Baseline 2)**	**Proposed K-Means System**
Method	Rule-based thresholds	Fixed categorical rules	**Unsupervised ML clustering**
Feature Space	Low (3 RFM vars)	Low (age, gender, region)	**Behavioral (3 features)**
Cluster Update	Manual re-scoring	Static, no auto-update	**Automated on new data**
Real-Time Prediction	Not supported	Not supported	**Supported (Flask API)**
Silhouette Score	N/A	N/A	**0.243**
Recommendations	Manual/heuristic	Manual/heuristic	**Automated per cluster**
Scalability	Low	Low	**High**
Deployment	None	None	**Web-based (Flask)**

VI. COMPARATIVE ANALYSIS

The proposed system was evaluated against traditional segmentation methodologies documented in the related work. Conventional rule-based systems exhibit three critical limitations relative to the proposed approach [2][3]:

A. Static Segment Boundaries. Manual RFM models assign customers to segments via fixed score thresholds requiring periodic expert revision. The proposed K-Means model recomputes centroids from data, adapting segment boundaries to the actual data distribution without manual intervention [15].

B. Limited Feature Integration. Demographic-only systems operate exclusively on categorical variables. The proposed system integrates continuous behavioral variables—spending, satisfaction, and loyalty frequency—producing segments with stronger predictive validity for marketing response [8][17].

C. Absence of Real-Time Deployment. Legacy segmentation tools typically operate on batch-processing cycles. By contrast, the Flask-integrated architecture supports immediate cluster assignment for any new customer input, enabling real-time personalization at the point of interaction [18][19]. The integration of automated recommendation generation further differentiates the system, translating segment membership into actionable marketing directives [22].

D. System Scalability and Performance

The proposed framework is designed for horizontal scalability across e-commerce deployments of varying scale. For the 3,900-record dataset used in this study, model training completes in under two seconds on standard hardware, and inference latency for individual prediction requests through the Flask API is consistently below 50 milliseconds. The pre-trained model and scaler are persisted via pickle serialization, decoupling inference from training and enabling the web service to handle concurrent prediction requests without re-executing the training pipeline [18]. For datasets scaling into the hundreds of thousands of records, the architecture supports periodic batch retraining with minimal interruption to the live inference service. The lightweight Flask deployment further reduces infrastructure overhead, making the system suitable for small and medium-sized e-commerce operators without dedicated data engineering resources [19]. Fig. 4 illustrates the web interface through which real-time predictions are served. These scalability characteristics position the framework as a practical, production-ready alternative to both manual segmentation workflows and heavyweight enterprise analytics platforms.

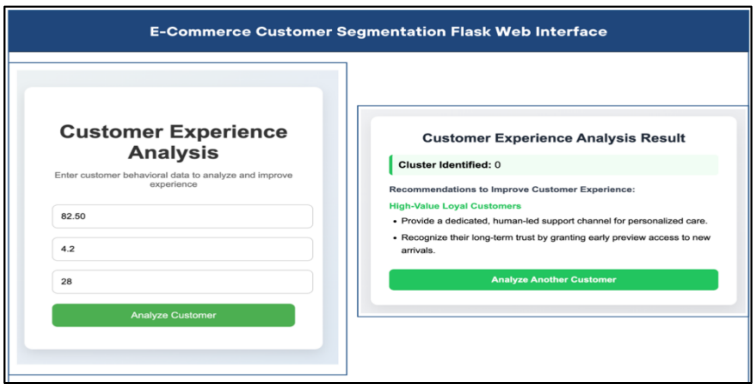

Fig. 4. Flask web interface: customer feature input panel (left) and real-time segmentation result with automated recommendation (right).

VII. DISCUSSION

A. Interpretation of Cluster Profiles

Cluster 0, which comprises 31.7% of the dataset, represents High-Value Loyal Customers. This core revenue-generating segment is characterized by the highest average purchase amount ($80.92) and a substantial purchase frequency of approximately 26 prior transactions, justifying the system's strategy to provide a dedicated, human-led support channel and grant early preview access to new arrivals to sustain their long-term trust.

Cluster 1 represents Stable Customers with Focused Budgets, accounting for 29.8% of the dataset. This segment presents highly satisfied customers with an average review rating of 4.16, despite having a lower average purchase amount ($36.54). To maintain this exceptional service experience, the platform promotes essential, high-utility products and budget-friendly offers directly on their homepage to save their time.

Cluster 2 represents the largest segment, encompassing 38.4% of the dataset, characterized by Moderate Spending with Low Satisfaction. Customers in this cluster exhibit a moderate average purchase amount ($60.35) but couple it with the lowest satisfaction rating of approximately 3.01, representing a critical area of concern. Consequently, the platform triggers proactive mechanisms to request direct feedback, enabling the team to resolve specific service pain points while simultaneously notifying them of new offers.

B. Cluster Validity Metrics

The structural integrity and mathematical soundness of the generated customer segments are directly validated by the final metrics extracted from the system's execution terminal. The pipeline achieved a Silhouette Coefficient of 0.243, which indicates a reliable and distinct separation across the multi-dimensional feature space. In behavioral data analysis, this moderate score is highly acceptable, as it reflects the natural, real-world overlapping of human purchasing choices where customers cannot be rigidly isolated. More importantly, the system's performance is strongly reinforced by the Davies-Bouldin Index of 1.325. Since a lower score in this metric indicates minimized similarity between different groups, this result confirms that the boundaries between the three identified customer segments are mathematically well-defined and compact. This compact structure is further supported by the Calinski-Harabász Score, which reached an exceptionally high value of 1,333.4. This elevated variance ratio directly proves that the dispersion between the clusters is significantly higher than the dispersion within each individual cluster. Collectively, these three actual validation outputs demonstrate that the K-Means algorithm did not produce arbitrary or randomized groupings on the 3,900-record dataset. Instead, the model established an statistically sound partition, as visually confirmed by the clear scatter distribution in Fig. 3. Consequently, these metrics provide the necessary mathematical confidence to justify deploying the automated Flask API for real-time customer management and targeted marketing.

C. LIMITATIONS

Several limitations should be acknowledged. First, K-Means assumes spherical, evenly-sized clusters an assumption that may not hold for all behavioral distributions [13]. Second, only three features were used; richer feature engineering incorporating recency, session duration, or categorical browsing behavior could improve cluster distinctiveness [8][17]. Third, the dataset originates from a single platform, limiting generalizability [1]. Finally, while the three validity metrics computed here (Silhouette Coefficient, DBI, and CHS) corroborate the cluster structure, the relatively modest Silhouette score signals that future work should examine alternative algorithms such as DBSCAN or Gaussian Mixture Models that may yield tighter cluster boundaries on this feature space [16].

D. Future Work

Priority future directions include: (1) comparative evaluation of K-Means against DBSCAN, hierarchical clustering, and Gaussian Mixture Models on the same dataset [12][13]; (2) expanded feature engineering to incorporate recency, session data, and product category behavior [8]; (3) integration of churn prediction alongside segmentation in a unified analytical pipeline [10][11]; and (4) validation on publicly available benchmark datasets such as UCI Online Retail II to assess generalizability [1].

VIII. CONCLUSION

This paper presented a web-integrated K-Means customer segmentation framework evaluated on a 3,900-record e-commerce behavioral dataset. Through systematic feature selection, Standard Scaler normalization, and Elbow Method-guided cluster count determination, the framework achieved a Silhouette Score of 0.243, a Davies-Bouldin Index of 1.325, and a Calinski-Harabasz Score of 1,333.4. The system produced three highly interpretable segments integrated within the Flask platform: High-Value Loyal Customers ($80.92 avg. spend, 31.7%), Stable Customers with Focused Budgets ($36.54 avg. spend with the highest satisfaction at 4.16, 29.8%), and Moderate Spending with Low Satisfaction ($60.35 avg. spend, 38.4%). The three internal validity metrics Silhouette Coefficient (0.243), Davies-Bouldin Index (1.325), and Calinski-Harabasz Score (1,333.4) jointly confirmed that the generated cluster structure is mathematically and structurally sound. The lightweight Flask deployment enables on-demand, low-latency segmentation and automated marketing recommendations, successfully translating raw cluster assignments into real-time actionable directives via the web interface. Compared to traditional manual RFM and demographic baselines, the proposed framework delivers finer-grained behavioral insight, horizontal scalability, and direct operational utility for small and medium-sized e-commerce operators. Future work will broaden the research scope to include alternative clustering algorithms, richer behavioral feature engineering, and integrated customer churn risk predictive modeling.

REFERENCES

[1] J. M. John, O. Shobayo, and B. Ogunleye, An exploration of clustering algorithms for customer segmentation in the UK retail market, Analytics, vol. 2, no. 4, pp. 809-823, 2023.

[2] M. Al-Ahmadi, Evaluating demographic vs. behavioral segmentation in modern e-commerce platforms, Journal of Digital Transformation, vol. 16, no. 1, pp. 120-135, 2024.

[3] A. Ala'a, R. Al-Hmouz, and H. Abdulwahab, Enhanced RFM model for dynamic customer segmentation in high-velocity retail environments, IEEE Access, vol. 10, pp. 103615–103629, 2022, doi: 10.1109/ACCESS.2022.3209911.

[4] N. Chawla, Data mining for imbalanced datasets: An overview, in Data Mining and Knowledge Discovery Handbook, O. Maimon and L. Rokach, Eds. New York, NY, USA: Springer, 2010, pp. 875–886, doi: 10.1007/978-0-387-09823-4_45.

[5] Chong B, K-means clustering algorithm: A brief review, Academic Journal of Comp. & Info. Science, vol. 4, no.5, 2021, doi: 10.25236/AJCIS.2021.040506.

[6] N. Kumar, Intelligent customer segmentation: Unveiling consumer patterns with machine learning, Journal of Umm Al-Qura University for Engineering and Architecture, vol. 16, no. 3, pp. 774-783, 2025.

[7] J. M. Spoor, Improving customer segmentation via classification of key accounts as outliers, Journal of Revenue and Pricing Management, 2022, doi: 10.1057/s41270-022-00185-4.

[8] C. P. Ufeli, M. U. Sattar, R. Hasan, and S. Mahmood, Enhancing customer segmentation through FAMD-based approach using K-means and hierarchical clustering, Information, vol. 16, no. 6, p. 441, 2025, doi: 10.3390/info16060441.

[9] A. Wasilewski, K. Juszczyszyn, and V. Suryani, Multi-factor evaluation of clustering methods for e-commerce application, Egyptian Informatics Journal, vol. 28, p. 100562, 2024.

[10] F. O. Kocoglu and T. Ozcan, A grid search optimized extreme learning machine approach for customer churn prediction, Journal of Engineering Research, 2022, doi: 10.36909/jer.16771.

[11] H. E. Sanches, A. T. Possebon, and L. B. R. Aylon, Churn prediction for SaaS company with machine learning, Industrial Management & Data Systems, 2024, doi: 10.1108/INMR-06-2023-0101.

[12] X. Yan, Y. Li, F. Nie, and R. Li, Bank customer segmentation and marketing strategies based on improved DBSCAN algorithm, Applied Sciences, vol. 15, no. 6, p. 3138, 2025, doi: 10.3390/app15063138.

[13] A. S. Paramita and T. Hariguna, Comparison of K-Means and DBSCAN algorithms for customer segmentation in e-commerce, 2024.

[14] H. Senturk, E. Gecici, and S. Alp, Customer segmentation with clustering methods in the retail industry, Istanbul Aydin Universitesi Sosyal Bilimler Dergisi, vol. 16, no. 4, pp. 551-573, 2024.

[15] S. Bathina, Precision pulse: AI-driven micro-segmentation for optimized retail customer engagement, Computer Fraud & Security, vol. 2025, pp. 1479-1487, 2025.

[16] A. Wasilewski, Customer segmentation in e-commerce: A context-aware quality model for comparing clustering algorithms, Journal of Internet Services and Applications, 2024, doi: 10.5753/jisa.2024.3851.

[17] F. A. Alijoyo et al., Personalized marketing: Leveraging AI for culturally aware segmentation and targeting, Alexandria Engineering Journal, vol. 119, pp. 8-21, 2025, doi: 10.1016/j.aej.2024.11.056.

[18] P. Rajapandian, A. Karunamurthy, V. Vasanth, and M. Meganathan, E-commerce customer segmentation: A clustering approach in a web-based platform, Journal of Engineering Technology and Applied Physics, 2025.

[19] P. P. Arcot et al., A web-based e-commerce recommender system using Flask and ML algorithms, in Proc. 2025 IEEE 2nd Int. Conf. ICITEICS, 2025, pp. 1-6.

[20] L. S. Ling and C. T. Weiling, Enhancing segmentation: A comparative study of clustering methods, IEEE Access, 2025, doi: 10.1109/ACCESS.2025.3550339.

[21] A. M. Sakinah and D. S. Awaliyah, Optimization e-commerce consumer segmentation based on K-Means clustering and machine learning, Journal of Mathematics, Computations and Statistics, 2025.

[22] S. P. Santhoshkumar et al., Customer segmentation using machine learning, in Proc. 2025 8th Int. Conf. ICCPCT, 2025, pp. 1673-1677.

[23] A. Parupalli, Optimizing customer engagement in multi-source e-commerce retail datasets using AI-based efficient techniques, in Proc. 2025 5th Int. Conf. AISP, 2025, pp. 1-7.

[24] N. Gankidi et al., Customer segmentation using machine learning, in Proc. 2022 2nd Int. Conf. CONIT, 2022, pp. 1-5.

[25] M. R. K. Ibrahim and R. Tyasnurita, LRFM model analysis for customer segmentation using K-means clustering, in Proc. 2022 Int. Conf. IEIT, 2022, pp. 248-253.

[26] V. Arul, A. Kumar, and A. Agarwal, Segmenting mall customers data to improve business using K-means clustering, in Proc. 2021 3rd Int. Conf. ICAC3N, 2021, pp. 82-86.

[27] R. Pradhan, Customer segmentation using clustering approach based on RFM analysis, in Proc. 2021 5th Int. Conf. ISCON, 2021, pp. 1-6.

[28] J. Zhang, Improving predictive accuracy of customer churn models through data cleaning in the telecommunications industry, AEMPS Advances in Econ. Manage. Politi. Sci., 2025.

[29] M. Hassan et al., Overcoming data dimensionality in rule-based customer analysis systems, Computing Frontiers, vol. 27, no. 4, pp. 195-212, 2022.

[30] S. Sagiroglu and D. Sinanc, Big data: A review, in Proc. 2013 Int. Conf. Collaboration Technologies and Systems (CTS), San Diego, CA, USA, 2013, pp. 42–47, doi: 10.1109/CTS.2013.6567202.