

A Comparative Analysis of Clustering Techniques for Customer Behaviour Analysis and Personalized Marketing

Anuradha K. Kulkarni¹, Anuja U. Shetty², Maithilee S. Kune³,
Sneha D. Harne⁴, Tanmay M. Kohinkar⁵

Department of Information Technology¹⁻⁵

Sinhgad Institute of Technology SIT, Lonavala

anuradhak.kulkarni@gmail.com¹, anujashetty0124@gmail.com²

maithileekune@gmail.com³, harnesneha79@gmail.com⁴, tanmay.kohinkar29@gmail.com⁵

Abstract: Understanding customer behaviour has become central to modern marketing, as it helps organizations identify differences in customer preferences, purchasing habits, and levels of engagement. As customer data continues to grow, machine learning-based clustering has become a popular approach for customer segmentation and targeted marketing. However, existing studies vary considerably in terms of the datasets used, behavioural features selected, preprocessing techniques, clustering algorithms, and evaluation measures, making it difficult to determine which approaches are most effective.

This paper presents a comparative analysis of clustering techniques for customer behaviour analysis and personalized marketing. Existing research on customer segmentation is reviewed to identify the methods, features, algorithms, and evaluation measures used across different studies. The reported results are then compared using measures such as the Silhouette Score, Davies-Bouldin Index, and Calinski-Harabasz Index. The customer segments identified in these studies are examined based on their behavioural characteristics and potential applications in personalized marketing. The analysis also discusses the strengths and limitations of different clustering approaches and identifies research gaps that remain in customer segmentation.

Keywords: Customer Segmentation, Customer Behaviour Analysis, K-Means Clustering, Machine Learning, RFM Analysis, Personalized Marketing

I. INTRODUCTION

Every transaction, online interaction, and digital touchpoint generates customer data. This data provides insights into customer preferences, purchasing behaviour, and levels of engagement. When organizations analyse this data effectively, they gain a clearer understanding of their customer groups and can design marketing strategies that better suit their needs.

Customer segmentation is the process of grouping customers who share similar characteristics and behaviours. Machine learning-based clustering techniques are well suited for this task because they can identify groups within data without requiring predefined labels. K-Means is one of the most commonly used approaches because of its simplicity, efficiency, and interpretability. Researchers have also explored other approaches, including fuzzy clustering, deep clustering, and deep learning.

However, directly comparing these studies is difficult because they differ in their datasets, selected features, preprocessing methods, algorithms, and evaluation measures. Many studies also focus primarily on identifying



customer segments without clearly showing how these segments can be used to develop personalized marketing strategies.

This paper reviews and compares existing clustering approaches for customer behaviour analysis and personalized marketing. It examines the results reported in earlier studies and highlights the research gaps that remain.

II. LITERATURE SURVEY

Previous studies have explored different clustering techniques, customer features, and evaluation measures for customer segmentation and personalized marketing. Table 1 presents a comparative summary of the selected studies based on their methodology, clustering or AI techniques, main findings, and evaluation measures.

Table 1. Comparative Analysis of Existing Customer Segmentation Studies

Year	Study / Approach	Features / Methodology	Clustering / AI Technique	Main Finding	Evaluation
2026	Deep learning-driven customer segmentation model for online retail industry based on extended RFM model [1]	Extended RFM → Feature Engineering → GMM → Prediction	GMM; Conv1D-LSTM	Extended RFM captured additional customer behaviour and improved segmentation and prediction	Silhouette: 0.633 vs 0.576 for K-Means
2026	Customer Segmentation Based on RFM, Deep Embedded Clustering, and Lifetime Value Analysis for Loyalty Strategy Optimization [2]	RFM → K-Means & DEC → CLV Analysis	DEC; K-Means	Identified VIP/high-value customers and supported loyalty strategy optimization	DEC outperformed K-Means using Silhouette, DBI and CH
2026	Customer Segmentation Using K-Means Clustering Algorithm for Digital Marketing Strategy [3]	RFM → Min-Max Normalization → Elbow → K-Means	K-Means	Identified 4 customer segments and developed different digital marketing strategies	Elbow Method: K = 4
2025	Personalized	Customer Data →	K-Means;	Generated meaningful	MSE: 0.9212;



Year	Study Approach /	Features Methodology /	Clustering / AI Technique	Main Finding	Evaluation
	Marketing: Leveraging AI for Culturally Aware Segmentation and Targeting [4]	K-Means → Elbow + Silhouette → LIME	Explainable AI (LIME)	segments and supported personalized and ethical marketing	MAE: 0.9874 (reported model performance)
2025	A Framework for Customer Segmentation to Improve Marketing Strategies Using Machine Learning [5]	RFM → K-Means, Mini-Batch, Spectral & Fuzzy K-Means	Multiple ML clustering methods (K-Means, Mini-Batch, Spectral, Fuzzy K-Means)	Compared clustering methods and identified K-Means as a strong approach	K-Means Silhouette: 0.432619
2025	Customer Segmentation Using K-Means Clustering for Personalized Marketing Campaigns [6]	Customer demographic and purchase data → Normalization → Feature Selection → Dimensionality Reduction → K-Means	K-Means	Identified distinct customer segments based on purchasing patterns and behavioural trends to support personalized marketing campaigns	Elbow Method; Silhouette Analysis
2025	Customer segmentation in the digital marketing using a Q-learning based differential evolution algorithm integrated with K-means clustering [7]	Customer Data → Correlation Analysis → PCA → Q-learning-based Optimization → K-Means	K-Means + Q-learning-based Differential Evolution	Improved customer segmentation by optimizing K-Means and identifying distinct customer groups for targeted marketing strategies	Elbow Method; >95% classification accuracy of clustering results
2024	RFMI-Based Customer Segmentation	RFM + Purchase Interval → AHP → K-Means	K-Means; AHP	Identified 6 customer segments and improved	Segmentation and consumption-pattern analysis



Year	Study Approach /	Features Methodology /	Clustering / AI Technique	Main Finding	Evaluation
	with K-Means [8]			behavioural/value analysis	
2021	RFM Ranking – An Effective Approach to Customer Segmentation [9]	RFM → K-Means + Fuzzy C-Means → Improved Clustering	K-Means; Fuzzy C-Means	Improved clustering initialization and reduced iterations/execution time	Silhouette: 0.33 (K-Means), 0.43 (Fuzzy C-Means), 0.49 (RM K-Means)
2021	Customer Segmentation & Personalized Marketing [10]	RFM → K-Means → Apriori/ECLAT	K-Means; Association Rule Mining (Apriori/ECLAT)	Combined segmentation with product and offer recommendations	Clustering and association-rule evaluation

The literature demonstrates that K-Means is one of the most frequently used methods for customer segmentation because of its simplicity and interpretability. RFM-based features are also widely used to represent customer purchasing behaviour.

More advanced approaches have also been investigated. For example, the 2026 Extended RFM and deep learning-based study reported a Silhouette Score of 0.633 for its proposed approach compared with 0.576 for K-Means. Similarly, the 2021 RFM Ranking study reported an improvement from 0.33 for K-Means and 0.43 for Fuzzy C-Means to 0.49 for its proposed approach.

The 2025 framework study reported a K-Means Silhouette Score of 0.432619. Other studies focused on identifying an appropriate number of customer groups, such as four segments in the 2026 digital marketing study.

The literature also shows that some studies extend customer segmentation beyond clustering. For example, Deep Embedded Clustering, Explainable AI using LIME, association-rule mining, and Customer Lifetime Value analysis have been combined with segmentation to support more targeted marketing decisions.

III. RESEARCH GAP

The review of existing studies identifies several research gaps in customer behaviour analysis and personalized marketing.

1. Lack of standardized comparison: Existing studies use different datasets, features, preprocessing methods, clustering algorithms, and evaluation metrics, making direct comparison difficult.
2. Dependence on RFM-based features: Many studies primarily use Recency, Frequency, and Monetary features or their extensions. Additional behavioural and demographic attributes may provide broader customer insights.
3. Limited comparative evaluation: Many studies focus on a single clustering algorithm or compare only a few approaches. A common evaluation framework can provide a clearer understanding of their strengths and limitations.
4. Limited connection with marketing actions: Several studies identify customer groups without clearly connecting them to actionable personalized marketing strategies.

Therefore, there is a need for a systematic comparison using a common dataset, consistent preprocessing, and appropriate clustering evaluation measures, while also considering how customer segments can support personalized marketing.



IV. METHODOLOGY

The methodology of this study focuses on systematically reviewing and comparing existing research on customer segmentation and personalized marketing. The selected studies are analysed based on their customer features, preprocessing methods, clustering techniques, AI approaches, findings, and evaluation measures.

1. Study Selection

Relevant studies on customer segmentation, customer behaviour analysis, clustering, and personalized marketing are selected for comparison.

The selected studies cover different clustering and AI-based approaches.

2. Feature and Methodology Comparison

The customer features and methodologies used in the selected studies are examined.

Particular attention is given to RFM-based features and their extensions.

3. Clustering Technique Comparison

Different clustering approaches, including K-Means, Fuzzy C-Means, GMM, Deep Embedded Clustering, and deep learning-based methods, are compared.

4. Evaluation Comparison

The evaluation measures reported by the selected studies are analysed.

These include Silhouette Score, Davies-Bouldin Index, Calinski-Harabasz Index, and other measures reported by individual studies.

5. Comparative Analysis

The reported results, strengths, and limitations of the different approaches are compared.

Since the studies use different datasets, features, preprocessing methods, and evaluation measures, their numerical results are interpreted within their respective study contexts.

6. Research Gap Identification

The comparison is used to identify limitations and gaps in existing customer segmentation research.

The identified gaps are considered in terms of standardized evaluation, broader customer features, and the connection between customer segments and personalized marketing strategies.

V. RESULTS AND COMPARATIVE ANALYSIS

The comparison of the selected studies shows that K-Means is widely used for customer segmentation, mainly because it is simple and easy to interpret. At the same time, researchers have explored other techniques, such as Fuzzy C-Means, Gaussian Mixture Models (GMM), Deep Embedded Clustering (DEC), and deep learning-based approaches.

Different studies reported different clustering evaluation results. In the 2026 Extended RFM study, the proposed approach achieved a Silhouette Score of 0.633, while K-Means achieved 0.576. The 2021 RFM Ranking study reported Silhouette Scores of 0.33 for K-Means, 0.43 for Fuzzy C-Means, and 0.49 for the proposed approach. In addition, the 2025 framework study reported a K-Means Silhouette Score of 0.432619.

However, these reported values should not be treated as a direct ranking of the different approaches because each study used different datasets, features, preprocessing procedures, and evaluation measures. The comparison suggests that alternative clustering techniques and broader behavioural features can provide additional opportunities for improving customer segmentation. It also highlights the need for a common dataset and consistent evaluation framework to make a fair comparison between different clustering approaches.

VI. DISCUSSION

The reviewed studies indicate that clustering techniques can effectively identify groups of customers with similar behavioural characteristics.



K-Means remains a popular approach because of its simplicity, speed, and ease of interpretation. At the same time, advanced methods such as Fuzzy C-Means, Deep Embedded Clustering, and Deep Learning provide additional approaches for customer segmentation.

Customer segmentation becomes more useful when combined with techniques such as recommendation systems, Customer Lifetime Value analysis, and Explainable AI. The main objective is not only to create customer groups but also to use these groups to support better and more personalized marketing decisions.

VII. CONCLUSION

Customer segmentation helps businesses understand different types of customers. K-Means is widely used because of its simplicity and ease of interpretation. Advanced techniques such as Fuzzy C-Means and Deep Learning provide additional approaches for improving customer segmentation. The comparison also highlights the need for consistent evaluation and broader customer features in customer segmentation research. Effective customer segmentation can further support personalized marketing strategies by enabling businesses to design more targeted and relevant customer interactions.

VIII. FUTURE SCOPE

Future work can focus on evaluating different clustering techniques using a common customer dataset and standardized preprocessing methods. Additional behavioural, demographic, and transaction-based features can also be incorporated to improve customer segmentation. Advanced approaches such as deep learning, explainable AI, and hybrid clustering techniques can be explored for more accurate and interpretable customer segments. Future studies can also connect identified customer segments with real-time recommendation systems and personalized marketing strategies to support better customer engagement and decision-making.

REFERENCES

- [1] R. S. Manikandan, S. K. Sundar, J. J. Amalraj, P. Selvam, and M. Faheem, "Deep learning-driven customer segmentation model for online retail industry based on extended RFM model," *Discover Computing*, vol. 29, article 507, 2026. doi: 10.1007/s10791-026-10421-5.
- [2] Y. R. Rahadian, D. Handayani, D. A. F. Ferlania, et al., "Customer segmentation based on RFM, deep embedded clustering, and lifetime value analysis for loyalty strategy optimization," *Procedia Computer Science*, vol. 284, pp. 1611–1618, 2026. doi: 10.1016/j.procs.2026.07.154.
- [3] P. S. Hasugian, A. Simangunsong, M. Manik, and H. Tarihoran, "Customer segmentation using K-Means clustering algorithm for digital marketing strategy," *Jurnal Armada Informatika*, vol. 10, no. 1, 2026. doi: 10.36520/jai.v10i1.320.
- [4] F. A. Alijoyo, T. S. A. Aziz, N. Omer, N. Yusuf, M. D. Kumar, A. V. N. Ramesh, Z. Ulmas, and Y. A. Baker El-Ebiary, "Personalized marketing: Leveraging AI for culturally aware segmentation and targeting," *Alexandria Engineering Journal*, vol. 119, pp. 8–21, 2025. doi: 10.1016/j.aej.2025.01.074.
- [5] A. Ashraf, C. A. Rayed, N. A. Awadallah, and H. M. Sabry, "A framework for customer segmentation to improve marketing strategies using machine learning," *Procedia Computer Science*, vol. 260, pp. 616–625, 2025. doi: 10.1016/j.procs.2025.03.240.
- [6] P. Perapu, "Customer segmentation using K-Means clustering for personalized marketing campaigns," *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, vol. 11, no. 3, pp. 810–815, 2025. doi: 10.32628/CSEIT25113344.
- [7] G. Wang, "Customer segmentation in the digital marketing using a Q-learning based differential evolution algorithm integrated with K-means clustering," *PLOS ONE*, vol. 20, no. 2, e0318519, 2025. doi: 10.1371/journal.pone.0318519.
- [8] W. Gan, P. Zhou, S. Wan, J. Zeng, and Z. Qi, "RFMI-based customer segmentation with K-means," in *2024 IEEE International Conference on Big Data (BigData)*, pp. 8286–8295, 2024. doi: 10.1109/BigData62323.2024.10825251.



- [9] A. J. Christy, A. Umamakeswari, L. Priyatharsini, and A. Neyaa, "RFM ranking – An effective approach to customer segmentation," *Journal of King Saud University - Computer and Information Sciences*, vol. 33, no. 10, pp. 1251–1257, 2021. doi: 10.1016/j.jksuci.2018.09.004.
- [10] K. Gayathri and R. Arunodhaya, "Customer segmentation and personalized marketing using K-Means and APRIORI algorithm," in *Proceedings of the First International Conference on Combinatorial and Optimization (ICCOP 2021)*, 2021. doi: 10.4108/eai.7-12-2021.2314561.

