Journal of Applied Informatics and Computing (JAIC)
Vol.9, No.5, October 2025, pp. 2418~2425
e-ISSN: 2548-6861

Enhancing E-Commerce Customer Segmentation with Fuzzy C-Means Soft Clustering Probabilities

Muhamad Iqbal Januadi Putra [1]*, Vincent Alexander [2], Ahmad Chusyairi [3]*, Raka Admiral Abdurrahman [4]***, Alexander Daniel Pratama [5]******
* Department of Computer Science, Universitas Siber Asia
** Department of Computer Science, Universitas Tarumanagara
*** Department of Computer Science, Politeknik Negeri Malang
**** Department of Geodesy, Institut Teknologi Bandung
2024.iqbal.januadi@student.unsia.ac.id [1], alex.535220149@stu.untar.ac.id [2], ahmadchusyairi@lecturer.unsia.ac.id [3], raka.admiral.abdurrahman@gmail.com [4], alexanderdan.pratama@gmail.com [5]

Article Info

Article history:

Received 2025-08-05
Revised 2025-08-31
Accepted 2025-09-10

Keyword:

Clustering,
Customer segmentation,
E-commerce,
Fuzzy C means,
Soft clustering probability.

ABSTRACT

Customer segmentation is of paramount importance in the e-commerce industry, enabling businesses to improve marketing strategies and customer engagement. This study compares the performance of two clustering algorithms, K-Means and Fuzzy C-Means (FCM), using Walmart's public e-commerce dataset of 550,068 transactions. After preprocessing and normalization, the elbow method was applied to determine the optimal number of clusters, yielding seven clusters for K-Means and eight for FCM. Experimental evaluation based on the silhouette score shows that FCM achieved 0.48, outperforming K-Means which scored 0.36, indicating that FCM generated clusters with stronger cohesion and separation. However, this improvement comes at a computational cost. K-Means consistently required less than 0.02 seconds per run, while FCM averaged 0.3 seconds and peaked at 1.38 seconds when the number of clusters increased, making it approximately 20–30 times slower. Cluster distribution analysis further revealed that K-Means produced an uneven segmentation dominated by a single large cluster, whereas FCM generated a more balanced distribution across its clusters. This demonstrates the advantage of FCM in capturing overlapping and multidimensional customer behaviors through partial memberships, in contrast to the rigid and oversimplify assignments of K-Means. These findings highlight the benefit of adopting FCM for e-commerce segmentation, as it provides more interpretable and actionable insights for personalized marketing. At the same time, the trade-off between clustering quality and computation time suggests that future research should explore optimization techniques such as parallelization, approximate fuzzy clustering, or hybrid models that combine the efficiency of hard clustering with the interpretability of soft clustering.

I. INTRODUCTION

E-commerce has become one of the top industries in global economy, enabling company to penetrate vast customer in wide geographic area [1]. The presence of e-commerce has changed the consumer behaviour from traditional shopping outlets to internet-based transaction [2]. This condition makes customers plays more important role in shaping the e-commerce ecosystem by independently showing their preferences in buying product [1, 3]. However, diverse customer shows diverse preferences, behaviousm and purchasing pattern [4]. As a result, to meet diverse customer needs, customer segmentation has played more vital role for company to gain more profit by becoming more client centric [5, 6]

Customer segmentation is an essential in industry in order to help a better understanding of their customer behaviour [7]. By segmenting their customer, company can provide a personalized products and services for each customer cluster [1]. The idea behind customer segmentation arises from the fact that each customer is unique. Therefore, different strategies may be required to keep retaining them [7]. Customer segmentation is performed by dividing all the customer into several groups based on their demographics, geographics, behaviour, and consumption attributes [8, 9]. This analysis can be undertaken by adopting unsupervised machine learning method such as K-Means, DBSCAN, and hierarchical clustering.

The adoption unsupervised clustering method for customer segmentation in e-commerce is not a new case. There are plenty of research has been done to perform clustering algorithm in customer segmentation, without any standardized guideline for systematic implementation [10]. One of the popular algorithms used for customer segmentation is K-Means [11], as the most efficient algorithm [12]. Kumar et al. [13] performed customer segmentation in e-commerce using K-Means and hierarchical cliustering to identify distinct customer segments. Similarly, Tabinan et al. [14] succeed implementing K-Means clustering method to divide the e-commerce customers from their demographic psychographic, behavioural, and geographic data. Rajput et al. [15] employed the K-means algorithm to analyze customer purchasing behavior, aiming to enhance value creation for businesses. Paramita & Hariguna [4] used K-Means and DBSCAN algorithms to segment the customer from e-commerce company. Othayoth & Muthalagu [16] performed various machine learning unsupervised technique, such as agglomerative clustering algorithm and filtering-based recommender system to segment customer. All these papers have clear goal in dividing their customer into distinct groups based on the pattern.

The customer segmentation using unssupervised machine learning algorithms such as K-Means, DBSCAN, and hierarchical clustering has been widely used and explored in many papers. These algorithms can divide the dataset into distinct groups based on the pattern. In the domain of e-commerce, K-Means, DBSCAN, and hierarchical clustering is performed to cluster the customer into several groups based on the demographics, geographic, behaviour, and transaction pattern [4, 13, 14].

While it seems clear that customer can be assigned into distinct group, in reality, customer in e-commerce may show overlapping patterns. For instance, one customer may show the purchasing frequency of a loyal buyer while simultaneously demonstrating the browsing habits of a bargain seeker. Forcing such customers into a single group oversimplifies their behavior and risks misclassification. This limitation reduces the effectiveness of customer segmentation in e-commerce, as it ignores cross-segment tendencies, oversimplifies multidimensional data, and may lead to less personalized marketing strategies.

Traditional clustering algorithms like K-Means are widely used due to their simplicity and efficiency, but they rely on the assumption that each customer belongs exclusively to one cluster. This “hard assignment” approach creates rigid boundaries that may not capture the complexity of real e-commerce behavior. There is always probability of customer showing ambiguous pattern which can belong to more than one cluster. This fact fails to be discussed in many papers that related to the customer segmentation, especially in the domain of e-commerce. Therefore, this paper investigates new point of view in adopting soft cluster probability in segmenting e-commerce customer using Fuzzy C Means cluster. Moreover, the adoption of Fuzzy C Means clustering for ecommerce customer segmentation, which provide soft clustering based on the probability, remained unexplored.

The Fuzzy C-Means method is chosen for this study because of its ability to handle the inherent uncertainty and fuzziness in customer data. Traditional clustering methods, such as K-Means, force customers into rigid groups, which can lead to oversimplification and loss of valuable insights. In contrast, FCM acknowledges the complexity of customer behavior by allowing for partial membership in multiple clusters. This flexibility makes FCM particularly suitable for e-commerce, where customer preferences are often fluid and multidimensional. Additionally, FCM provides a probabilistic framework that can be leveraged to interpret the segmentation results more effectively, enabling businesses to design more personalized and targeted marketing strategies.

The aim of this paper is to explore the application of the Fuzzy C-Means method in e-commerce customer segmentation, with a focus on improving the interpretation of segmentation results through soft clustering probabilities. By adopting FCM, this study seeks to address the limitations of traditional hard clustering methods and provide a more accurate representation of customer behavior. Therefore, this paper also will compare the output of FCM clustering method with the common clustering algorithm, K-Means [11], as a benchmark due to the its simplicity [12].

II. METHOD

A. Fuzzy C Means

The Fuzzy C-Means (FCM) method is a soft clustering technique that addresses the limitations of traditional hard clustering methods, such as K-Means, by allowing data points to belong to multiple clusters with varying degrees of membership and probabilities (Figure 1) [17, 18]. This method is based on fuzzy logic method [19]. Unlike hard clustering, where each customer is assigned to a single, distinct group, FCM assigns a probability or membership value to each group [20,21], indicating their likelihood of belonging to each cluster. This approach is particularly useful in e-commerce customer segmentation, where customers often exhibit overlapping behaviors, preferences, and purchasing patterns that cannot be neatly categorized into a single group. For example, a customer might frequently purchase both electronics and fashion items, making them a potential member of both "tech-savvy" and "fashion-forward" segments. FCM captures this ambiguity by providing a more nuanced understanding of customer behavior. FCM method produces row of clusters with degree of membership and probabilities [19].

The Fuzzy C-Means (FCM) algorithm is an iterative clustering method that minimizes an objective function to partition data into clusters. The algorithm assigns membership values to data points, indicating their degree of belonging to each cluster. Below is the mathematical formulation of the FCM algorithm:

$$J = \sum_{i=1}^{n}\sum_{j=1}^{c}(u_{ij})^m . ||x_i - v_j||^2 \qquad (1)$$

where n is total number of data points (customers), c is total number of clusters, uij is membership value of the i-th data point to the j-th cluster, where 0≤uij≤10≤uij≤1, m is fuzziness parameter (typically m≥1m≥1), which controls the degree of fuzziness in the clustering, xi is the i-th data point (e.g., customer features), vj is the centroid of the j-th cluster, and ∥xi−vj∥2 is squared Euclidean distance between the i-th data point and the j-th cluster centroid.

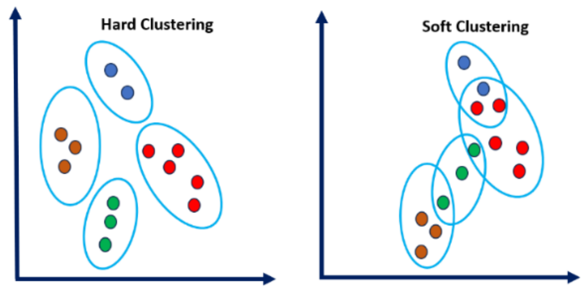

Figure 1. Difference between hard and soft clustering

The membership values uijuij are updated iteratively using the following formula:

$$u_{ij} = \frac{1}{\sum_{k=1}^{c}(\frac{||x_i - v_j||}{||x_i - v_k||})^{\frac{2}{m-1}}} \qquad (2)$$

where uij is membership value of the i-th data point to the j-th cluster and k is index for clusters (ranging from 1 to cc).

The cluster centroids vj are updated iteratively using the following formula:

$$v_j = \frac{\sum_{i=1}^{n}(u_{ij})^m . x_i}{\sum_{i=1}^{n}(u_{ij})^m} \qquad (3)$$

where vj is the centroid of the j-th cluster and xi is the i-th data point.

B. Datasets

This study uses Walmart's public e-commerce dataset which can be accessed in Kaggle. The dataset consists of 550,068 entries and 10 columns, stored in a structured format (e.g., CSV or database) and contains both numerical and categorical variables. The dataset is a mix of both numerical and categorical variables, providing a robust foundation for a variety of analytical tasks.

The dataset offers valuable insights into Walmart's customer base and purchasing behavior, allowing for analysis of customer demographics, product preferences, and spending patterns. This is paramount important to build customer segmentation. The detail of column in this dataset can been in Table 1.

TABEL I
DETAIL OF DATASETS

No	Column Name	Detail	Data Type
1	User_ID	ID of User	int64
2	Product_ID	ID of Product	object
3	Gender	Gender in code M or F	object
4	Age	Category of Age (0-17, 18-25, … , 55+)	object
5	Occupation	Category of Occupation (1, 2, 3, 4 …., 18)	int64
6	City_Category	Category of City (A, B, C)	object
7	Stay_In_Current_City_Years	Category of length time (0, 1, 2, 3, 4+)	object
8	Marital_Status	Category of marital status (0, 1)	int64
9	Product_Category	Product Category	int64
10	Purchase	Purchase	int64

C. Data Processing

The data processing phase is a critical step in preparing the dataset for customer segmentation using the Fuzzy C-Means (FCM) and K-Means algorithm, ensuring the dataset is clean from the noise, outlier, and desired format [22]. This section outlines the steps taken to clean, transform, and preprocess the data to ensure its suitability for clustering analysis. The e-commerce Walmart data has no missing value. Therefore, this step includes data transformation, data normalization/scaling, and feature engineering. The workflow of this study is portrayed in the Figure 2.

Data transformation is performed to prepare the dataset by converting one format into another format [22]. Categorical variables, such as gender and city category, are converted into numerical values using one-hot encoding technique. Some columns store numerical value in object or string format, such as occupation and length of stay. Therefore, converting the data type into numerical is performed for these columns. In addition, age column is stored in categorical format with the format of range value. Binning method is performed to map this categorical age value into numeric value. Then, numerical features are normalized or standardized to bring all variables to a common scale, as clustering algorithms are sensitive to the magnitude of features. Normalization rescales features to a range of 0-1 using min-max scaling [21, 23]. This step ensures that all features contribute equally to the clustering process. Feature selection is then conducted to improve the quality of clustering and reduce computational complexity.

After completing the data processing phase, the next step is to perform customer segmentation using clustering algorithms. In this study, two clustering methods are employed Fuzzy C-Means (FCM) and K-Means. Both algorithms are unsupervised learning techniques used to group customers into clusters based on their similarities. However, they differ in their approach: K-Means assigns each customer to a single, distinct cluster (hard clustering), while FCM allows customers to belong to multiple clusters with varying degrees of membership (soft clustering). The modeling process involves several steps, including determining the optimal number of clusters, training the models, and evaluating their performance. There are hyperparameters or initial value for FCM algorithm, such as fuzziness parameter (m), stopping criterion, and maximum number of iterations [21]. This study used m = 1.7, stopping criterion = 0.005, and maximum number of interations = 1000.

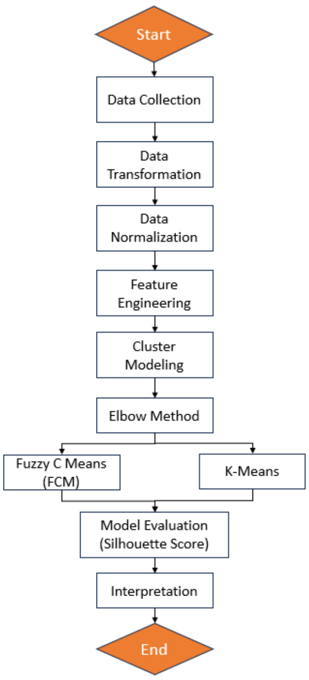

Figure 1. Workflow of the study

The first step in the modelling process is to determine the optimal number of clusters (k). This is a critical step because the number of clusters directly impacts the quality of segmentation. To identify the optimal k value, the elbow method is used. The elbow method involves running the K-Means algorithm for a range of k values (e.g., from 2 to 10) and calculating the Within-Cluster-Sum of Squares (WCSS) for each k. WCSS measures the total variance within clusters, and the goal is to minimize this value. As k increases, WCSS decreases, but the rate of decrease slows down after a certain point. The optimal k value is chosen at the "elbow" of the WCSS curve, where the rate of decrease sharply changes. This point represents a balance between the number of clusters and the compactness of the clusters. Once the optimal k value is determined, it is used for both K-Means and FCM to ensure a fair comparison.

To compare the performance of K-Means and FCM, the silhouette score is used as an evaluation metric. The silhouette score measures how similar a customer is to their own cluster compared to other clusters [24]. It ranges from -1 to 1, where a higher score indicates better-defined clusters. A score close to 1 suggests that customers are well-matched to their clusters, while a score close to 0 indicates overlapping clusters. A negative score implies that customers may have been assigned to the wrong cluster. The silhouette score is calculated for both K-Means and FCM, and the algorithm with the higher score is considered to provide better segmentation.

III. Result And Discussion

This section presents the results of the customer segmentation analysis using both K-Means and Fuzzy C-Means (FCM) clustering algorithms. The discussion focuses on the optimal number of clusters, the characteristics of each cluster, and the comparative performance of the two methods based on the silhouette score. Additionally, the unique advantages of FCM in providing a probabilistic interpretation of customer segments are highlighted.

Both clustering methods, K-Means and FCM, were evaluated to determine the optimal number of clusters using the elbow method. The Within-Cluster-Sum of Squares (WCSS) was calculated for k values ranging from 2 to 10. The results showed a noticeable “elbow” at k = 7 for K-Means and k = 8 for FCM, indicating that these values provided the best balance between minimizing intra-cluster variance and maintaining interpretability for each respective algorithm. Therefore, K-Means was modeled using seven clusters, while FCM was modeled using eight clusters, to ensure that both methods operated under their optimal configurations (Figure 3).

The formed clusters are examined to uncover shared traits and trends among customers in each group. These insights are then used to design tailored marketing strategies [23]. The sample detail of each cluster is shown in Table 2. The clusters provide a detailed segmentation of the customer base, highlighting distinct demographic and behavioral patterns. Clusters 2, 3, and 6 represent high-spending groups, with variations in age, marital status, and occupation levels. Clusters 0 and 1 consist of moderate spenders, while Clusters 4 and 5 represent low-spending groups, primarily older and married individuals. These insights enable businesses to tailor marketing strategies, product offerings, and customer engagement initiatives to meet the unique needs of each segment.

TABEL II
CLUSTERS DETAIL

Cluster	Detail
0	• Demographics: All female, middle-aged (average age = 2.25), with middle-to-high occupation levels (average occupation code = 6.4). All are unmarried. • Spending: Moderate average purchase amount (593.847). • Interpretation: This group represents productive, unmarried females with moderate spending habits, likely balancing work and personal life.
1	• Demographics: All male, middle-aged (average age = 2.27), with higher occupation levels (average occupation code = 8.7). All are unmarried. • Spending: Moderate average purchase amount (570.000). • Interpretation: This cluster consists of productive, professional unmarried males with moderate spending patterns, possibly focused on career growth.
2	• Demographics: Mostly male, middle-aged (average age = 2.99), with higher occupation levels (average occupation code = 8.4). All are married. • Spending: High average purchase amount (1,230,000). • Interpretation: This group represents married, professional males with high spending habits, likely driven by family needs and higher disposable income.
3	• Demographics: Mostly male, young (average age = 2.13), with middle-to-high occupation levels (average occupation code = 7.9). All are single. • Spending: High average purchase amount (1,245,000). • Interpretation: This cluster consists of young, single individuals with high spending habits, likely due to their higher occupational levels and fewer financial responsibilities.
4	• Demographics: All male, older (average age = 3.36), with high occupation levels (average occupation code = 9). All are married. • Spending: Low average purchase amount (552,000). • Interpretation: This group represents older, married male professionals with low spending habits, possibly due to conservative financial behavior or savings-focused lifestyles.
5	• Demographics: All female, older (average age = 3.3), with moderate occupation levels (average occupation code = 6.8). All are married. • Spending: Low average purchase amount (597,000). • Interpretation: This cluster consists of older, married females with limited purchasing power, likely prioritizing essential expenses over discretionary spending.
6	• Demographics: Mostly male, young (average age = 2.4), with moderate-to-high occupation levels (average occupation code = 8.3). Mostly single. • Spending: Very high average purchase amount (1,346,000). • Interpretation: This group represents young males with moderate-to-high job profiles and very high spending habits, likely driven by a combination of disposable income and lifestyle preferences.

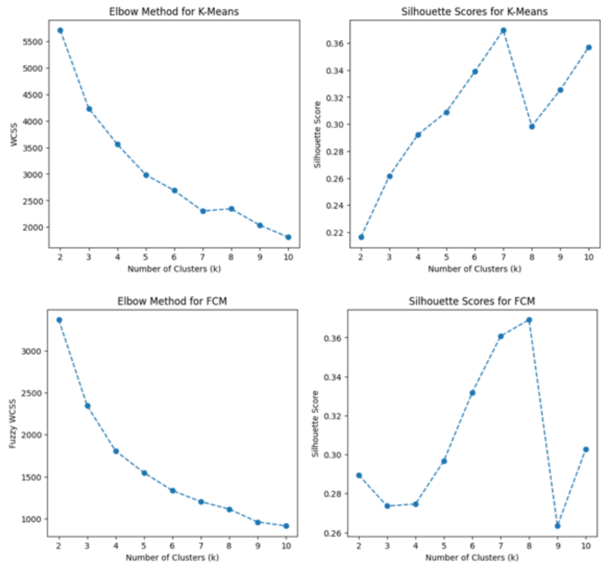

Figure 3. Performance result of FCM and K-Means

The performance of the two clustering methods was evaluated using the silhouette score, which measures how well-defined the clusters are. The silhouette score ranges from -1 to 1, with higher values indicating better clustering results. The experimental results show that the Fuzzy C-Means (FCM) model achieved a higher silhouette score (0.48) compared to the K-Means model (0.36). This indicates that the clusters generated by FCM are more cohesive and better separated, providing a clearer structure for customer segmentation. The improvement can be attributed to FCM's ability to assign customers partial memberships across multiple clusters, which better captures the overlapping and multidimensional nature of e-commerce customer behavior. In contrast, K-Means applies hard assignments that force each customer into a single cluster, often oversimplifying complex purchasing patterns and leading to lower cluster quality. These findings highlight the advantage of soft clustering probabilities in delivering more accurate and actionable customer segmentation insights for e-commerce platforms. Additionally, FCM has more ability ability to handle overlapping customer behaviors, which K-Means, as a hard clustering method, fails to capture effectively.

The distribution of customers across clusters revealed notable differences between K-Means and FCM. As shown in Figure 4, K-Means with seven clusters produced an uneven distribution, with one dominant cluster (Cluster 1) containing more than 1,200 customers, while other clusters such as Cluster 5 contained fewer than 500 customers. This imbalance suggests that K-Means tends to oversimplify customer behavior by forcing boundary cases into a single cluster, which reduces the granularity of segmentation. In contrast, FCM with eight clusters resulted in a more balanced distribution, with most clusters ranging between 500 and 1,000 customers and only Cluster 7 exceeding 1,200 customers. This outcome reflects FCM's ability to assign partial memberships, thereby reducing extreme dominance of a single cluster and capturing overlapping patterns more effectively. From a business perspective, FCM provides finer segmentation granularity, which can support the design of more personalized marketing strategies compared to the less balanced grouping produced by K-Means.

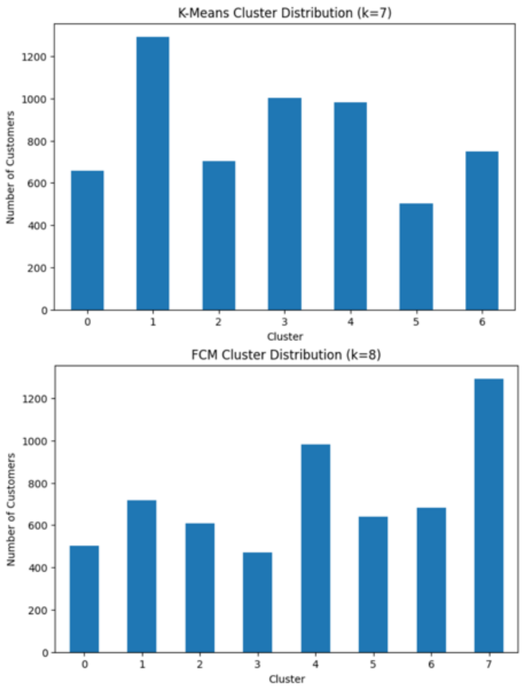

Figure 4. Cluster distribution of FCM and K-Means

One of the key advantages of FCM over K-Means is its ability to assign customers to multiple clusters based on membership probabilities. Unlike K-Means, where each customer is assigned to a single, distinct cluster, FCM provides a soft clustering approach. Each customer is assigned a membership value for all clusters, reflecting their degree of belonging to each segment. For example, consider the following two customers (Table 3):

a) User ID 1000002: Assigned to Cluster 1 by K-Means, but FCM reveals a more nuanced picture. This customer has a 13.2% probability of belonging to Cluster 0, 6% to Cluster 1, 24% to Cluster 2, and 38% to Cluster 4. This indicates that while K-Means categorizes them strictly into one group, FCM shows they exhibit behaviors aligning with multiple clusters, particularly Cluster 4. As a result, company can predict the groups of customers belonging to.

b) User ID 1000003: Assigned to Cluster 6 by K-Means, but FCM shows a 8% probability for Cluster 0, 21% for Cluster 1, 18% for Cluster 2, and 25% for Cluster 4. This suggests that the customer's behavior is not confined to a single cluster but overlaps with several, especially Cluster 4.

TABEL III
EXAMPLE OF CLUSTER MEMBERSHIP AND PROBABILITY

User ID	**K-Mean Clusters**	**FCM Cluster 0 Prob**	**FCM Cluster 1 Prob**	**FCM Cluster 2 Prob**	**FCM Cluster 3 Prob**	**FCM Cluster 4 Prob**	**FCM Cluster 5 Prob**	**FCM Cluster 6 Prob**
1000002	1	0.13	0.06	0.24	0.07	0.38	0.046	0.049
1000003	6	0.08	0.21	0.18	0.057	0.25	0.09	0.11
1000008	4	0.57	0.03	0.09	0.1	0.11	0.053	0.027
1000027	4	0.87	0.007	0.03	0.032	0.034	0.014	0.006

This probabilistic interpretation offers a more nuanced understanding of customer behavior, as it acknowledges that customers often exhibit characteristics of multiple segments. This flexibility is particularly valuable for e-commerce businesses, as it allows for more personalized and targeted marketing strategies. For instance, a customer with significant membership probabilities in multiple clusters, such as Cluster 2 and Cluster 4, could be targeted with promotions tailored to both high-spending and moderate-spending behaviors, maximizing the likelihood of engagement. In contrast, K-Means' rigid assignment of customers to a single cluster may oversimplify their behavior and lead to less effective strategies.

While the interpretation of FCM outperformed K-Means in this study, it is important to acknowledge its limitations. FCM requires careful tuning of the fuzziness parameter m, distance metrics [26] and its computational complexity is higher than that of K-Means. In addition, FCM tends to create high membership values for the outliers [27]. Future work could explore the impact of different values of m on clustering results and investigate hybrid approaches that combine the strengths of both hard and soft clustering methods. Additionally, integrating external data sources, such as social media activity or customer feedback, could further enhance the segmentation results.

Another limitation observed in this study is the significantly higher computation time of FCM compared to K-Means, as shown in Figure 5. While K-Means maintained nearly constant runtime across different cluster sizes (approximately 0.01–0.02 seconds per run), FCM required substantially more processing time, ranging from ~0.07 seconds at k = 2 to ~1.38 seconds at k = 9. On average, FCM was nearly 20–30 times slower than K-Means in this experiment. The increasing trend with larger k reflects FCM's iterative process of updating membership probabilities for every data point across all clusters, rather than assigning each point to a single cluster as in K-Means. This leads to higher computational complexity and slower convergence, especially in large-scale datasets.

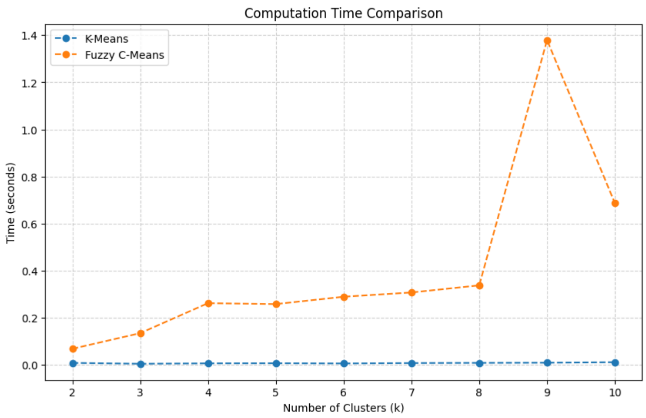

Figure 5. Comparison of computation time between FCM and K-Means

Although this additional computation enables a more nuanced segmentation with improved silhouette scores (0.48 vs. 0.36), it highlights a trade-off between segmentation quality and computational efficiency. In real-world e-commerce applications where datasets can reach millions of records, the higher runtime of FCM may pose scalability challenges. This limitation suggests the need for optimization techniques, such as parallelization, approximate fuzzy clustering, or hybrid approaches that combine the speed of hard clustering with the interpretability of soft clustering.

IV. CONCLUSION

This study explored the application of K-Means and Fuzzy C-Means (FCM) clustering algorithms for customer segmentation in the e-commerce domain. Both methods identified seven distinct customer segments, each characterized by unique demographic and behavioral patterns. However, the evaluation results revealed clear differences in clustering quality. The FCM model achieved a higher silhouette score (0.48) compared to K-Means (0.36), indicating that FCM produced clusters that were more cohesive and better separated. Unlike K-Means, which assigns each customer to a single cluster and may oversimplify complex behavioral traits, FCM provides a probabilistic framework that allows customers to hold partial memberships across multiple clusters. This ability to capture overlapping and multidimensional customer behaviors makes FCM more suitable for e-commerce segmentation, where customer preferences are often fluid and heterogeneous.

Beyond quality, the analysis of cluster distribution highlighted further differences between the two methods. K-Means produced an uneven grouping with one dominant cluster containing the majority of customers, while FCM yielded a more balanced segmentation across its eight clusters, enabling finer granularity and more actionable marketing insights.

Nevertheless, it is important to acknowledge the limitation of FCM in terms of computational efficiency. The experiments showed that FCM required significantly longer computation time compared to K-Means, with runtime increasing as the number of clusters grew. On average, FCM was found to be 20–30 times slower than K-Means due to its iterative updates of membership probabilities across all clusters. While this trade-off yields richer and more interpretable segmentation results, it may hinder scalability in real-world applications involving very large datasets.

The findings of this study underscore the importance of adopting advanced clustering techniques like FCM for customer segmentation in e-commerce. By providing a more flexible and interpretable framework, FCM enables businesses to gain deeper insights into customer behavior and develop data-driven strategies that enhance customer engagement and drive profitability. Due to this limitation, future research should explore optimization techniques such as parallelization, approximate fuzzy clustering, or hybrid approaches that combine the speed of hard clustering with the interpretability of soft clustering, further refine segmentation accuracy.

REFERENCES

[1] L. Li, "Analysis of e-commerce customers' shopping behavior based on data mining and machine learning," *Soft Computing*, Jul. 2023, doi: 10.1007/s00500-023-08903-5.

[2] R.-S. Wu and P.-H. Chou, "Customer segmentation of multiple category data in e-commerce using a soft-clustering approach," *Electronic Commerce Research and Applications*, vol. 10, no. 3, pp. 331–341, May 2011, doi: 10.1016/j.elerap.2010.11.002.

[3] R. Punhani, V. P. S. Arora, A. S. Sabitha, and V. K. Shukla, "Segmenting e-commerce customer through data mining techniques," *J. Phys.: Conf. Ser.*, vol. 1714, no. 1, pp. 012026, Jan. 2021, doi: 10.1088/1742-6596/1714/1/012026.

[4] A. S. Paramita and T. Hariguna, "Comparison of K-Means and DBSCAN algorithms for customer segmentation in e-commerce," *Journal of Digital Market and Digital Currency.*, vol. 1, no. 1, pp. 43–62, Jun. 2024, doi: 10.47738/jdmdc.v1i1.3.

[5] R. Punhani, V. P. S. Arora, S. Sabitha and V. Kumar Shukla, "Application of clustering algorithm for effective customer segmentation in e-commerce," *2021 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE)*,

Dubai, United Arab Emirates, 2021, pp. 149-154, doi: 10.1109/ICCIKE51210.2021.9410713.

[6] R. J. Kuo, M. N. Alfareza, and T. P. Q. Nguyen, “Genetic based density peak possibilistic fuzzy c-means algorithms to cluster analysis-a case study on customer segmentation,” *Engineering Science and Technology, an International Journal*, vol. 47, pp. 101525, Nov. 2023, doi: 10.1016/j.jestch.2023.101525.

[7] M. Sivaguru, “Dynamic customer segmentation: a case study using the modified dynamic fuzzy c-means clustering algorithm,” *Granular Computing*, vol. 8, no. 2, pp. 345–360, Mar. 2023, doi: 10.1007/s41066-022-00335-0.

[8] N. P. P. Yuliari, I. K. G. D. Putra, and N. K. D. Rusjayanti, “Customer segmentation through fuzzy c-means and fuzzy RFM method,” *Journal of Theoretical and Applied Information Technology*, vol. 3, no. 78, pp. 380, 2015.

[9] Y. Li, X. Chu, D. Tian, J. Feng, and W. Mu, “Customer segmentation using K-means clustering and the adaptive particle swarm optimization algorithm,” *Applied Soft Computing*, vol. 113, pp. 107924, Dec. 2021, doi: 10.1016/j.asoc.2021.107924.

[10] B. Shen, “E-commerce customer segmentation via unsupervised machine learning,” in *The 2nd International Conference on Computing and Data Science*, Stanford CA USA: ACM, Jan. 2021, pp. 1–7, doi: 10.1145/3448734.3450775.

[11] M. Alves Gomes and T. Meisen, “A review on customer segmentation methods for personalized customer targeting in e-commerce use cases,” *Information Systems and e-Business Management*, vol. 21, no. 3, pp. 527–570, Sep. 2023, doi: 10.1007/s10257-023-00640-4.

[12] S. Koul and T. M. Philip, “Customer segmentation techniques on e-commerce,” in *2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, Greater Noida, India, 2021, pp. 135-138, doi: 10.1109/ICACITE51222.2021.9404659.

[13] S. Kumar, R. Rani, S. K. Pippal, and R. Agrawal, “Customer segmentation in e-commerce: K-means vs hierarchical clustering,” *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 23, no. 1, pp. 119, Feb. 2025, doi: 10.12928/telkomnika.v23i1.26384.

[14] K. Tabianan, S. Velu, and V. Ravi, “K-means clustering approach for intelligent customer segmentation using customer purchase behavior data,” *Sustainability*, vol. 14, no. 12, pp. 7243, Jun. 2022, doi: 10.3390/su14127243.

[15] L. Rajput and S. N. Singh, “Customer segmentation of e-commerce data using K-means clustering algorithm,” in *2023 13th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, Noida, India: IEEE, Jan. 2023, pp. 658–664, doi: 10.1109/Confluence56041.2023.10048834.

[16] S. P. Othayoth and R. Muthalagu, “Customer segmentation using various machine learning techniques,” *International Journal of Business Intelligence and Data Mining,* vol. 20, no. 4, pp. 480, 2022, doi: 10.1504/IJBIDM.2022.123218.

[17] F. Afrin, M. Al-Amin, and M. Tabassum, “Comparative performance of using PCA with K-means and fuzzy c means clustering for customer segmentation,” *International Journal of Scientific and Technology Research*, vol. 4, no. 8, pp. 70-74, 2015.

[18] S. E. Hashemi, F. Gholian-Jouybari, and M. Hajiaghaei-Keshteli, “A fuzzy c-means algorithm for optimizing data clustering,” *Expert Systems with Applications*, vol. 227, pp. 120377, Oct. 2023, doi: 10.1016/j.eswa.2023.120377.

[19] O. N. Purba, D. N. Sitompul, T. H. Harahap, S. R. D. Saragih, and R. F. Siregar, “Application of fuzzy c-means algorithm for clustering customers,” *Hanif Journal of Information Systems*, vol. 1, no. 1, pp. 26–36, Agu. 2023, doi: 10.56211/hanif.v1i1.8.

[20] K. E. Setiawan, A. Kurniawan, A. Chowanda, and D. Suhartono, “Clustering models for hospitals in Jakarta using fuzzy c-means and k-means,” *Procedia Computer Science*, vol. 216, pp. 356–363, 2023, doi: 10.1016/j.procs.2022.12.146.

[21] A. Chusyairi and P. R. N. Saputra, “Fuzzy c-means clustering algorithm for grouping health care centers on diarrhea disease,” *International Journal of Artificial Intelligence Research*, vol. 5, no. 1, Jan. 2021, doi: 10.29099/ijair.v5i1.191.

[22] K. Singh, P. M. Booma, and U. Eaganathan, “E-commerce system for sale prediction using machine learning technique,” *Journal of Physics Conference Series*, vol. 1712, no. 1, pp. 012042, Dec. 2020, doi: 10.1088/1742-6596/1712/1/012042.

[23] A. Chusyairi and P. R. N. Saputra, “Pengelompokan data puskesmas Banyuwangi dalam pemberian imunisasi menggunakan metode K-means clustering,” *Telematika*, vol. 12, no. 2, pp. 139–148, Agu. 2019, doi: 10.35671/telematika.v12i2.848.

[24] C. Irfiyanda, R. Andreswari and F. Hamami, “Customer segmentation using fuzzy c-means algorithm in telco industry,” in *2022 International Conference of Science and Information Technology in Smart Administration (ICSINTESA)*, Denpasar, Bali, Indonesia, 2022, pp. 1-4, doi: 10.1109/ICSINTESA56431.2022.10041585.

[25] B. Song, “A path to implementing a fresh produce e-commerce customer segmentation method based on clustering algorithms,” in *2023 IEEE 7th Information Technology and Mechatronics Engineering Conference (ITOEC)*, Chongqing, China: IEEE, Sep. 2023, pp. 1047–1050, doi: 10.1109/ITOEC57671.2023.10291762.

[26] S. Munusamy and P. Murugesan, “Modified dynamic fuzzy c-means clustering algorithm – Application in dynamic customer segmentation,” *Applied Intelligence*, vol. 50, no. 6, pp. 1922–1942, Feb. 2020, doi: 10.1007/s10489-019-01626-x.U.

[27] Rusdiana, I. Ernawati, N. Falih, and A. Arista, “Comparison of distance metrics on fuzzy C-Means algorithm through customer segmentation,” *2022 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*, vol. 4, pp. 307–311, Oct. 2021, doi: 10.1109/icimcis53775.2021.9699206.