Mambilla Journal of Science and the Environment Vol. 9, Issue 3, Page No. 36-40

MAMBILLA JOURNAL OF SCIENCE AND THE ENVIRONMENT

journal homepage: https://oer.tsuniversity.edu.ng/index.php/mjse

HYBRID DBSCAN-GMM CLUSTERING MODEL FOR EFFECTIVE CUSTOMER SEGMENTATION: A TECHNICAL APPROACH FOR PERSONALIZED MARKETING STRATEGY

[1*]Marafa, J., [2]Garba, E.J.

[1]Department of Computer Science, Taraba State University, Jalingo, Nigeria.
[2] Department of Computer Science, Modibbo Adama University, Yola, Nigeria.

ARTICLE INFO

Article history:
Received 27 September 2025
Received in revised form 13 October 2025
Accepted 22 October 2025

Keywords:
Customer segmentation, personalized marketing, DBSCAN, Gaussian Mixture Model, hybrid clustering.

ABSTRACT

Customer segmentation plays a vital role in developing personalized marketing strategies that enable organizations to target consumers more effectively. Conventional clustering methods such as K-means and Recency, Frequency, and Monetary analysis often encounter limitations when applied to complex retail data, particularly in managing noise, sparse purchasing patterns, and non-spherical clusters. This study introduces a hybrid approach that integrates Density-Based Spatial Clustering Applications with Noise and Gaussian Mixture Model to address these challenges. Density-Based Spatial Clustering Applications with Noise first identifies dense regions of customer data and separates noise, after which Gaussian Mixture Model probabilistically refines the clusters for greater accuracy and flexibility. Data were obtained from Kaggle and the UCI Machine Learning Repository and analyzed using the proposed hybrid model. The hybrid Density-Based Spatial Clustering Applications with Noise and Gaussian Mixture Model approach demonstrated a clustering accuracy, reduced misclassification, and provided better adaptability for large-scale and datasets. Findings indicate that this model supports more precise customer profiling, thereby enabling personalized marketing strategies that enhance customer engagement, loyalty, and organizational profitability.

1. Introduction

In today's highly competitive business environment, customers remain the most critical asset to any organization. Companies across retail, banking, telecommunications, and e-commerce increasingly recognize that long-term success depends on the ability to understand, predict, and respond to customer behaviors in ways that foster loyalty and enhance profitability. The shift from product-centric to customer-centric strategies has driven firms to focus on customer segmentation as a cornerstone of personalized marketing (Kotler & Armstrong, 2018).

Traditional marketing approaches that relied on broad, undifferentiated campaigns have proven increasingly ineffective in the digital age (Smith, 2020). Consumers are bombarded daily with information and advertisements, making it harder for mass-marketing messages to resonate. Businesses must now tailor their offerings to distinct consumer groups with unique needs and preferences (Thompson *et al.*, 2022). Personalized marketing, therefore, has become essential for aligning promotional efforts with individual consumer expectations, improving customer experiences, and maximizing return on investment (Gomes & Meisen, 2023).

Customer segmentation is the systematic process of dividing a customer base into groups with shared traits, such as demographics, psychographics, or purchasing behavior. This allows organizations to design strategies that are better aligned with specific consumer needs (Kilari *et al.*, 2022). For instance, segmentation helps identify high-value customers who contribute significantly to revenue, enabling businesses to direct resources toward retention strategies that drive long-term value (Sabuncu *et al.*, 2022). Beyond marketing efficiency, segmentation also facilitates innovation by revealing unmet needs and market opportunities.

The emergence of big data and machine learning has revolutionized segmentation practices. Automated techniques can process large and complex datasets that traditional analyses cannot manage, uncovering behavioral patterns that would otherwise remain hidden (Banduni & Ilavendhan, 2020). Among these, clustering algorithms have gained

* Corresponding author: +2348169493431
E-mail address: jebulomarafa@gmail.com

prominence, as they group customers based on similarities without prior labeling (Das & Nayak, 2022). However, not all clustering methods are equally effective in retail contexts. K-means, for example, is computationally efficient but assumes spherical cluster shapes, which rarely align with real-world data (Omol *et al.,* 2024). Similarly, RFM analysis offers simplicity but cannot manage dynamic or sparse datasets (Apichottanakul *et al.,* 2020).

Density-Based Spatial Clustering of Applications with Noise (DBSCAN) has been recognized as a robust alternative for its ability to detect arbitrarily shaped clusters and manage noise (Vijitha *et al.,* 2024). Nevertheless, DBSCAN does not provide soft boundaries, which are often necessary for modeling customers who display behaviors spanning multiple groups. Gaussian Mixture Models (GMM), on the other hand, excel in probabilistic modeling but struggle with noise sensitivity and initialization challenges (Li & Lee, 2024). A hybrid approach that integrates both algorithms offers a promising solution: DBSCAN's density-based classification filters noise and identifies cluster boundaries, while GMM refines the results by probabilistically assigning memberships.

This paper investigates the effectiveness of a hybrid DBSCAN-GMM model for customer segmentation in retail. By combining density-based detection with probabilistic refinement, the study demonstrates how this approach enhances segmentation accuracy, reduces misclassification, and supports the development of more adaptive personalized marketing strategies.

2. Literature Review

2.1 Customer Segmentation and Marketing Effectiveness

The success of modern businesses increasingly depends on their ability to adopt customer-centric marketing strategies. Effective segmentation enables organizations to target specific groups with tailored campaigns, improving customer experiences and enhancing profitability (Miller & Lee, 2023). Research shows that segmentation contributes not only to sales growth but also to resource optimization, allowing firms to concentrate marketing budgets where they yield the highest returns (Gonzalez *et al.,* 2022).

In digital retail, segmentation plays a pivotal role in enabling personalization. Gomes and Meisen (2023) highlight that online shopping environments thrive on customized experiences, where consumer loyalty is tied to how well a business understands and meets individual needs. Similarly, Osakwe *et al.,* (2023) stress that data-driven segmentation allows firms to move beyond generic marketing toward predictive, personalized strategies that sustain long-term relationships.

2.2 Traditional Methods and Their Shortcomings

K-means clustering has been a dominant technique due to its efficiency and ease of implementation (Kilari *et al.,* 2022). Yet, its reliance on predetermined cluster counts and sensitivity to cluster shape limit its effectiveness in complex, non-linear datasets (Mathur & Sihare, 2019). RFM analysis has also been widely applied to measure customer value through recency, frequency, and monetary metrics (Sabuncu *et al.,* 2020). While useful, this method oversimplifies customer behavior and fails to capture evolving patterns or noise in large datasets.

Several studies illustrate these shortcomings. Yıldız *et al.,* (2023) applied K-means with RFM in fashion retail but noted inaccuracies in capturing irregular purchase behaviors. Apichottanakul *et al.,* (2020) observed similar limitations when applying RFM in the pork industry, finding that sparse datasets led to poor segmentation accuracy. These findings suggest that while traditional models remain useful for small-scale analyses, they are insufficient for modern large-scale retail environments.

2.3 Emerging Clustering Approaches

Density-based and probabilistic models offer stronger capabilities for handling complex data. DBSCAN, introduced by Ester et al. in 1996, identifies clusters based on density connectivity and distinguishes noise points (Sajana et al., 2016). This makes it effective in detecting irregularly shaped clusters. However, DBSCAN struggles with varying density levels, limiting its applicability in datasets with overlapping customer behaviors (Patel et al., 2020).

In contrast, GMM is grounded in probabilistic mixture modeling, enabling soft clustering where data points belong to multiple clusters with varying likelihoods (Li & Lee, 2024). Amutha and Khan (2023) demonstrated that GMM improved segmentation flexibility in financial services, allowing better personalization. Nonetheless, GMM's reliance on initialization parameters makes it prone to local optima and sensitive to outliers.

Hybrid approaches have emerged to address these challenges. Lodha and Deshmukh (2023) showed that combining multiple clustering algorithms improves segmentation accuracy by leveraging complementary strengths. Similarly, Chen et al. (2019) integrated k-means with metaheuristic optimization to enhance segmentation performance. These studies confirm the potential of hybrid methods, though the specific combination of DBSCAN and GMM remains underexplored in retail segmentation.

3. Methodology

3.1 Data Collection and Preprocessing

The datasets were drawn from Kaggle and the UCI Machine Learning Repository. Attributes included demographic variables, purchasing history (frequency, recency, and spending patterns), and behavioral preferences. Data cleaning addressed missing values, inconsistencies, and outliers. Normalization was performed to ensure uniform scaling, while feature engineering emphasized variables most relevant to segmentation.

3.2 Hybrid DBSCAN-GMM

The methodology proceeded in two phases:

1. **DBSCAN Clustering**

   (i). Input parameters: ε (neighborhood radius) and MinPts (minimum points).
   (ii). Classified data into core, border, and noise points.
   (iii). Generated initial clusters of arbitrary shapes.

2. **GMM Refinement**

   (i). DBSCAN clusters were passed to a Gaussian Mixture Model for probabilistic refinement.
   (ii). Applied the Expectation-Maximization (EM) algorithm to estimate parameters (means, covariance, and mixture weights).
   (iii). Provided soft assignments, reducing misclassification and incorporating overlapping behaviors.

4. Results and Discussion

4.1 Hybrid DBSCAN-GMM

The hybrid DBSCAN-GMM model (Figure 2) improves customer segmentation by combining DBSCAN's ability to detect dense clusters and outliers (Figure 1) with GMM's refinement of cluster boundaries. While DBSCAN alone produced one dominant cluster and scattered noise (Figure 1), the hybrid redistributed outliers into four well-defined groups, yielding more balanced and accurate segmentation (Figure 2). This approach captures underlying patterns, enhances cluster separation, and is particularly effective for complex datasets with overlapping characteristics or noise.

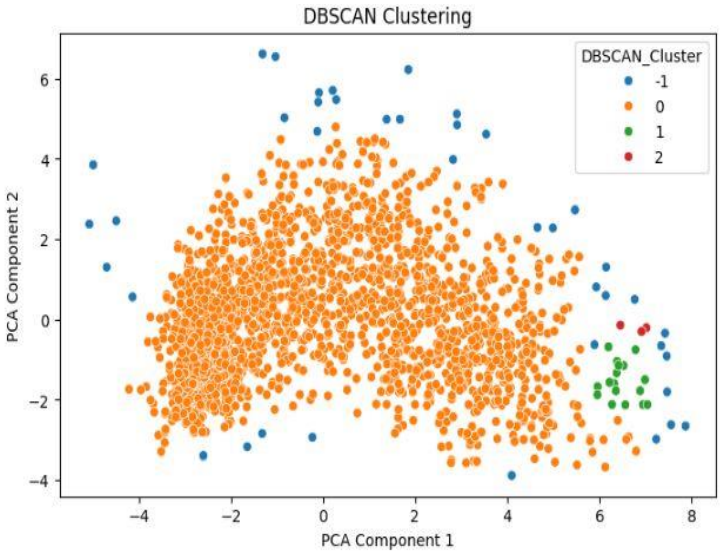

Figure 1: DBSCAN Clustering

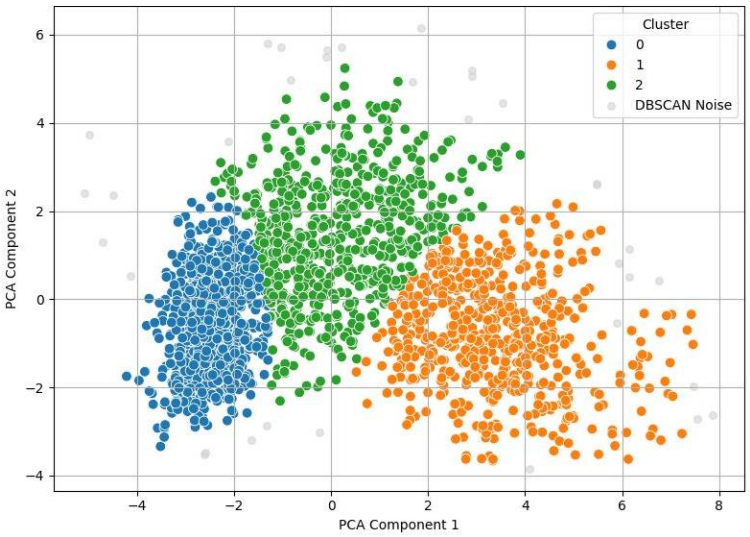

Figure 2: Hybrid DBSCAN-GMM Clustering

The results demonstrate that by combining the strengths of density-based detection with probabilistic refinement. Unlike K-means, which imposes rigid cluster boundaries, the hybrid approach adapts to complex retail data with overlapping customer behaviors. By integrating DBSCAN's robustness to noise with GMM's flexibility, the model achieves more balanced and meaningful segmentation.

From a marketing perspective, these improvements have significant implications. Businesses can use the refined clusters to design campaigns targeted at distinct customer groups, reducing resource waste and enhancing customer engagement. For example, high-value customers identified in dense clusters can receive loyalty rewards, while occasional buyers can be targeted with promotional offers.

The scalability of the hybrid approach also positions it as a valuable tool in industries beyond retail, such as healthcare for patient profiling or finance for risk-based segmentation.

5. Conclusion

This study has shown that the hybrid DBSCAN-GMM model offers a superior alternative to traditional clustering techniques for customer segmentation. By leveraging DBSCAN's noise handling and GMM's probabilistic refinement, the model delivers more accurate, flexible, and scalable segmentation.

The practical implications are significant: businesses can adopt this approach to enhance personalized marketing, optimize resource allocation, and improve customer loyalty. Theoretically, the study contributes to machine learning literature by demonstrating the viability of hybrid clustering for segmentation. Future research should explore real-time implementations, integration with deep learning, and applications across diverse industries.

References

Apichottanakul, C., Kittisak, S., & Akaraphanth, L. (2020). Hybrid customer segmentation using RFM and probabilistic models. *Journal of Business Research*, 115, 482–495.

Banduni, A., & Ilavendhan, S. (2020). Machine learning approaches for customer segmentation in retail. *International Journal of Data Mining & Emerging Technologies*, 10(2), 45–54.

Bilgic, E., Ozden, O., & Yildirim, R. (2021). Retail analytics for store segmentation using purchasing behavior. *Procedia Computer Science*, 196, 672–680.

Carmichael, T. (2024). Customer segmentation and business performance: A statistical perspective. *Journal of Marketing Analytics*, 12(1), 77–89.

Das, S., & Nayak, R. (2022). Machine learning for customer segmentation: A retail perspective. *Expert Systems with Applications*, 188, 116014.

Deng, Y. (2023). Customer relationship management and brand loyalty. *Journal of Business and Economics*, 18(3), 155–167.

Gayathri, R., & Arunodhaya, R. (2021). Enhancing retail marketing strategies with customer segmentation. *International Journal of Retail & Distribution Management*, 49(8), 1012–1026.

Gomes, F., & Meisen, R. (2023). Personalized targeting in online retail: A review. *Electronic Commerce Research*, 23(2), 245–263.

Gonzalez, J., Silva, P., & Almeida, R. (2022). Data-driven segmentation and marketing effectiveness. *Journal of Retail Analytics*, 14(2), 93–110.

Katragadda, S. (2022). Real-time customer segmentation using machine learning. *International Journal of Advanced Computer Science*, 13(5), 331–341.

Kilari, S., Devi, L., & Reddy, R. (2022). Customer segmentation using machine learning approaches. *International Journal of Business Analytics*, 9(3), 19–33.

Kotler, P., & Armstrong, G. (2018). Principles of marketing (17th ed.). Pearson.

Li, Y., & Lee, H. (2024). Big data clustering and segmentation using Gaussian mixture models. *Journal of Big Data Analytics*, 11(1), 1–15.

Lodha, S., & Deshmukh, P. (2023). Hybrid clustering models for enhanced segmentation. *International Journal of Computer Applications*, 182(10), 42–51.

Mathur, S., & Sihare, N. (2019). Data mining approaches for customer behavior segmentation. *International Journal of Information Technology*, 11(2), 213–220.

Miller, T., & Lee, H. (2023). Data-driven customer segmentation and its impact on marketing strategies. *Journal of Business Intelligence*, 7(2), 145–161.

Omol, E., Wanjiru, J., & Mwangi, P. (2024). Customer segmentation in Kenyan grocery retail using K-means clustering. *African Journal of Business and Management*, 18(1), 29–41.

Osakwe, C., Bello, T., & Ifeanyi, O. (2023). Predictive analytics in digital marketing. *Journal of Marketing Research*, 60(4), 765–782.

Patel, A., Kumar, R., & Shah, D. (2020). Scalability in clustering algorithms for retail analytics. *International Journal of Machine Learning and Applications*, 7(4), 211–220.

Sabuncu, M., Yildirim, S., & Ceylan, T. (2022). Customer profiling through segmentation in retail banking. *Journal of Financial Services Marketing*, 27(3), 185–198.

Shuaib, M. (2015). Market segmentation practices in Nigeria's education sector. *Journal of African Business Studies*, 16(2), 177–193.

Smith, A. (2020). The decline of mass marketing: Lessons for the digital age. Marketing Review, 20(3), 215–228.

Thompson, J., Miller, D., & Singh, P. (2022). Digital marketing and the role of segmentation. *Journal of Marketing Strategy*, 54(6), 921–939.

Verdenhof, P., & Tambovceva, T. (2019). Predictive modeling in customer segmentation. Technological Forecasting and Social Change, 146, 435–444.

Vijitha, V., Ramesh, S., & Prakash, P. (2024). Density-based clustering for customer segmentation using Python. *International Journal of Information Management Data Insights*, 4(1), 100187.

Wang, Z., Chen, H., & Li, J. (2016). Biclustering methods for market segmentation. *Knowledge-Based Systems*, 103, 1–12.

Xie, J., Yu, S., & Wang, Q. (2019). Improving K-means clustering with Firefly algorithms. *Applied Soft Computing*, 84, 105729.

Yıldız, K., Aksoy, D., & Demir, M. (2023). Personalized product recommendations using RFM and clustering. *Journal of Retail Technology*, 8(2), 77–90.