JIKO (Jurnal Informatika dan Komputer) Accredited KEMDIKTISAINTEK, No.0173/C3/DT.05.00/2025
Vol. 9, No. 1, April 2026, pp. 1-9 p-ISSN: 2614-8897
DOI: 10.33387/jiko.v9i1.10859 e-ISSN: 2656-1948

PERFORMANCE EVALUATION OF HYBRID CLUSTERING K-MEANS AND DBSCAN WITH FEATURE WEIGHT OPTIMIZATION

Vic Devlin[1], Robet[2*], Octara Pribadi[3]

[1,2,3] Teknik Informatika, STMIK TIME, Medan, 20234, Indonesia
Email: [1]vicdevlin2004@gmail.com, [*2]robertdetime@gmail.com, [3]octarapribadi@gmail.com

(Received: 14 October 2025, Revised: 30 October 2025, Accepted: 17 December 2025)

Abstract

This study evaluates the performance of a hybrid clustering model that integrates K-Means and DBSCAN, enhanced through Feature Weight Optimization (FWO) using a Genetic Algorithm (GA), to achieve more accurate consumer data segmentation. Two benchmark datasets, Customer Personality Analysis (CPA) and Online Retail (OR), were employed to assess the adaptability of clustering techniques to various data structures. The feature weighting process was optimized using GA due to its strong global search capability and ability to avoid local minima, enabling the identification of optimal feature contributions that improve cluster discrimination. In this optimization, GA dynamically adjusts feature weights through iterative selection, crossover, and mutation operations, allowing the clustering model to assign appropriate importance to each variable. The Silhouette Score was utilized as the main evaluation metric to measure intra-cluster cohesion and inter-cluster separation. Experimental findings show that for the CPA dataset, the Hybrid + FWO model achieved the highest performance with a Silhouette Score of 0.9600, while the K-Means + FWO model obtained the best score of 0.9804 on the OR dataset. Across all scenarios, the inclusion of FWO consistently enhanced clustering stability and interpretability. These results indicate that applying GA-based feature optimization not only strengthens the robustness of hybrid clustering but also ensures more meaningful and actionable segmentation insights in consumer behavior analytics.

Keywords: *K-Means, DBSCAN, Hybrid Clustering, Feature Weight Optimization, Genetic Algorithm*

**Corresponding Author: Robet*

1. INTRODUCTION

The acceleration of digital transformation over the past two decades has fundamentally changed how organizations manage information and make strategic decisions. Data is now regarded as a valuable asset that can provide a competitive advantage when systematically managed and analytically processed [1]. In the modern business landscape, particularly in the retail and e-commerce sectors, customer data utilization has become crucial for understanding consumer behavior, identifying market needs, and developing targeted marketing strategies. One effective analytical approach for leveraging such data is customer segmentation, which involves dividing consumers into groups based on similarities in attributes, transactional behavior, or preference patterns. Through segmentation, organizations can prioritize marketing strategies, deliver targeted promotions, and enhance customer loyalty [2]. Data-driven segmentation strategies have also proven effective in optimizing customer retention and minimizing marketing campaign costs [1], [2].

In the realm of unsupervised learning, clustering has emerged as one of the most widely used techniques for identifying hidden patterns within unlabeled data. Among the various algorithms, K-Means clustering occupies a significant position due to its ability to efficiently partition data based on Euclidean distances from centroid positions [3]. Its main advantages include computational efficiency and ease of implementation across analytical platforms. However, K-Means also exhibits notable limitations, such as sensitivity to initial centroid placement, inability to detect clusters of arbitrary shapes, and poor performance in handling outliers [4]. To overcome these drawbacks, the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm was developed. DBSCAN groups data points based on density, offering the advantage of not requiring a predefined number of clusters and automatically detecting noise [5]. Recent studies have introduced

variants such as Adaptive Multi-Density DBSCAN (AMD-DBSCAN), which adapts to varying data densities, particularly in large-scale datasets [4].

As a further advancement, several studies have proposed hybrid clustering approaches that combine two or more algorithms to enhance the accuracy and stability of segmentation results. One of the most commonly adopted combinations is K-Means and DBSCAN, as these two algorithms complement each other: K-Means excels in computational efficiency, while DBSCAN performs better in identifying irregularly shaped clusters and handling noise [2], [5]. The hybrid approach produces more consistent and adaptive outcomes than individual methods, making it especially valuable in consumer segmentation tasks such as identifying loyal, potential, and inactive customers [1], [2]. However, existing studies rarely address how feature weighting and hybrid clustering can be integrated to improve cluster quality in large and heterogeneous consumer datasets. This gap highlights the need for a systematic approach that jointly optimizes clustering structure and feature contribution to achieve higher segmentation accuracy.

Beyond algorithm selection, another factor that significantly affects clustering performance is the relative contribution of each feature within the dataset. Unequal feature weights can degrade clustering accuracy by disproportionately influencing the formation of certain centroids. To mitigate this, Feature Weight Optimization (FWO) techniques are applied to adjust the weight of each attribute according to its significance within the data structure [6]. A widely adopted approach for this task is the Genetic Algorithm (GA), an optimization method inspired by the process of natural evolution that uses selection, mutation, and crossover to find near-optimal solutions. GA is known for its ability to explore large solution spaces and identify optimal feature weight combinations without assuming linearity [7]. Integrating GA into clustering has been shown to improve performance, as evidenced by higher evaluation metrics such as the Silhouette Score, which measures intra-cluster cohesion and inter-cluster separation [8]. Therefore, the use of GA for feature weight optimization represents an effective means of enhancing clustering accuracy, particularly for high-dimensional datasets.

Clustering performance is typically evaluated using internal metrics such as the Silhouette Score, Calinski–Harabasz Index, and Davies–Bouldin Index. Among these, the Silhouette Score is the most widely adopted due to its intuitive representation of cluster quality, ranging from –1 to 1 [9]. A score close to 1 indicates that data points are well assigned to their clusters with strong separation between them. In large-scale data analysis, the Silhouette Score has been adapted to distributed computing environments such as Hadoop and Spark for efficient evaluation [9]. Principal Component Analysis (PCA) is also often employed to visualize clustering results, facilitating easier interpretation. Therefore, this study aims to evaluate the performance of a hybrid clustering approach that combines K-Means and DBSCAN with Feature Weight Optimization based on a Genetic Algorithm. The evaluation is conducted using the Silhouette Score as the primary metric and PCA for visual validation. The findings are expected to contribute to the development of more accurate, adaptive, and efficient customer segmentation methods for large-scale data environments.

2. RESEARCH METHOD

This research was designed using a quantitative experimental approach aimed at analyzing and evaluating the performance of a hybrid clustering algorithm. The proposed method integrates two widely used clustering techniques, K-Means and DBSCAN, which are subsequently enhanced through Feature Weight Optimization (FWO) using a Genetic Algorithm (GA) to enhance segmentation precision and clustering robustness. The overall workflow of this study includes problem identification, data collection, preprocessing, clustering implementation, feature weight optimization, and model evaluation.

Each stage of the research was structured according to the methodological frameworks proposed by Sridevi and Rajanna [5] as well as Kouser et al. [6]. These studies demonstrated that integrating partition-based methods (such as K-Means) with density-based approaches (such as DBSCAN) can yield more stable and adaptive cluster structures, especially when dealing with datasets characterized by varying data densities and heterogeneous feature distributions.

The research process is illustrated as a flow diagram shown in Figure 1, which presents the sequential relationship among the major stages ranging from data input and preprocessing to clustering, feature weight optimization, and model evaluation.

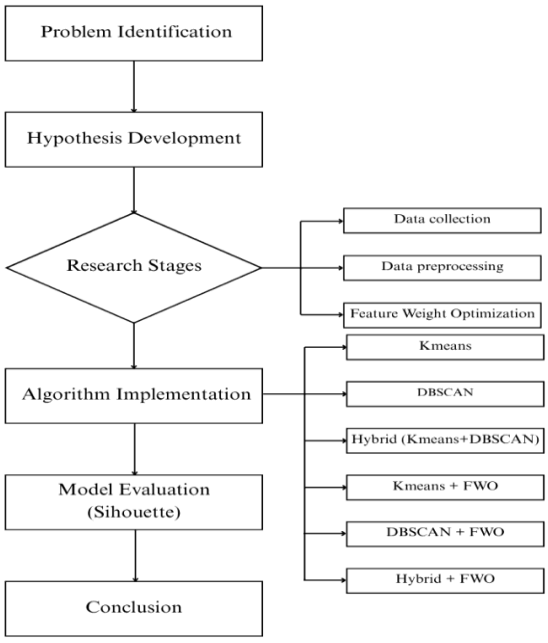

Figure 1. Research Flowchart

2.1 Dataset Description

This study utilized two publicly available datasets relevant to customer segmentation experiments, namely the Marketing Campaign Dataset and the Online Retail Dataset. These datasets were selected to represent two distinct but complementary consumer contexts behavioral marketing and transactional purchasing patterns thereby providing a valid foundation for cross-domain evaluation [2], [10].

The Marketing Campaign Dataset comprises 2,240 customer records with a total of 29 attributes, including demographic variables (e.g., Year_Birth, Education, Marital_Status), economic indicators (Income), and behavioral attributes such as MntWines, MntMeatProducts, and NumWebPurchases. This dataset was selected because it reflects real-world consumer marketing behavior, making it suitable for behavior-based segmentation analysis. The data is publicly accessible on the Kaggle platform [1].

The Online Retail Dataset consists of e-commerce transaction records from a UK-based retail company, containing 541,909 entries and eight primary attributes InvoiceNo, StockCode, Quantity, InvoiceDate, UnitPrice, CustomerID, and Country. Since the data is transactional, aggregation was performed based on CustomerID to construct unique customer profiles. The aggregated features include: TotalQuantity (Total number of items purchased per customer), Frequency (Total number of unique transactions (InvoiceNo)), TotalSpent (Total expenditure of each customer), calculated using Equation (1).

$$TotalSpent = \sum_{i=1}^{n} (Quantity_i \times UnitPrice_i) \quad (1)$$

Both datasets were chosen because they represent two distinct perspectives of consumer behavior marketing engagement and transactional purchasing enabling a comprehensive evaluation of the proposed hybrid clustering model across different data characteristics [2], [10].

2.2 Data Preprocessing

The preprocessing stage is conducted to ensure the integrity, consistency, and analytical readiness of the datasets before applying clustering algorithms. This stage consists of three main steps: data cleaning, categorical transformation, and feature normalization.

Data Cleaning, Records containing missing values or duplicate entries are removed to eliminate bias and distortion during analysis. This process ensures that all retained records are valid and representative of the dataset's underlying distribution.

Categorical Attribute Transformation, Categorical variables are converted into numerical representations using two techniques: Label Encoding for ordinal variables and One-Hot Encoding for nominal variables. This transformation enables clustering algorithms to process data numerically while preserving categorical distinctions [11].

Feature Normalization, All numerical attributes are normalized within the range [0, 1] using the Min–Max Scaling technique. This step prevents attributes with larger magnitudes from dominating distance calculations during clustering and ensures that each feature contributes proportionally to the clustering process.

2.3 Clustering Method Implementation

K-Means Algorithm, the K-Means algorithm **is** employed to generate the initial customer segmentation by minimizing the Euclidean distance between data points and their respective centroids. The process begins by defining the number of clusters (k), followed by iterative centroid updates until the Sum of Squared Errors (SSE) objective function reaches a minimum value [3], [12]. For both datasets, the number of clusters is set to two (n_clusters = 2), aligning with the initial assumption of binary customer segmentation. The parameter random_state = 42 is applied to ensure reproducibility and consistency across multiple runs. This configuration serves as a balanced baseline for comparative evaluation with other clustering approaches.

DBSCAN Algorithm, the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm **is** utilized to identify cluster structures based on data density. Two key parameters are configured: Epsilon (ε), representing the neighborhood radius, and min_samples (MinPts), indicating the minimum number of points required to form a dense region [4]. The parameter tuning for DBSCAN is conducted separately for each dataset through exploratory testing, seeking combinations that achieve the highest Silhouette Score with a valid number of clusters (excluding the noise label "–1").

Online Retail Dataset: The optimal configuration is found at (eps = 1.0, min_samples = 3), producing five clusters (including noise). Although this configuration achieves a Silhouette Score of 0.9400, the score is partially affected by the presence of outliers labeled as noise. Nevertheless, the visual distribution of the resulting clusters indicates representative separability, validating the chosen parameters.

Customer Personality Analysis (CPA) Dataset: The optimal parameters are determined as (eps = 0.7, min_samples = 10), generating three valid clusters (including noise). This configuration effectively separates customer groups based on income and total expenditure, despite the existence of minor noise points. These parameter selections provide an optimal balance between inter-cluster separation and noise reduction, aligning with the recommendations of recent clustering optimization studies [13], [14].

Hybrid Clustering (K-Means + DBSCAN), the hybrid clustering approach is designed to combine the complementary strengths of K-Means and DBSCAN, leveraging DBSCAN's ability to detect noise and K-Means' efficiency in handling spherical cluster

structures. The implementation procedure comprises the following steps:

Noise Relabeling: Data points labeled "–1" (noise) by DBSCAN are reassigned cluster labels predicted by K-Means, Final Labeling: The final cluster labels are derived from the DBSCAN output, with revised labels applied to former noise points as determined by K-Means, Stability Enhancement: This integration reduces the adverse effects of excessive noise while maintaining the structural consistency of cluster formation. Both datasets apply the same parameter configurations as defined in the respective K-Means and DBSCAN implementations. The hybrid strategy improves cluster stability and minimizes information loss caused by noise overrepresentation [15], [16].

2.4 Feature Weight Optimization (FWO) using Genetic Algorithm (GA)

The adjustment of feature weights in this study **is** performed through a Genetic Algorithm (GA), which is widely recognized for its ability to explore complex and high-dimensional search spaces while avoiding local optima. GA represents an evolutionary optimization technique inspired by the process of natural selection, where a population of candidate solutions evolves iteratively through genetic operators such as selection, crossover, and mutation [6]. This approach proves effective for feature-weighting tasks, as it allows dynamic adjustment of each feature's contribution to improve clustering performance.

The primary objective of this optimization process **is** to adaptively tune the relative contribution of each feature to the final clustering structure. Each individual in the GA population is represented as a weight vector whose dimension corresponds to the number of features used in the dataset. The overall procedure consisted of several key phases, as illustrated in Figure 2, which visualizes the complete workflow of the Genetic Algorithm based Feature Weight Optimization process.

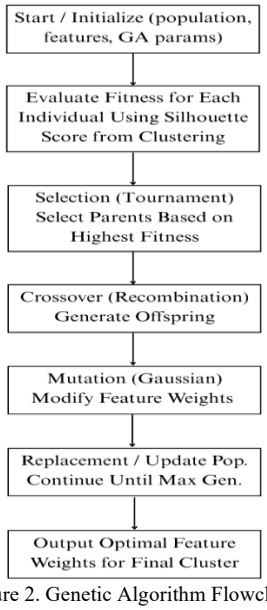

Figure 2. Genetic Algorithm Flowchart

These stages include the initialization of the population, evaluation of individual fitness values using the Silhouette Score, selection of the best individuals, genetic operations such as crossover and mutation, and population updates through evolutionary iteration. The detailed steps are described as follows:

Initialization, an initial population of 20 individuals is generated, where each individual contains randomly assigned weights within the range [0.1, 1.0]. These values serve as the initial solution set to be evolved through successive generations.

Fitness Evaluation each individual (i.e., weight vector w) is evaluated by computing the Silhouette Score obtained from the clustering results. The fitness function is defined as:

$$f(w) = Silhouette\,(X \times w, labels) \tag{2}$$

where X denotes the normalized dataset, w represents the weight vector applied to each feature, and labels refer to the cluster assignments produced by K-Means, DBSCAN, or the Hybrid model.

The fitness function evaluates how well the clustering structure achieves both intra-cluster cohesion and inter-cluster separation. A higher fitness value indicates better clustering quality and more optimal feature weighting.

Selection, the Tournament Selection method is applied to identify individuals with superior fitness values. The best-performing individuals are chosen to propagate to the next generation, ensuring the survival of optimal solutions while maintaining population diversity.

Mutation, to introduce variability and prevent premature convergence, random perturbations are applied to individual weights using Gaussian Mutation, defined as:

$$w_i' = w_i + \mathcal{N}(0, \sigma^2) \tag{3}$$

where w_i is the original feature weight, w_i' is the mutated weight, and $N(0, \sigma^2)$ represents a Gaussian random variable with mean 0 and variance σ^2. This mechanism allows small random adjustments in the feature space, promoting exploration and maintaining diversity within the population.

Crossover, for the Online Retail dataset, a Two-Point Crossover strategy is implemented to enhance solution variety. Two parent individuals exchange partial segments of their chromosomes between crossover points $\mathcal{P}_1$ and $\mathcal{P}_2$, producing new offspring. The crossover function is defined as:

$$w_i' = \begin{cases} w_i^{(1)}, & i \in [\mathcal{P}_1, \mathcal{P}_2] \\ w_i^{(2)}, & otherwise \end{cases} \tag{4}$$

where $w_i^{(1)}$ and $w_i^{(2)}$ are the feature weights of two parent individuals, while $\mathcal{P}_1$ and $\mathcal{P}_2$ denote crossover points. This mechanism is selectively

applied to the Online Retail dataset as it improves feature diversity and exploration capability, particularly for transactional data structures.

Evolutionary Iteration, the GA evolutionary cycle consisting of selection, mutation, crossover, and fitness evaluation is executed iteratively across multiple generations until convergence or until the optimal weight configuration maximizing the fitness function is achieved. The parameter settings used in this study are as follows: population size = 20, generations = 10, crossover probability (cxpb) = 0.5, mutation probability (mutpb) = 0.2, mutation type = Gaussian Mutation, and selection = Tournament Selection (size = 3).

This optimization process aims to maximize the Silhouette-based objective function, enabling the model to dynamically adapt feature contributions for improved inter-cluster separation and intra-cluster cohesion. The results confirm the effectiveness of GA in enhancing clustering performance, particularly for high-dimensional datasets [6], [17].

2.5 Experimental Design of Clustering Methods

This study evaluates six clustering scenarios divided into two main categories: methods without feature weight optimization and methods with Feature Weight Optimization (FWO) based on the Genetic Algorithm (GA). Each clustering technique is tested both independently and in a hybrid configuration, as summarized in Table 1.

Table 1. Experimental Scenarios

Clustering Method	Feature Weight Optimization (FWO)
K-Means	Not Applied
DBSCAN	Not Applied
Hybrid	Not Applied
K-Means	Applied
DBSCAN	Applied
Hybrid	Applied

All experiments are evaluated using the Silhouette Score metric, which measures intra-cluster cohesion and inter-cluster separation [9]. To complement the quantitative evaluation, the clustering results are visualized using Principal Component Analysis (PCA), which projects the multi-dimensional data into two dimensions. This visualization provides a clearer understanding of the spatial distribution of clusters, aiding in the interpretation of how effectively each method separates distinct data groups [7], [18].

3. RESULT AND DISCUSSION

This section presents the experimental results obtained from three clustering approaches namely K-Means, DBSCAN, and the Hybrid method (K-Means + DBSCAN) evaluated both without and with Feature Weight Optimization (FWO) using the Genetic Algorithm (GA). The performance of each method is assessed using the Silhouette Score, while Principal Component Analysis (PCA) is employed to illustrate the spatial distribution and cohesion of clusters.

3. 1 Customer Personality Analysis (CPA) Dataset

The experimental results for the Customer Personality Analysis (CPA) dataset are summarized and presented in Table 2, which reports the Silhouette Scores obtained from each clustering configuration.

Table 2. Silhouette Score for CPA dataset

Clustering Method	FWO	Silhouette Score
K-Means	Not Applied	0.5951
DBSCAN	Not Applied	0.7478
Hybrid	Not Applied	0.7614
K-Means	Applied	0.9592
DBSCAN	Applied	0.8979
Hybrid	Applied	0.9599

As shown in Table 2, the implementation of Feature Weight Optimization (FWO) using the Genetic Algorithm (GA) significantly improved clustering performance across all methods. The Hybrid + FWO configuration achieved the highest Silhouette Score of 0.9599, indicating superior intra-cluster cohesion and inter-cluster separation compared to other approaches. In contrast, the baseline K-Means without FWO recorded the lowest score of 0.5951, suggesting that unweighted features contributed to less distinct cluster boundaries. These findings confirm that optimizing feature contributions using GA enhances the structural clarity and overall quality of the generated clusters.

The PCA projections for all six experimental scenarios are illustrated in Figures 3-8, demonstrating how each algorithm separates the consumer data in two-dimensional space.

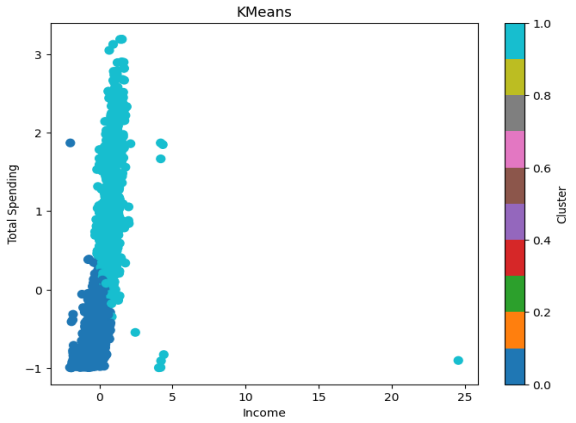

Figure 3. PCA Visualization of K-Means (CPA Dataset)

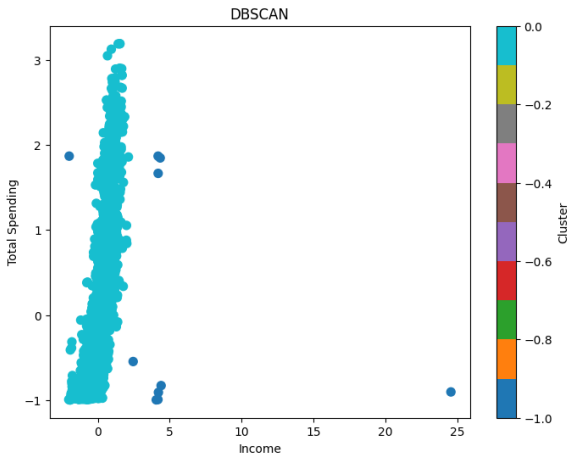

Figure 4. PCA Visualization of DBSCAN (CPA Dataset)

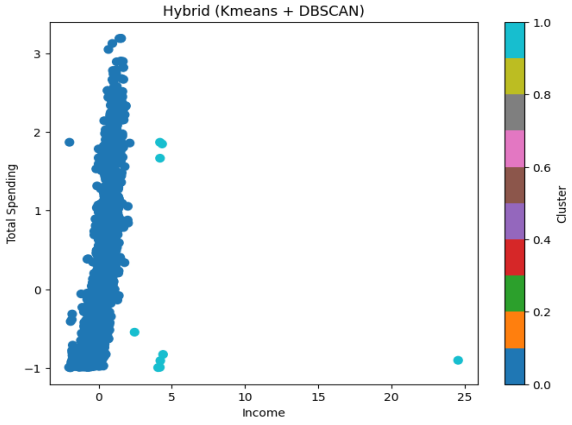

Figure 5. PCA Visualization of Hybrid Method (CPA Dataset)

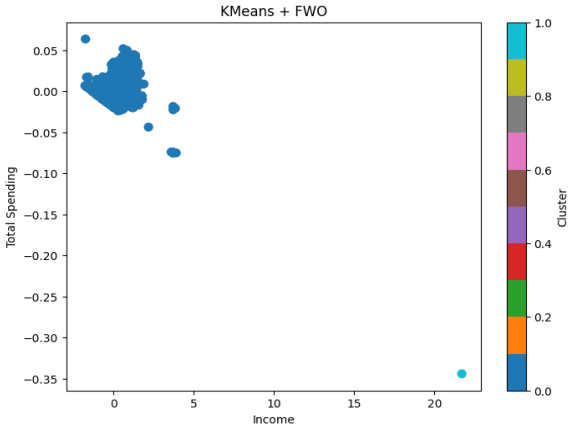

Figure 6. PCA Visualization of K-Means + FWO (CPA Dataset)

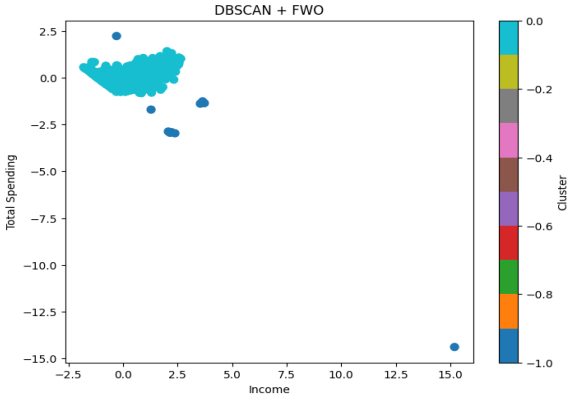

Figure 7. PCA Visualization of DBSCAN + FWO (CPA Dataset)

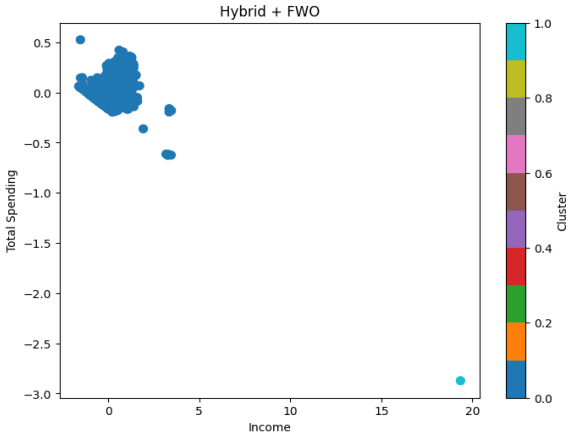

Figure 8. PCA Visualization of Hybrid Method + FWO (CPA Dataset)

The distribution in Figure 8 shows the most optimal separation, with minimal noise and well-defined cluster boundaries. This result indicates that integrating FWO significantly improves the hybrid model's ability to form compact and well-separated clusters.

3. 2 Online Retail (OR) Dataset

The experimental results for the Online Retail (OR) dataset are summarized and presented in Table 2, which reports the Silhouette Scores obtained from each clustering configuration.

Table 3. Silhouette Score for OR dataset

Clustering Method	FWO	Silhouette Score
K-Means	Not Applied	0.9801
DBSCAN	Not Applied	0.9400
Hybrid	Not Applied	0.9452
K-Means	Applied	0.9804
DBSCAN	Applied	0.9435
Hybrid	Applied	0.9521

As shown in Table 3, the K-Means algorithm achieved the highest Silhouette Score of 0.9804 after applying Feature Weight Optimization (FWO), slightly surpassing the Hybrid + FWO configuration with 0.9521. This outcome suggests that while the hybrid model benefitted from noise reduction and structural stability, the Online Retail dataset's relatively homogeneous transactional features favored the centroid-based clustering of K-Means. Consequently, the impact of FWO in the hybrid method was less pronounced than in the CPA dataset, indicating that the effectiveness of feature optimization depends on the underlying data characteristics.

The PCA projections for all six experimental scenarios are illustrated in Figures 9-14, demonstrating how each algorithm separates the consumer data in two-dimensional space.

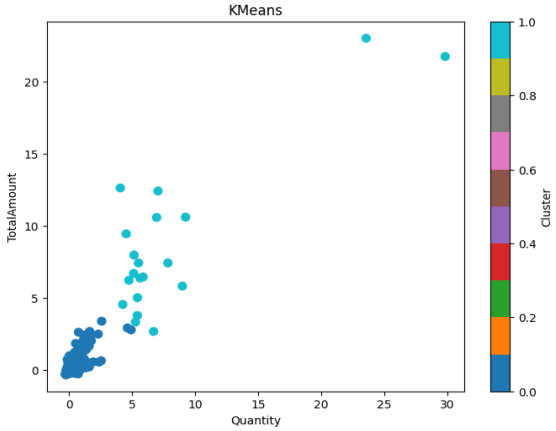

Figure 9. PCA Visualization of K-Means (OR Dataset)

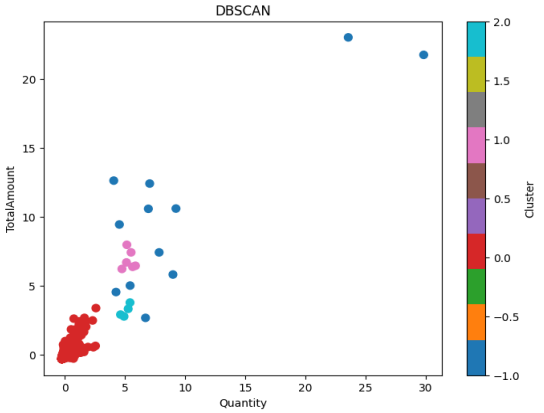

Figure 10. PCA Visualization of DBSCAN (OR Dataset)

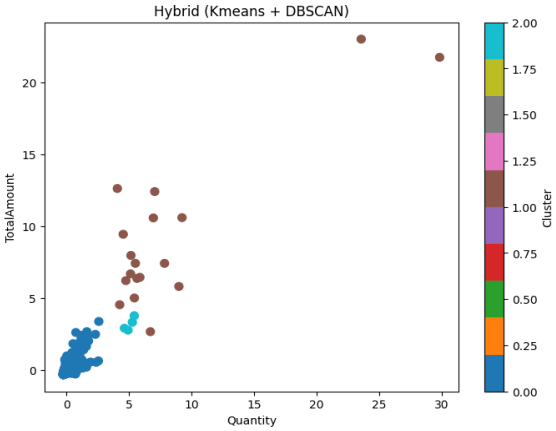

Figure 11. PCA Visualization of Hybrid Method (OR Dataset)

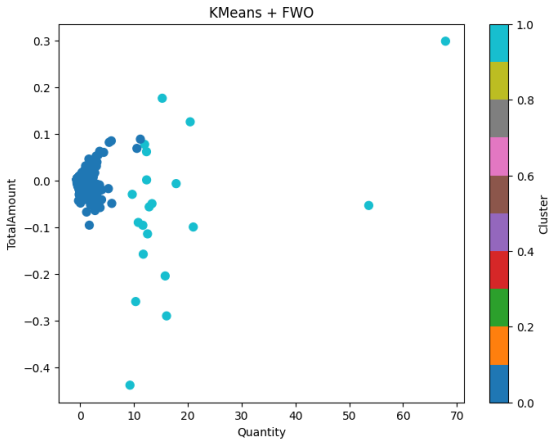

Figure 12. PCA Visualization of K-Means + FWO (OR Dataset)

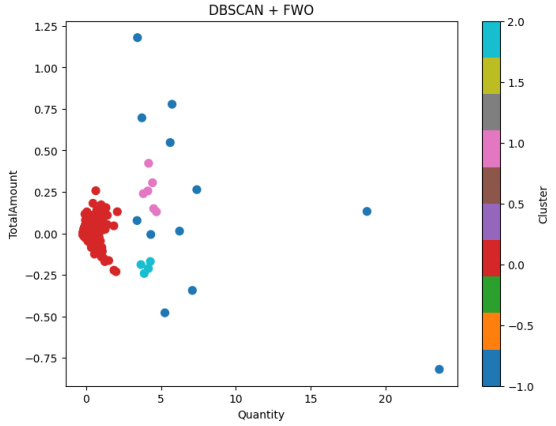

Figure 13. PCA Visualization of DBSCAN + FWO (OR Dataset)

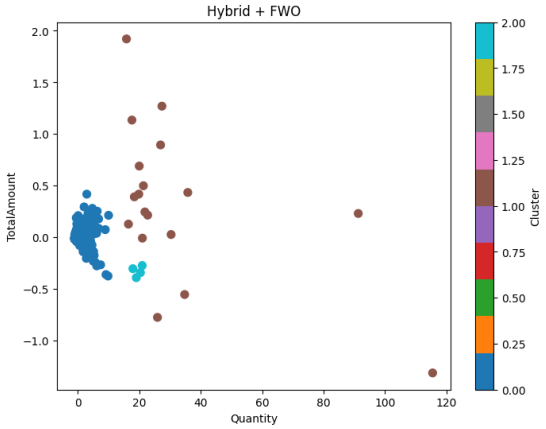

Figure 14. PCA Visualization of Hybrid Method + FWO (OR Dataset)

The plot in Figure 11 shows the most optimal separation pattern and highest cluster compactness corresponding to the K-Means + FWO configuration. Although a small number of outliers remain, the overall structure shows improved cohesion and reduced overlap among clusters after applying Feature Weight Optimization (FWO). This indicates that, for transaction-oriented datasets with a pronounced centroid-based structure, a centroid-based method augmented with adaptive feature weighting achieves superior segmentation performance.

3. 3 Discussion

The experimental results show that applying Feature Weight Optimization (FWO) using a Genetic Algorithm (GA) consistently improves clustering quality across both datasets Customer Personality Analysis (CPA) and Online Retail (OR). The Hybrid Clustering + FWO approach achieves the highest Silhouette Score on the CPA dataset (0.9599), while the K-Means + FWO method produces the best result on the OR dataset (0.9804).

These improvements indicate the adaptive capability of GA in adjusting feature weights proportionally to their importance, ensuring that each attribute contributes optimally to cluster formation. The PCA visualizations support these findings, showing clearer inter-cluster separations and stronger intra-cluster cohesion after optimization.

Compared to previous studies, this approach represents a substantial advancement. Kouser et al. [6] validated the effectiveness of GA in optimizing feature weighting, while Alsmadi et al. [19] highlighted the superiority of hybrid fuzzy clustering for high-dimensional data. The results of this study extend these findings by applying a hybrid FWO-based model to large-scale consumer segmentation tasks. Additionally, this study aligns with Salman and Fauziah [11], who emphasized the computational efficiency of K-Means for homogeneous datasets.

However, certain limitations persist most notably in the empirical selection of DBSCAN parameters (epsilon and min_samples), which remain dataset dependent and may influence cluster consistency.

Overall, the findings confirm that the integration of Hybrid Clustering and Feature Weight Optimization is an effective approach to improving segmentation accuracy in customer analytics. This study supports the hypothesis that adaptive feature weighting enhances cluster separation without sacrificing computational efficiency.

The significance of this research lies in the incorporation of evolutionary optimization principles into consumer data segmentation an area rarely explored in existing literature. Future studies could further extend this work by developing deep learning–based adaptive FWO or evolutionary hybrid clustering methods to handle real-time data within distributed computing environments such as Apache Spark or Hadoop. Such developments would further advance clustering intelligence and support automated decision-making systems in industrial and e-commerce domains.

4. CONCLUSION

This study analyzes the performance of K-Means, DBSCAN, and Hybrid Clustering algorithms combined with Feature Weight Optimization (FWO) based on a Genetic Algorithm (GA) to improve the accuracy of customer data segmentation. The experimental results show that applying FWO significantly improves clustering quality, as reflected by higher Silhouette Scores on both datasets Customer Personality Analysis (CPA) and Online Retail (OR). The Hybrid + FWO method achieves the best performance on the CPA dataset with a Silhouette Score of 0.9599, while K-Means + FWO records the highest score of 0.9804 on the OR dataset, indicating that data characteristics affect algorithmic effectiveness.

The findings confirm that feature weight optimization using GA enhances both cohesion and separation among clusters, resulting in more adaptive and stable segmentation outcomes. The results also suggest that evolutionary optimization provides a viable approach for refining clustering performance in consumer analytics. Future research should investigate the integration of deep learning–based FWO and distributed clustering frameworks to further enhance efficiency and scalability for large-scale data processing.

5. REFERENCES

[1] A. A. Rahma, A. Faqih, and A. R. Rinaldi, "Optimalisasi Strategi pemasaran melalui segmentasi pelanggan dengan analisis RFM dan algoritma K-Means untuk bisnis ritel," *JIKO (J. Inform. dan Komput.)*, vol. 9, no. 2, pp. 338-338, Jun. 2025, doi: 10.26798/jiko.v9i2.1737.

[2] S. D. K. Wardani, A. S. Ariyanto, M. Umroh, and D. Rolliawati, "Perbandingan hasil metode clustering K-Means, DBSCAN, dan hierarchical untuk analisa segmentasi pasar," *JIKO (J. Inform. dan Komput.)*, vol. 7, no. 2, pp. 191-191, Sep. 2023, doi: 10.26798/jiko.v7i2.796.

[3] Rahmati r and Wijayanto A, "Analisis cluster dengan algoritma K-Means, Fuzzy C-Means, dan hierarchical clustering," *JIKO (J. Inform. dan Komput.)*, vol. 5, no. 2, pp. 1-7, Mar. 2021.

[4] Z. Wang *et al.*, "AMD-DBSCAN: An adaptive multi-density DBSCAN for datasets of extremely variable density," *arXiv preprint arXiv:2210.08162*, 2022, doi: 10.48550/arXiv.2210.08162.

[5] K. N. Sridevi and M. Rajanna, "Hybrid clustering framework for scalable and robust query analysis: Integrating Mini-Batch K-Means with DBSCAN," *Int. J. Adv. Comput. Sci. Appl.*, vol. 16, no. 1, pp. 87–95, Jan. 2025, doi: 10.14569/IJACSA.2025.0160187.

[6] K. Kouser, A. Priyam, M. Gupta, S. Kumar, and V. Bhattacharjee, "Genetic algorithm–based optimization of clustering algorithms for the healthy aging dataset," *Appl. Sci.*, vol. 14, no. 13, p. 5530, Jul. 2024, doi: 10.3390/app14135530.

[7] G. Feng, "Feature selection algorithm based on optimized genetic algorithm and the application in high-dimensional data processing," *PLoS One*, vol. 19, no. 5, May. 2024, doi: 10.1371/journal.pone.0303088.

[8] A. G. Oskouei *et al.*, "Feature-weighted fuzzy clustering methods: An experimental review," *Neurocomputing*, vol. 619, p. 129176, Jan. 2025, doi: 10.1016/j.neucom.2024.129176.

[9] M. Gaido, "Distributed silhouette algorithm: Evaluating clustering on big data," *arXiv preprint arXiv:2303.14102*, 2023, [Online]. Available: https://arxiv.org/abs/2303.14102

[10] A. S. Paramita and T. Hariguna, "Comparison of K-Means and DBSCAN algorithms for customer segmentation in e-commerce," *J. Digit. Mark. Digit. Commer.*, vol. 1, no. 1, pp. 43–62, Mar. 2024, doi: 10.47738/jdmdc.v1i1.3.

[11] F. Salman and F. Fauziah, "Comparison analysis of K-Means and DBSCAN algorithms for improving budget absorption efficiency in EIS," *Brilliance: Res. Artif. Intell.*, vol. 3, no. 2, pp. 378–383, Jun. 2023, doi: 10.47709/brilliance.v3i2.3373.

[12] Q.-V. Doan, T. Amagasa, T.-H. Pham, T. Sato, F. Chen, and H. Kusaka, "Structural k-means (Sk-means) and clustering uncertainty evaluation framework (CUEF) for mining climate data," *Geosci Model Dev*, vol. 16, pp. 2215–2233, May. 2023, doi: 10.5194/gmd-16-2215-2023.

[13] R. Tinós, L. Zhao, F. Chicano, and D. Whitley, "NK hybrid genetic algorithm for clustering," *arXiv preprint arXiv:2402.03813*, 2024, doi: 10.48550/arXiv.2402.03813.

[14] S. Chowdhury, N. Helian, and R. de Amorim, "Feature weighting in DBSCAN using reverse nearest neighbours," *Pattern Recognit*, vol. 137,

p. 109314, Jun. 2023, doi: 10.1016/j.patcog.2023.109314.

[15] R. Mussabayev and R. Mussabayev, "Comparative analysis of optimization strategies for K-Means clustering in big data contexts: A review," *arXiv preprint arXiv:2310.09819*, 2023, doi: 10.48550/arXiv.2310.09819.

[16] T. Bezdan, Y. Zhang, and Y. Zhang, "Fruit-fly algorithm-based hybrid K-Means clustering method for text document clustering," *Math.*, vol. 9, no. 16, p. 1929, Aug. 2021, doi: 10.3390/math9161929.

[17] P. Bansal *et al.*, "GGA-MLP: A greedy genetic algorithm to optimize weights and biases in MLP," *Contrast Media Mol. Imaging*, vol. 2022, p. 4036035, Dec. 2022, doi: 10.1155/2022/4036035.

[18] V. V. Baligodugula and F. Amsaad, "Unsupervised learning: Comparative analysis of clustering techniques on high-dimensional data," *arXiv preprint arXiv:2503.23215*, 2025.

[19] M. K. Alsmadi *et al.*, "A hybrid topic modeling method based on Dirichlet multinomial mixture and fuzzy matching algorithm for short text clustering," *J. Big Data*, vol. 11, no. 68, Feb. 2024, doi: 10.1186/s40537-024-00930-9.