



Metaheuristic-based Clustering Algorithms with Principal Coordinate Analysis for Shoe Market Segmentation

Ridho Ananda¹, Samuel Sinaga², Budi Pratikno³, Nur Afrina Huda Zulkainain⁴, Tri A. Sundara⁵

^{1,2}Industrial Engineering, Telkom University, Purwokerto, Indonesia

³Universitas Jenderal Soedirman, Banyumas, Indonesia

^{4,5}The National University of Malaysia, Selangor, Malaysia

*ridhoa@telkomuniversity.ac.id

Abstract

Micro, Small, and Medium Enterprises (MSMEs) are vital to Indonesia's economy, yet many struggle with ineffective marketing strategies, as seen in Bintang Sepatu Purwokerto, an MSME shoemaker facing stagnant sales. To overcome this, this study proposed a metaheuristic-based clustering (MBC) algorithm combined with principal coordinate analysis (PCoA) for optimal customer segmentation. The developed algorithm successfully overcomes the limitations of Kmeans in mixed datasets, containing categorical and numeric data. Furthermore, the procedure for updating centroids in Kmeans that risks falling into a local optimum is also solved. In this study, six MBC algorithms were developed based on six state-of-the-art metaheuristic optimizations utilized. Then, the developed MBC algorithms are compared with benchmark algorithms, namely Kmeans and KmeansQLDE, based on the near-optimal clustering obtained, t-test, and required running time. The comparison results show that the MBC clustering using gray wolf optimization (GWOCustering) outperforms the benchmark algorithms, achieving an average Silhouette score of 0.7491. In addition, this algorithm significantly outperforms Kmeans based on t-test conducted and results in relatively low runtime. The GWOCustering simulation yielded four near-optimal clusters in the customer segmentation of Bintang Shoes Purwokerto MSME. The analysis of each cluster's characteristics indicates the need for distinct marketing strategies for this MSME. Marketing based on offline purchasing services, store conditions, and product layout is appropriate for consumers in Cluster 1. Meanwhile, digital marketing with attractive, informative graphic content is suitable for consumers in Clusters 2 and 4. Furthermore, strategies with competitive pricing, discounts, or bundling strategies are appropriate for Cluster 3.

Keywords: clustering; customer segmentation; metaheuristic; MSMEs; PCoA

How to Cite: R. Ananda, S. Sinaga, B. Pratikno, N. A. H. Zulkainain, and T. A. Sundara, "Metaheuristic-based Clustering Algorithms with Principal Coordinate Analysis for Shoe Market Segmentation", *J. RESTI (Rekayasa Sist. Teknol. Inf.)*, vol. 10, no. 3, pp. 702 - 711, Jun. 2026. [Permalink/ doi: https://doi.org/10.29207/resti.v10i3.7369](https://doi.org/10.29207/resti.v10i3.7369)

Received: December 3, 2025

Accepted: June 1, 2026

Available Online: June 16, 2026

This is an open-access article under the [CC BY 4.0 License](#)

Published by [Ikatan Ahli Informatika Indonesia](#)

1. Introduction

Micro, Small, and Medium Enterprises (MSMEs) are very important to Indonesia's economy. They account for 99% of all businesses and contribute 61% to the country's GDP [1], [2]. Among these, the footwear industry has experienced significant growth due to increased demand for shoes [3]. Although they are crucial to the economy, many MSMEs struggle to maximize their sales due to ineffective marketing strategies. One of them is Bintang Sepatu Purwokerto, a shoemaker in Banyumas Regency, Central Java, Indonesia. This MSME struggles to match its products with customer demand, resulting in underutilized production capacity and stagnant sales growth.

Figure 1 shows that the gap between actual sales and monthly targets at Bintang Sepatu Purwokerto has remained unchanged over time. The target sales were raised in October and November after hiring more salespeople, which was a strategic move to increase revenue. However, even after this intervention, sales performance continued to fall short of targets, with an average monthly decline of 4%. This downward trend shows that simply adding salespeople was not enough to close the gap. This suggests we should focus on other aspects of the marketing strategy, such as customer segmentation.

Clustering is a popular approach used in marketing analytics, especially in identifying market segments [4-7]. The most popular clustering algorithm is k-means

[8]. The K-means algorithm, combined with PCA, can improve customer segmentation in dynamic retail settings [9]. In addition, compared with other algorithms such as agglomerative and DBSCAN, the combination of K-means and PCA shows better performance in examining the behaviors and preferences of online shopping customers [10].

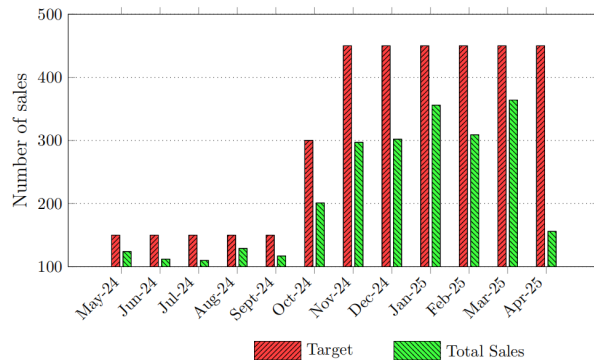


Figure 1. Sales and target data at Bintang Sepatu Purwokerto

Despite its advantages, the combination of K-means and PCA has essential limitations. Both K-means and PCA cannot handle mixed data types, such as numeric and categorical data, which are common in customer segmentation datasets [11]. This is because K-means updates the centroids in each iteration using the mean, which is not suitable for categorical data. Simultaneously, PCA aims to elucidate the variance of numerous variables by linearly combining them, leveraging the Euclidean distance [12]. This distance is not appropriate for implementation in the categorical dataset [13]. Furthermore, the K-means algorithm iteratively updates the positions of the cluster centroids by computing the mean of the data points assigned to each centroid. This process continues until convergence is reached. However, this strategy often leads to local minima, which makes clustering less effective [14].

To overcome the limitations of PCA in handling categorical data, several alternative algorithms can be utilized with K-means, such as Multiple Correspondence Analysis (MCA) [15], Factor Analysis of Mixed Data (FAMD) [16], t-distributed Stochastic Neighbor Embedding (t-SNE) [17], Uniform Manifold Approximation and Projection (UMAP) [18], or Principal Coordinate Analysis (PCoA) [19]. The PCoA algorithm offers a competitive advantage for determining the configuration of objects in mixed, categorical, and numerical data using the Gower distance [4], [20]. Also, the PCoA procedure does not involve transformations that can reduce the information in the data, unlike MCA, which uses a discretization approach to transform categorical data [21], [22]. Additionally, PCoA operates directly on the adjacency matrix, which is certainly better at representing the data's reality than FAMD, which imposes assumptions of linearity and variance structure on the dataset [23]. In terms of efficiency, PCoA outperforms t-SNE and UMAP because it does not require hyperparameters that are often sensitive and computationally expensive [24-

26]. Furthermore, PCoA preserves global distances [27], whereas t-SNE and UMAP focus only on local structures [28]. Based on the advantages described, PCoA is appropriate to be used when the analysis goal is to mine information from the structure of dissimilarities between objects in the dataset.

Furthermore, the limitations of K-means due to the centroid-updating procedure, which can fall into a local optimum, have been attempted to be addressed. Some solutions, such as diffusion k-means [29] and the integration of Radial Basis Function and k-means (RBF K-means) [30], have been proposed to overcome k-means' limitations. Diffusion k-means involves a random walk, namely a Markov chain, and the combination of Euclidean and diffusion distances to maximize the within-cluster connectedness. Meanwhile, integrating differentiable Radial Basis Functions with k-means yields an exact gradient-based update rule for k-means centroids. Despite their reliability, both approaches face challenges, namely that they are often computationally intractable and provide exact solutions only in particular cases [29], [31]. Another approach to optimizing k-means is metaheuristic optimization. A metaheuristic combination with k-means, called Kmeans Q-learning differential evolution (KmeansQLDE), has been proposed for customer segmentation data [32]. Despite its low computation, the metaheuristic in K-meansQLDE only determines the optimal initial centroids, which are then used in the regular K-means. This combination does not address the limitations of K-means in its centroid-updating procedure. Therefore, there is a need for another strategy in combining metaheuristics and k-means in this problem.

Based on the aforementioned statements, this study proposes metaheuristic-based clustering (MBC) algorithms that perform clustering using updated centroids, leveraging a nature-inspired procedure. The proposed MBC algorithms do not use k-means; instead, they directly improve the performance of metaheuristic optimization methods for clustering. The metaheuristics used are Aquila Optimizer (AO) [33], Differential Search Algorithm (DSA) [34], Grey Wolf Optimizer (GWO) [35], Komodo Mlipir Algorithm (KMA) [36], Particle Swarm Optimization (PSO) [37], and Whale Optimization Algorithm (WOA) [38]. Those metaheuristics were selected based on their state-of-the-art methods and demonstrated success in solving complex optimization problems [39], [40]. In addition, dimensionality reduction using PCoA was implemented to handle both numerical and categorical data in the customer segmentation dataset from MSME Bintang Shoes Purwokerto and other datasets. The key contributions of this work are (1) to develop the MBC algorithms that enhance centroid refinement and mitigate local optimum convergence issues, (2) to provide a comparative evaluation among the proposed MBC algorithms and the benchmark algorithms, such as Kmeans and KmeansQLDE, and (3) to demonstrate the practical applicability of the best MBC algorithm

through a real-world case study, namely MSMEs Bintang Shoes Purwokerto, offering actionable insights for data-driven marketing strategy formulation.

This study is structured as follows: Section 2 presents the materials and methods used. Section 3 provides the results of the clustering analysis experiment and discusses the clustering results, including their comparative analysis and performance. Lastly, Section 4 concludes the paper by summarizing the key insights obtained.

2. Methods

This section includes six subsections. The first subsection describes the dataset employed in this study, including the primary dataset and the public datasets. The second subsection presents the proximity measure utilized to quantify the dissimilarity among objects. Subsequently, the third subsection outlines the procedural steps of the Principal Coordinates Analysis (PCoA) algorithm. The fourth subsection introduces the proposed metaheuristic-based clustering framework. The fifth subsection outlines the performance evaluation metrics employed. Finally, the sixth subsection presents the research flow.

2.1. The dataset used

The primary dataset used in this study is the customer dataset of MSME Bintang Sepatu Purwokerto. In addition, open-source Kaggle datasets are used to perform experiments and evaluate the proposed algorithms against benchmark algorithms. Table 1 lists the datasets used in this study, which contain both categorical and numerical data.

Table 1. The dataset used

Dataset	ID
Predict Customer Purchase Behavior	D1
The Ship Performance	D2
Shop Customer Data	D3
MSME of Bintang Sepatu Purwokerto	D4

Datasets D1, D2, and D3 were obtained from the open-source Kaggle dataset, which has been widely used to evaluate the performance of the developed algorithm [41-43]. These four datasets have varying numbers of objects (n), variables (p), and contain both categorical and numeric variables, detailed information is presented in Table 2.

Table 2. Characteristics of the dataset used

ID	n Object	p Variables	The number of variables	
			Categoric	Numeric
D1	1500	9	5	4
D2	2127	14	3	11
D3	2000	7	2	5
D4	179	16	14	2

Meanwhile, dataset D4 is secondary data on real sales, collected and selected based on marketing mix variables [44-46]. The selected variables are presented in Table 3.

Table 3. The selected variables

Variable	Description
Address [47]	Residence location of customers
Gender [48]	Customer gender identity
Age [49]	Customer age in years
Occupation [48]	Customer occupation
Income[48]	Monthly income range bracket
Motivation [50]	Motivation for purchase decisions
Frequency [51]	Number of transactions
Model [52]	Preferred shoe style/design
Size [52]	Purchased shoe sizing
Color [52]	Selected color preferences
Price [47]	Shoe prices
Place of purchase [52]	Primary purchase platform (online/offline)
Social Media Engagement [47]	Daily social platform usage duration
Negotiation [47]	Price bargaining during transactions
Information Sources [47]	Sources of information obtained by consumers regarding products
Marketing Consent [47]	Permission for product promotion communications

2.2. The proximity measure

To determine the configuration of each object, the proximity measures are required. For numerical variables, we utilize Euclidean distance, which has proven effective in optimization research [53] and data mining applications [54-56]. Formula to calculate the Euclidean distance between x_i and x_j is demonstrated in Equation 1.

$$d_E(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{(\mathbf{x}_i - \mathbf{x}_j)^T (\mathbf{x}_i - \mathbf{x}_j)} \quad (1)$$

where refers to the Euclidean distance between object- i and object- j . Meanwhile, refers to the vector whose elements are the information of the i -th object for each variable.

Meanwhile, for categorical data, this study employs the Gower distance [57] as a proximity measure. This approach has proven particularly useful for analyzing categorical datasets in several fields, such as education [58] and imputation data [59]. The Gower distance framework transforms similarity measures into proximity measures using the complementarity concept. Formula of the Gower distance and its corresponding similarity measure between and are showed in Equations 2 and 3, respectively.

$$d_G(x_i, x_j) = 1 - s(x_i, x_j) \quad (2)$$

$$s_{ij} = \frac{1}{p} \sum_{k=1}^p (1 - |x_{ik} - x_{jk}|) \quad (3)$$

$d_G(x_i, x_j)$ denotes the Gower distance between x_i and x_j . Meanwhile, is the similarity measure between them. is the number of variables in the dataset.

For the mixed (numerical and categorical) dataset, the combination of Euclidean and Gower distances is used by leveraging a weighted combination method [40]. The complete computational procedure for handling the mixed dataset is carried out through a series of systematic steps. Initially, dataset of objects and variables, containing both numerical and categorical types, is given. The dataset is subsequently partitioned

into two subsets, $nX_p = (nY_{p1}|nZ_{p2})$, based on variable type, yielding a numerical and a categorical dataset (nZ_{p2}). This separation enables each data type to be processed using an appropriate similarity or dissimilarity measure. For numerical data (nY_{p1}), the Euclidean distance between x_i and x_j , for all i, j , namely is calculated. Meanwhile, the Gower distance between x_i and x_j , for all i, j , namely is conducted for categorical data. After obtaining the respective proximity matrices, both matrices are standardized to ensure comparability and to prevent any particular data type from dominating the proximity assessment. Finally, the standardized numerical and categorical proximity matrices are integrated to construct a unified proximity measure between objects using Equation 4.

$$d_{ij} = \frac{d_E(y_i y_j)}{p_1} + \frac{d_G(z_i z_j)}{p_2}, \quad \forall i, j \quad (4)$$

2.3. Principal Coordinate Analysis (PCoA)

This study employs PCoA to obtain the object configurations in a low-dimensional Euclidean space while preserving dissimilarity from the original proximity matrix [60]. PCoA was selected for this study due to its ability to analyze mixed data types (both numerical and categorical). In addition, PCoA maintains geometric interpretability in Euclidean space, which is required by the clustering algorithms.

The procedure of PCoA begins by transforming the proximity matrix, $nD_n = (d_{ij})$, into a scalar product matrix through a distance-based transformation, namely $nA_n = (a_{ij})$ where $a_{ij} = -\frac{1}{2}d_{ij}^2, \forall i, j \in [1, n]$. Then, is calculated, where $J = 11'$. Subsequently, spectral decomposition is performed, yielding $B = V\Lambda V$. These components of the decomposition results are then used to construct the object configurations in a k -dimensional Euclidean space using formula $\tilde{X} = V_k \Lambda_k^{1/2}$.

To evaluate the quality of the obtained representation, from is calculated using monotonic regression. Then, compute STRESS using Equation 5. If the STRESS value is high then determine the object configurations in another dimension. Otherwise, the object configurations are obtained with the good quality.

$$STRESS = \frac{\sum_{i < j} (\tilde{d}_{ij} - d_{ij})^2}{\sum_{i < j} d_{ij}^2} \quad (5)$$

2.4. Metaheuristic-based clustering (MBC)

This study develops a metaheuristic-based clustering (MBC) algorithm to avoid the tendency of the clustering process to become trapped in local optima during centroid updates [10]. The developed algorithm dynamically updates cluster centroids by leveraging the natural search mechanisms of the metaheuristic algorithms used. Figure 2 provides the general framework of the MBC algorithms.

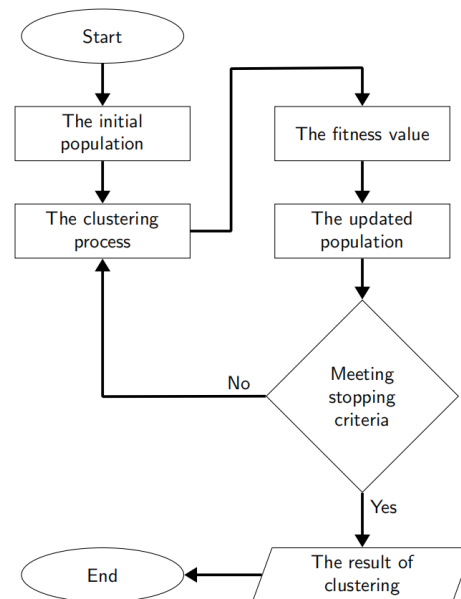


Figure 2. The general framework of the MBC algorithms

The proposed metaheuristic-based clustering framework begins by generating the initial population consisting of n candidate solutions, where each individual represents a set of k cluster centroids randomly. For every individual, the clustering process is performed by assigning each data object to its nearest centroid based on the Euclidean distance measure. Subsequently, the fitness value of each individual is calculated using the silhouette score. Based on these fitness values, the population is updated by adjusting the centroid positions according to the naturally inspired search methodology of the metaheuristic used. If the convergence criterion or iteration limit is reached, the clustering results are close to optimal. Otherwise, the iterative process is conducted until the termination criterion is met.

Table 4 provides an overview of the MBC algorithms implemented using various metaheuristic. Furthermore, the centroids are updated iteratively using the standard procedure for each metaheuristic. Moreover, the Kmeans and KmeansQLDE algorithms are also utilized as benchmark algorithms.

Table 4. The develop metaheuristic-based clustering algorithm

Metaheuristic	MBC algorithm
Aquila optimizer (AO) [16]	AOCustering
Differential search algorithm (DSA) [17]	DSAClustering
Grey wolf optimization (GWO) [18]	GWOClustering
Komodo mlpir algorithm (KMA) [19]	KMAClustering
Particle swarm optimization (PSO) [20]	PSOClustering
Whale optimization algorithm (WOA) [21]	WOAClustering

2.5. The performance measurements

The primary performance measure used in this study is the Silhouette score, which measures the quality of each algorithm's clustering results. The formula of the silhouette score is demonstrated in Equation 6.

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (6)$$

In the formula, is the Silhouette score of the i th object. Then, $a(i)$ is the average distance between the i th object and all objects in its cluster. Meanwhile, is the average distance between the i th object and objects in its nearest cluster. The total Silhouette score is then calculated from the average converted to $[0,1]$, using Equation 7 [61]. Cluster quality is determined based on the silhouette score, with quality criteria presented in Table 5.

$$\text{Silhouette score} = \frac{1}{2n} \sum_{i=1}^n [s(i) + 1] \quad (7)$$

Table 5. Clustering quality criteria

Interval	Characteristics
0.710 - 1.000	Strong cluster
0.510 - 0.709	Reasonable cluster
0.260 - 0.509	Weak cluster
< 0.259	wrong cluster

In addition, this study also uses a t-test statistical test to compare simulation results between the developed algorithms and the benchmark algorithm. The normality test used is the Shapiro-Wilk test because the sample size is less than 50 [62]. Furthermore, the t-test used is the Student's t-test, when the data distribution is normal. Conversely, the Wilcoxon test is utilized when the data distribution is not normal. When $p_{value} < 0.05$, reject H_0 , indicating a significant difference in the performance of the developed and benchmark algorithms.

2.6. Research flow

This study contains four main stages as shown in Figure 3. Stage 1 includes data collection, data preprocessing, proximity matrix calculation, and object configuration determination using PCoA. Stage 2 is the process of simulating the MBC and benchmark algorithms to obtain optimal clustering. Furthermore, stage 3 compares the results between the simulated algorithms and interprets the best clustering on the MSME dataset of Bintang Sepatu Purwokerto. The final stage contains conclusions.

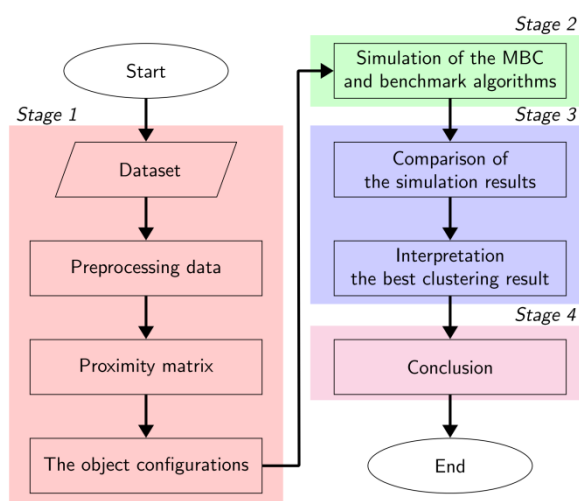


Figure 3. Research flow in this study

The simulations used in this research were written in MATLAB 2021b programming language. The studies were conducted on an AMD Ryzen AI 7 350 w/ Radeon 860M, 16.0 GB of RAM, and Windows 11 Home Single Language.

3. Results and Discussions

This section presents the simulation results obtained from the proposed metaheuristic-based clustering and provides a comprehensive discussion of its effectiveness in the dataset used. The analysis begins with the preprocessing stage, where the mixed-type dataset is prepared and transformed to ensure its suitability for subsequent analysis. Following this, the performance of the proposed approach is evaluated through a comparative simulation study involving several metaheuristic-based clustering algorithms and another state-of-the-art clustering algorithm used, enabling a thorough assessment of clustering quality and optimization capability. Finally, the best-performing clustering model is applied to the customer data of Bintang Sepatu Purwokerto MSMEs to identify meaningful customer segments, thereby providing valuable insights to support targeted marketing strategies and customer relationship management.

3.1. Preprocessing dataset

In stage 1, preprocessing and labeling categorical data as numbers were performed. Subsequently, the proximity matrix was calculated for each dataset using the procedure described in Subsection 2.2. Furthermore, the objects in the datasets were projected into a lower-dimensional space using PCoA to obtain their configurations.

The selection of the lowest dimensionality was based on the STRESS value obtained. Datasets D1, D2, and D3 are presented in dimension seven with STRESS values around 0.1 (10%), indicating that the resulting configuration adequately represents the proximity matrix [63]. Meanwhile, the configuration of the D4 dataset is in dimension two with a STRESS value of 0.091 (9.1%) which indicates a good fit between the derived configuration and its proximity matrix.

3.2. Comparison of the simulation results

This section presents and compares the simulation data, convergence rates, and near-optimal outcomes of the MBC and benchmark algorithms. The optimal number of clusters is determined using Silhouette criterion provided by MATLAB. The optimal number of clusters in datasets D1 and D3 is two clusters. Simultaneously, datasets D2 and D4 provide four clusters.

Furthermore, Table 6 presents the computational parameters employed by the metaheuristic optimization within the metaheuristic-based clustering (MBC) algorithms throughout the clustering procedure. Meanwhile, the metaheuristic of KmeansQLDE algorithm employs parameters with values as delineated in the referenced literature [32].

Table 6. The computational parameters utilized

Alg.	Parameter	Value	Ref.
-	population	100	[64]
-	The number of iteration	100	[64]
AO	Exploitation adjustment (α)	0.1	[33]
	Exploitation adjustment (γ)	0.1	[33]
DSA	Perturbation	$0.3 \times \text{rand}$	[34]
GWO	Adaptive parameter (a)	[0,2]	[35]
KMA	Portion (p)	0.5	[36]
	Mrpir rate (r)	0.2	[36]
PSO	Weight 1 (w_1)	0.8	[37]
	Weight 2 (w_2)	0.2	[37]
	Personal learning (c_1)	2	[37]
	Global learning (c_2)	2	[37]
WOA	Adaptive parameter (a)	[0,2]	[38]

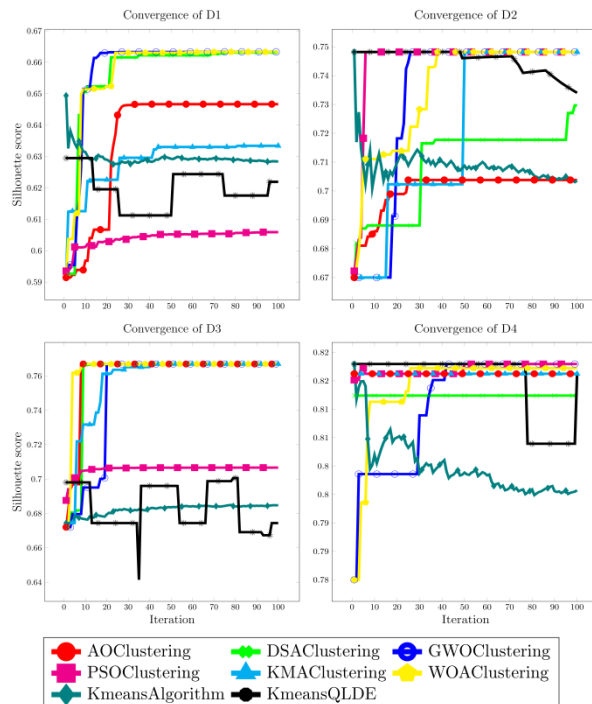


Figure 4. Convergence graphs of the simulated algorithms

Figure 4 presents the convergence graphs of the MBC algorithms and the benchmark algorithms. On dataset D1, GWOClustering, DSAClustering, and WOAclustering achieve near-optimal clustering performance, outperforming benchmark algorithms. Meanwhile, PSOClustering is trapped in local optima due to premature convergence. KmeansQLDE shows an interesting phenomenon, with fluctuating Silhouette scores. This is because the metaheuristic in KmeansQLDE only finds optimal initial centroids based on the sum of squared errors (SSE), which cannot guarantee good clustering results in Kmeans. On dataset D2, the best near-optimal clustering is obtained by GWOClustering, PSOClustering, KMAclustering, and WOAclustering, which outperform KMeans and KMeansQLDE. Meanwhile, DSAClustering shows moderate performance. In contrast, AOClustering shows premature convergence. Meanwhile, in D3, each MBC algorithm performs better than KMeans and KMeansQLDE. However, PSOClustering exhibits premature convergence, resulting in lower performance than other MBC algorithms. Finally, in the D4 dataset, it is known that GWOClustering, AOClustering,

PSOClustering, KMAclustering, and WOAclustering produce better near-optimal clustering compared to the benchmark algorithms. Meanwhile, DSAClustering shows performance below that of KMeansQLDE and has premature convergence. In general, the GWOClustering and WOAclustering algorithms demonstrate consistent performance with relatively high near-optimal clustering.

Furthermore, Table 7 presents the near-optimal clustering obtained. The table highlights the results of GWOClustering, which indicate that the algorithm consistently produces the best near-optimal clustering measures across datasets. Meanwhile, WOAclustering performs slightly below GWOClustering. Therefore, GWOClustering produces the best near-optimal clustering in the simulation conducted.

Table 7. The near-optimal clustering of each algorithm simulated

Algorithm	D1	D2	D3	D4
AOClustering	0.6467	0.7038	0.7670	0.8162
DSAClustering	0.6631	0.7297	0.7670	0.8124
GWOClustering	0.6633	0.7482	0.7670	0.8180
KMAclustering	0.6334	0.7482	0.7670	0.8162
PSOClustering	0.6059	0.7482	0.7067	0.8180
WOAclustering	0.6633	0.7482	0.7670	0.8172
Kmeans	0.6284	0.7035	0.6849	0.7957
KmeansQLDE	0.6196	0.7340	0.6837	0.8149

Subsequently, the Student t-test was used to determine whether the MBC algorithms significantly outperform the benchmark algorithms. The t-test results showed that GWOClustering and WOAclustering performed significantly better than Kmeans, with p-values of 0.0376 and 0.0388, respectively.

Running time for each algorithm

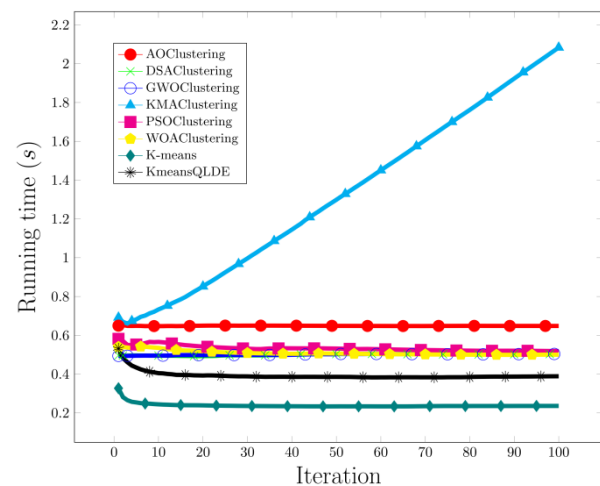


Figure 5. Running time of each algorithm simulated

In addition, Figure 5 presents the average running time obtained. Based on the figure, the developed algorithms have running times that are not much different from those of the benchmark algorithms. The difference in running time of the developed algorithms against Kmeans and KmeansQLDE ranges from 0.209 to 0.311 seconds, respectively, indicating a relatively low difference. Meanwhile, KMAclustering shows a unique increase in running time, a consequence of its added

population procedure that activates when the solution stagnates [65]. Despite these variations, the overall computational demand for obtaining the near-optimal clustering remains manageable.

3.3. Customer Segmentation of Bintang Sepatu Purwokerto MSMEs

The simulation results show that GWOCustering is the best algorithm. This algorithm shows the highest Silhouette value, namely 0.7491, on the Bintang Sepatu Purwokerto MSME customer segmentation dataset. Figure 6 presents a visualization of the near-optimal clustering obtained from the GWOCustering algorithm. The first and second principal components are valued at 47.16% and 18.79%, respectively, indicating that the visualization is able to retain 65.95% of the total variance [66]. This proves that most of the structural information from the dataset in the proximity matrix is well retained in two-dimensional space.

The visualization in Figure 6 shows four well-defined clusters. The distribution of objects in Clusters 1, 2, 3, and 4 is 46 (25.70%), 36 (20.11%), 45 (25.14%), and 52 (29.05%), respectively, which indicates that the objects

in the dataset are distributed relatively evenly across the clusters. Table 8 demonstrates the profile and characteristics of customer segmentation from each cluster obtained.

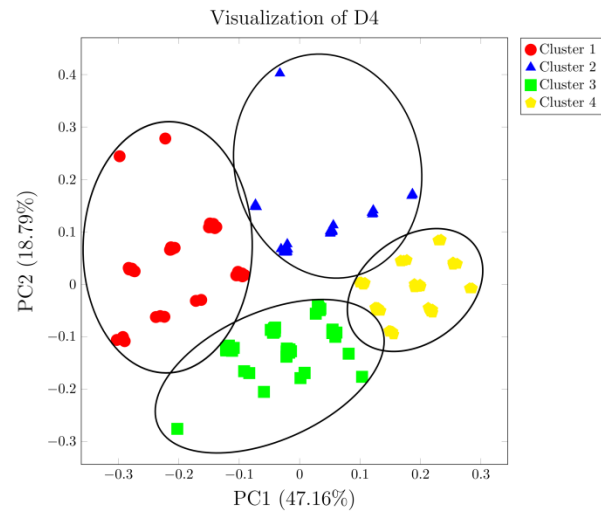


Figure 6. Visualization of the GWOCustering result

Table 8. Consumer Segmentation Profile and Characteristics of each Cluster

Characteristics	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Gender	Male-dominated ($\approx 80\%$)	Mixed (Male & Female)	Mixed (Male & Female)	Dominated by Women
Age	> 36 years old	26 - 35 Years (Young Adult)	among 15 to 45 (Very Varied)	26 - 45 Years
Income	1 - 6 Million (Middle)	1 - 6 Million	1 - 3 Million (Lower Middle Class)	1 - 6 Million (Variety)
Shoe Price	High (Rp250,000 - 400,000)	Upper Middle Class (Rp250,000)	Middle Class (Rp150,000 - 200,000)	Flexible (Rp125,000 - 250,000)
Shopping Location	Offline (Physical Store)	Online & Offline (Mixed)	Online & Offline	Online (E-commerce)
Bargaining	No Bargaining	Bargaining	Bargaining	Actively Bargain
Purchase Frequency	1-3 Times	Varied	Varied	Two or More Times (Loyal)
Information Source	Friends/Neighbors, Facebook	Instagram, TikTok	Shopee, Facebook	Shopee, TikTok, Lazada
Main Behavior	Conventional, averse to bargaining, purchase out of necessity.	Technologically proficient, prepared to invest additional funds for superior quality, and seeks knowledge via social media.	Seeking cost-effective, price-conscious, and utilitarian options.	Modern, proactively pursues deals and information, adept at online buying, impetuous, and emotional.

Based on the clustering characteristic analysis conducted by [4], it is known that Cluster 1 constitutes a conventional buyer segment predominantly composed of males aged 36 and above, with incomes ranging from 1 to 6 million rupiah. Consumers in this segment make purchases based on necessity, exhibit loyalty to physical retail establishments, and generally do not favor bargaining. The selected footwear often features a macho design and is priced between Rp 250,000 and 400,000. They generally acquire information from acquaintances, neighbors, or social media sites such as Facebook. Cluster 2 comprises young individuals aged 26 to 35, characterized by an equitable gender distribution and middle-income demographics. They are pragmatic yet prioritize quality, with shoe costs starting at approximately 250,000 rupiah. This group

exhibits adaptability in selecting shopping channels (both online and offline), participates in minimal bargaining, and predominantly utilizes visual social media platforms such as Instagram and TikTok for information acquisition.

Subsequently, Cluster 3 comprises functional purchasers from varied age demographics and lower-middle-income brackets (1–3 million rupiah). Their motivations are predominantly sensible and price-sensitive, selecting shoes within the Rp150,000–200,000 range. They utilize a combination of offline and online shopping methods, engage in minimal negotiation, and obtain information via e-commerce sites such as Shopee and Facebook. Meanwhile Cluster 4, the most substantial portion, predominantly comprises women aged 26–45 and demonstrates the

most active behavior. Their justifications encompass both rational and emotional factors, accompanied by a variable price range of Rp125,000–250,000. They are highly active in the digital realm, making online shopping their primary choice, actively seeking information from various sources (Shopee, Lazada, TikTok), actively bargaining, and tend to make repeat purchases and discuss them with their partners.

These four clusters demonstrate a unique range of customer behavior. Cluster 1 and Cluster 4 differ in their shopping channel and incentive preferences. Cluster 1 exhibits a strong adherence to traditional and rational values, demonstrating commitment to physical retail establishments. On the other hand, Cluster 4 is characterized by modernity, emotional engagement, and allegiance to the online buying environment. Cluster 2 and Cluster 3 occupy an intermediary position on channel flexibility, though they differ in their priorities: Cluster 2 emphasizes quality and is influenced by visual material, whereas Cluster 3 is more concerned with value for money and price sensitivity. Regarding information seeking, Cluster 4 and Cluster 2 exhibit significant adaptability to contemporary digital platforms (TikTok, e-commerce), but Cluster 1 continues to depend on traditional social networks, while Cluster 3 employs e-commerce effectively. The disparities underscore the need for tailored marketing strategies: (a) targeted in-store methods for Cluster 1, (b) graphic content marketing for Clusters 2 and 4, and lastly (c) competitive pricing for Cluster 3.

4. Conclusions

This study develops the metaheuristic-based clustering (MBC) algorithms by running several metaheuristics. The comparison of simulation results shows that GWOCustering produces the best near-optimal clustering, with an average Silhouette score of 0.7491, indicating strong cluster formation. In addition, GWOCustering significantly outperforms Kmeans with relatively low runtime requirements. Furthermore, in the customer segmentation dataset obtained from Bintang Sepatu Purwokerto MSME, the GWOCustering simulation yielded four clusters, with relatively uniform distribution.

The analysis of each cluster's characteristics indicates the need for distinct marketing strategies for this MSME. Marketing based on offline purchasing services, store conditions, and product layout is appropriate for consumers in Cluster 1. Meanwhile, digital marketing with attractive, informative graphic content is suitable for consumers in Clusters 2 and 4. Furthermore, strategies with competitive pricing, discounts, or bundling strategies are appropriate for Cluster 3. The simulation code of this study is presented at the link <https://github.com/ridho-cell/Metaheuristic-based-clustering-algorithm>

Acknowledgements

The authors are grateful for Telkom University Campus Purwokerto support. The author also extends sincere thanks to UMKM Bintang Sepatu Purwokerto for their valuable contribution as the subject of this research. The author declares no conflict of interest.

References

- [1] A. A. Siburian and D. P. Ompusunggu, "Pengaruh Pertumbuhan Usaha Mikro, Kecil dan Menengah Terhadap Pertumbuhan Ekonomi Di Palangka Raya," *JUEB*, vol. 2, no. 2, pp. 30–34, Jun. 2023, doi: [10.57218/jueb.v2i2.604](https://doi.org/10.57218/jueb.v2i2.604).
- [2] Kominfo, "Transformasi Digital UMKM Jadi Prioritas Penguatan Fondasi Ekonomi." [Online]. Available: <https://kominfo.ngawikab.go.id/transformasi-digital-umkm-jadi-prioritas-penguatan-fondasi-ekonomi/>
- [3] L. Yulia and W. Setianingsih, "Studi Manajemen Marketing Berbasis Online (Penelitian Pada UMKM Produksi Mebel di Babakan Muncang Tamansari Kota Tasikmalaya)," *Jurnal Maneksi*, vol. 9, no. 1, pp. 346–354, Jul. 2020, doi: [10.31959/jm.v9i1.397](https://doi.org/10.31959/jm.v9i1.397).
- [4] S. Sinaga, R. Ananda, H. Q. Karima, and A. M. Tazuddin, "Marketing Analysis of Shoe Products Using Principal Coordinates Analysis and K-Means Clustering Based on the Marketing Mix at Bintang Sepatu Purwokerto MSME," *J. Tek. Inform. (JUTIF)*, vol. 6, no. 3, pp. 1405–1418, Jun. 2025, doi: [10.52436/1.jutif.2025.6.3.4687](https://doi.org/10.52436/1.jutif.2025.6.3.4687).
- [5] M. F. Nur and A. Siregar, "Exploring the Use of Cluster Analysis in Market Segmentation for Targeted Advertising," *itsdi*, vol. 5, no. 2, pp. 158–168, Feb. 2024, doi: [10.34306/itsdi.v5i2.665](https://doi.org/10.34306/itsdi.v5i2.665).
- [6] W. A. Prastyabudi, A. N. Alifah, and A. Nurdin, "Segmenting the Higher Education Market: An Analysis of Admissions Data Using K-Means Clustering," *Procedia Computer Science*, vol. 234, pp. 96–105, 2024, doi: [10.1016/j.procs.2024.02.156](https://doi.org/10.1016/j.procs.2024.02.156).
- [7] 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, p. 7243, Jun. 2022, doi: [10.3390/su14127243](https://doi.org/10.3390/su14127243).
- [8] A. Kusdaryanto, C. D. Wijanarko, P. D. Widyantara Usat, and A. Prabowo, "The The Use of the K-Means Algorithm in Analyzing E-Commerce Consumer Segmentation: A Case Study of the Online Retail Dataset (UK)," *JTOS*, vol. 8, no. 2, pp. 612–622, Dec. 2025, doi: [10.36378/jtos.v8i2.4798](https://doi.org/10.36378/jtos.v8i2.4798).
- [9] K. K. Patibandla, R. Daruvuri, and P. Mannem, "Enhancing Online Retail Insights: K-Means Clustering and PCA for Customer Segmentation," in *2025 3rd International Conference on Advancement in Computation & Computer Technologies (InCACCT)*, Gharuan, India: IEEE, Apr. 2025, pp. 388–393, doi: [10.1109/InCACCT65424.2025.11011448](https://doi.org/10.1109/InCACCT65424.2025.11011448).
- [10] R. M. Hussein, K. W. Khaw, A. Gaber, and X. Chew, "Examining the Behaviors and Preferences of Online Shopping Customers Using Clustering Techniques," *JSCDM*, vol. 5, no. 1, Jun. 2024, doi: [10.30880/jscdm.2024.05.01.009](https://doi.org/10.30880/jscdm.2024.05.01.009).
- [11] P. Liu, H. Yuan, Y. Ning, B. Chakraborty, N. Liu, and M. A. Peres, "A modified and weighted Gower distance-based clustering analysis for mixed type data: a simulation and empirical analyses," *BMC Med Res Methodol*, vol. 24, no. 1, p. 305, Dec. 2024, doi: [10.1186/s12874-024-02427-8](https://doi.org/10.1186/s12874-024-02427-8).
- [12] M. Greenacre, P. J. F. Groenen, T. Hastie, A. I. D'Enza, A. Markos, and E. Tuzhilina, "Principal component analysis," *Nat Rev Methods Primers*, vol. 2, no. 1, p. 100, Dec. 2022, doi: [10.1038/s43586-022-00184-w](https://doi.org/10.1038/s43586-022-00184-w).
- [13] T.-H. T. Nguyen, D.-T. Dinh, S. Sriboonchitta, and V.-N. Huynh, "A method for k-means-like clustering of categorical data," *J Ambient Intell Human Comput*, vol. 14, no. 11, pp. 15011–15021, Nov. 2023, doi: [10.1007/s12652-019-01445-5](https://doi.org/10.1007/s12652-019-01445-5).
- [14] H. Hu, J. Liu, X. Zhang, and M. Fang, "An Effective and Adaptable K-means Algorithm for Big Data Cluster Analysis," *Pattern Recognition*, vol. 139, p. 109404, Jul. 2023, doi: [10.1016/j.patcog.2023.109404](https://doi.org/10.1016/j.patcog.2023.109404).

- [15] W. Atkinson, "Charting fields and spaces quantitatively: from multiple correspondence analysis to categorical principal components analysis," *Qual Quant*, vol. 58, no. 1, pp. 829–848, Feb. 2024, doi: [10.1007/s11135-023-01669-w](https://doi.org/10.1007/s11135-023-01669-w).
- [16] D. Soemitro and J. F. S. R. Neto, "Spectral Clustering of Categorical and Mixed-Type Data via Extra Graph Nodes," in *2025 IEEE International Conference on Big Data (BigData)*, Macau, China: IEEE, Dec. 2025, pp. 609–617. doi: [10.1109/BigData66926.2025.11401856](https://doi.org/10.1109/BigData66926.2025.11401856).
- [17] K. Olobatuyi, M. R. P. Parker, and O. Ariyo, "Cluster weighted model based on TSNE algorithm for high-dimensional data," *Int J Data Sci Anal*, vol. 17, no. 3, pp. 261–273, Apr. 2024, doi: [10.1007/s41060-023-00422-8](https://doi.org/10.1007/s41060-023-00422-8).
- [18] J. Healy and L. McInnes, "Uniform manifold approximation and projection," *Nat Rev Methods Primers*, vol. 4, no. 1, p. 82, Nov. 2024, doi: [10.1038/s43586-024-00363-x](https://doi.org/10.1038/s43586-024-00363-x).
- [19] S. Komaki et al., "Dimension reduction of microbiome data linked Bifidobacterium and Prevotella to allergic rhinitis," *Sci Rep*, vol. 14, no. 1, p. 7983, Apr. 2024, doi: [10.1038/s41598-024-57934-x](https://doi.org/10.1038/s41598-024-57934-x).
- [20] J. T. Hidrobo-Morales, J. C. Marcillo-Delgado, and A. García-Carrillo, "Analysis of innovation drivers with a mapping innovation approach Practical application in Ecuador," *Int. J. Innov.*, vol. 12, no. 1, pp. 1–51, Mar. 2024, doi: [10.5585/2024.23974](https://doi.org/10.5585/2024.23974).
- [21] N. Kosaraju, S. R. Sankepally, and K. Mallikharjuna Rao, "Categorical Data: Need, Encoding, Selection of Encoding Method and Its Emergence in Machine Learning Models—A Practical Review Study on Heart Disease Prediction Dataset Using Pearson Correlation," in *Proceedings of International Conference on Data Science and Applications*, vol. 551, M. Saraswat, C. Chowdhury, C. Kumar Mandal, and A. H. Gandomi, Eds., Singapore: Springer Nature Singapore, 2023, pp. 369–382. doi: [10.1007/978-981-19-6631-6_26](https://doi.org/10.1007/978-981-19-6631-6_26).
- [22] D. L. Aguilar, M. A. Medina-Pérez, O. Loyola-González, K.-K. R. Choo, and E. Bucheli-Susarrey, "Towards an Interpretable Autoencoder: A Decision-Tree-Based Autoencoder and its Application in Anomaly Detection," *IEEE Trans. Dependable and Secure Comput.*, vol. 20, no. 2, pp. 1048–1059, Mar. 2023, doi: [10.1109/TDSC.2022.3148331](https://doi.org/10.1109/TDSC.2022.3148331).
- [23] C. K. J. Hou and K. Behdinan, "Dimensionality Reduction in Surrogate Modeling: A Review of Combined Methods," *Data Sci. Eng.*, vol. 7, no. 4, pp. 402–427, Dec. 2022, doi: [10.1007/s41019-022-00193-5](https://doi.org/10.1007/s41019-022-00193-5).
- [24] P. Chourasia, T. Murad, S. Ali, and M. Patterson, "Enhancing t-SNE Performance for Biological Sequencing Data Through Kernel Selection," in *Bioinformatics Research and Applications*, vol. 14248, X. Guo, S. Mangul, M. Patterson, and A. Zelikovsky, Eds., Singapore: Springer Nature Singapore, 2023, pp. 442–452. doi: [10.1007/978-981-99-7074-2_35](https://doi.org/10.1007/978-981-99-7074-2_35).
- [25] M. Allaoui, S. B. Belhaouari, R. Hedjam, K. Bouanane, and M. L. Kherfi, "t-SNE-PSO: Optimizing t-SNE using particle swarm optimization," *Expert Systems with Applications*, vol. 269, p. 126398, Apr. 2025, doi: [10.1016/j.eswa.2025.126398](https://doi.org/10.1016/j.eswa.2025.126398).
- [26] D. Chicco, S. Melzi, F. Gasparini, and G. Jurman, "The advantages of our proposed Saturn coefficient over continuity and trustworthiness for UMAP dimensionality reduction evaluation," *PeerJ Computer Science*, vol. 12, p. e3424, Feb. 2026, doi: [10.7717/peerj-cs.3424](https://doi.org/10.7717/peerj-cs.3424).
- [27] Y. Wang, F. Sun, W. Lin, and S. Zhang, "AC-PCoA: Adjustment for confounding factors using principal coordinate analysis," *PLoS Comput Biol*, vol. 18, no. 7, p. e1010184, Jul. 2022, doi: [10.1371/journal.pcbi.1010184](https://doi.org/10.1371/journal.pcbi.1010184).
- [28] H. Jeon, J. Park, S. Shin, and J. Seo, "Stop Misusing t-SNE and UMAP for Visual Analytics," 2025, arXiv. doi: [10.48550/ARXIV.2506.08725](https://doi.org/10.48550/ARXIV.2506.08725).
- [29] X. Chen and Y. Yang, "Diffusion K-means clustering on manifolds: Provable exact recovery via semidefinite relaxations," *Applied and Computational Harmonic Analysis*, vol. 52, pp. 303–347, May 2021, doi: [10.1016/j.acha.2020.03.002](https://doi.org/10.1016/j.acha.2020.03.002).
- [30] F. de J. F. Arredondo, A. Ucan-Puc, and C. A. Noguez, "K-Means as a Radial Basis function Network: a Variational and Gradient-based Equivalence," 2026, arXiv. doi: [10.48550/ARXIV.2603.04625](https://doi.org/10.48550/ARXIV.2603.04625).
- [31] J. Mo, H. Shou, and W. Chen, "Implicit Surface Reconstruction via RBF Interpolation: A Review," *ENG*, vol. 16, no. 5, p. e070721194551, Sep. 2022, doi: [10.2174/1872212115666210707110903](https://doi.org/10.2174/1872212115666210707110903).
- [32] 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, p. e0318519, Feb. 2025, doi: [10.1371/journal.pone.0318519](https://doi.org/10.1371/journal.pone.0318519).
- [33] L. Abualigah, D. Yousefi, M. Abd Elaziz, A. A. Ewees, M. A. A. Al-qaness, and A. H. Gandomi, "Aquila Optimizer: A novel meta-heuristic optimization algorithm," *Computers & Industrial Engineering*, vol. 157, p. 107250, Jul. 2021, doi: [10.1016/j.cie.2021.107250](https://doi.org/10.1016/j.cie.2021.107250).
- [34] V. Kumar, J. K. Chhabra, and D. Kumar, "Differential Search Algorithm for Multiobjective Problems," *Procedia Computer Science*, vol. 48, pp. 22–28, 2015, doi: [10.1016/j.procs.2015.04.105](https://doi.org/10.1016/j.procs.2015.04.105).
- [35] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey Wolf Optimizer," *Advances in Engineering Software*, vol. 69, pp. 46–61, Mar. 2014, doi: [10.1016/j.advengsoft.2013.12.007](https://doi.org/10.1016/j.advengsoft.2013.12.007).
- [36] S. Suyanto, A. A. Ariyanto, and A. F. Ariyanto, "Komodo Mlipir Algorithm," *Applied Soft Computing*, vol. 114, p. 108043, Jan. 2022, doi: [10.1016/j.asoc.2021.108043](https://doi.org/10.1016/j.asoc.2021.108043).
- [37] M. F. Dzulkalnine, M. S. Mohamad, Y. W. Choon, M. A. Remli, and H. Alashwal, "Optimizing Ethanol Production in Escherichia Coli Using a Hybrid of Particle Swarm Optimization and Artificial Bee Colony," in *2022 The 6th International Conference on Advances in Artificial Intelligence*, Birmingham United Kingdom: ACM, Oct. 2022, pp. 140–146. doi: [10.1145/3571560.3571581](https://doi.org/10.1145/3571560.3571581).
- [38] S. Mirjalili and A. Lewis, "The Whale Optimization Algorithm," *Advances in Engineering Software*, vol. 95, pp. 51–67, May 2016, doi: [10.1016/j.advengsoft.2016.01.008](https://doi.org/10.1016/j.advengsoft.2016.01.008).
- [39] R. Ananda, K. M. Daud, and S. Zainudin, "RBI: a novel algorithm for regulatory-metabolic network model in designing the optimal mutant strain," *PeerJ Computer Science*, vol. 11, p. e2880, May 2025, doi: [10.7717/peerj-cs.2880](https://doi.org/10.7717/peerj-cs.2880).
- [40] R. Ananda, K. M. Daud, and S. Zainudin, "Non-Dominated Sorting Differential Search Algorithm for Optimizing Regulatory-Metabolic Networks by Using Probabilistic Approach," in *2023 International Conference on Electrical Engineering and Informatics (ICEEI)*, Bandung, Indonesia: IEEE, Oct. 2023, pp. 1–6. doi: [10.1109/ICEEI59426.2023.10346837](https://doi.org/10.1109/ICEEI59426.2023.10346837).
- [41] Y. F. Rachman, A. Syarif, D. L. Riatma, R. Khoirunnisa, N. A. Haqimi, and A. Faisal Sani, "Analyzing In-Game Purchase Behavior Using Predictive Analytics and Machine Learning," in *2025 5th International Conference of Science and Information Technology in Smart Administration (ICSINTESA)*, Yogyakarta, Indonesia: IEEE, Nov. 2025, pp. 317–322. doi: [10.1109/ICSINTESA68165.2025.11413625](https://doi.org/10.1109/ICSINTESA68165.2025.11413625).
- [42] S. Sharma and N. Desai, "Data-Driven Customer Segmentation Using Clustering Methods for Business Success," in *2023 4th IEEE Global Conference for Advancement in Technology (GCAT)*, Bangalore, India: IEEE, Oct. 2023, pp. 1–7. doi: [10.1109/GCAT59970.2023.10353367](https://doi.org/10.1109/GCAT59970.2023.10353367).
- [43] T. Wahyuningsih, D. Manongga, and I. Sembiring, "Clustering student argumentation types by implementing a multi-document clustering model: a combination of Pytorch and BERT," *PeerJ Computer Science*, vol. 12, p. e3431, Jan. 2026, doi: [10.7717/peerj-cs.3431](https://doi.org/10.7717/peerj-cs.3431).
- [44] Primadona, "DIGITAL MSME DEVELOPMENT IN INDONESIA DURING THE COVID-19 PANDEMIC," *jinn*, vol. 2, no. 2, pp. 145–154, Aug. 2020, doi: [10.54783/jinn.v2i2.505](https://doi.org/10.54783/jinn.v2i2.505).
- [45] C. Sianturi and J. M. Simanjuntak, "Pengaruh Strategi Pemasaran Terhadap Peningkatan Volume Penjualan Lahan di PT. Kawasan Industri Modern (Persero) Medan," *Jurnal Ilmu Sosial Dan Politik*, vol. 1, no. 2, pp. 205–224, Dec. 2021, doi: [10.51622/jispol.v1i2.418](https://doi.org/10.51622/jispol.v1i2.418).
- [46] D. Darmawan and E. Grenier, "Competitive Advantage and Service Marketing Mix," *jos3*, vol. 1, no. 2, pp. 75–80, Jul. 2021, doi: [10.56348/jos3.v1i2.9](https://doi.org/10.56348/jos3.v1i2.9).
- [47] R. Muqarrabin, Y. Zebua, and S. L. Nasution, "Analysis of Product Quality, Price, Promotion, and Location Toward

- Purchase Decision,” *ijjse*, vol. 5, no. 2, pp. 720–730, Jul. 2022, doi: [10.31538/ijjse.v5i2.2088](https://doi.org/10.31538/ijjse.v5i2.2088).
- [48] H. Daoudi and E. I. Develi, “Impact of gender and monthly income on consumer buying behavior,” *Journal of International Trade, Logistics and Law*, vol. 9, no. 1, pp. 241–250, May 2023, Accessed: Mar. 18, 2026. [Online]. Available: <https://www.jital.org/index.php/jital/article/view/367>
- [49] F. Alfianur and Y. Kadono, “The effects of gender and age on factors that influence purchase intentions and behaviours of e-commerce consumers in Indonesia,” *IJIL*, vol. 31, no. 4, p. 474, 2022, doi: [10.1504/IJIL.2022.123178](https://doi.org/10.1504/IJIL.2022.123178).
- [50] Andara Selviranih, Siti Aidini Khoerunnisa, Hilda Sapitri, Ibnu Muhammad Sabilli, Muhammad Syehan, and Nurul Aulia Fitriani, “Pengaruh Store Layout dan Motivasi Rasional terhadap Keputusan Pembelian Konsumen di Toko Abu Rohmah,” *JUPSIM*, vol. 4, no. 2, pp. 419–429, Jun. 2025, doi: [10.55606/jupsim.v4i2.4541](https://doi.org/10.55606/jupsim.v4i2.4541).
- [51] S. Mulyana, “Pengaruh Harga dan Ulasan Produk Terhadap Keputusan Pembelian Produk Fashion Secara Online Pada Shopee di Pekanbaru,” *dayasaing*, vol. 7, no. 2, pp. 185–195, Jul. 2021, doi: [10.35446/dayasaing.v7i2.665](https://doi.org/10.35446/dayasaing.v7i2.665).
- [52] E. Purwanto, “PENGARUH BAURAN PEMASARAN TERHADAP KEPUTUSAN PEMBELIAN SPONTAN,” *MISSY*, vol. 4, no. 1, pp. 77–83, Aug. 2023, doi: [10.24929/missy.v4i1.2760](https://doi.org/10.24929/missy.v4i1.2760).
- [53] Y. Fei, Y. Liu, C. Jia, Z. Li, X. Wei, and M. Chen, “A Survey of Geometric Optimization for Deep Learning: From Euclidean Space to Riemannian Manifold,” *ACM Comput. Surv.*, vol. 57, no. 5, pp. 1–37, May 2025, doi: [10.1145/3708498](https://doi.org/10.1145/3708498).
- [54] R. Ananda and A. Prasatiadi, “Hierarchical and K-means Clustering in the Line Drawing Data Shape Using Procrustes Analysis,” *JOIV : Int. J. Inform. Visualization*, vol. 5, no. 3, p. 306, Sep. 2021, doi: [10.30630/joiv.5.3.532](https://doi.org/10.30630/joiv.5.3.532).
- [55] C. M. Sha, J. Wang, and N. V. Dokholyan, “Predicting 3D RNA structure from the nucleotide sequence using Euclidean neural networks,” *Biophysical Journal*, vol. 123, no. 17, pp. 2671–2681, Sep. 2024, doi: [10.1016/j.bpj.2023.10.011](https://doi.org/10.1016/j.bpj.2023.10.011).
- [56] R. Ananda, D. Rachmawaty, B. Pratikno, O. A. Hamid, and M. B. Mohd Amin, “Improved Clustering-Based Feature Selection Using Feature Extraction Based on Principal Component Analysis,” in *Innovations of Intelligent Informatics, Networking, and Cybersecurity*, vol. 2329, S. O. Al-Mamory, A. Al-Sherbaz, T. Kanakis, A. S. Albahri, W. S. Bhaya, E. S. Alshamery, A. A. Abdullah, A. Al-Ajeli, and S. Z. Alrashid, Eds., Cham: Springer Nature Switzerland, 2025, pp. 15–38. doi: [10.1007/978-3-031-81065-7_2](https://doi.org/10.1007/978-3-031-81065-7_2).
- [57] N. Sumbherwal, B. K. Hooda, and P. K. Vinit, “Performance Analysis of Distance Measures for Mixed-Variables Data,” Dec. 28, 2023. doi: [10.21203/rs.3.rs-3749138/v1](https://doi.org/10.21203/rs.3.rs-3749138/v1).
- [58] H. Liu and Y. Yao, “Enhancing Student Performance Prediction with Graph Representation Learning Based on Gower Distance,” in *2024 5th International Conference on Information Science and Education (ICISE-IE)*, Zhanjiang, China: IEEE, Dec. 2024, pp. 20–23. doi: [10.1109/ICISE-IE64355.2024.11025450](https://doi.org/10.1109/ICISE-IE64355.2024.11025450).
- [59] I. El Badisy, N. Graffeo, M. Khalis, and R. Giorgi, “Multi-metric comparison of machine learning imputation methods with application to breast cancer survival,” *BMC Med Res Methodol*, vol. 24, no. 1, p. 191, Aug. 2024, doi: [10.1186/s12874-024-02305-3](https://doi.org/10.1186/s12874-024-02305-3).
- [60] P. Delicado and C. Pachón-García, “Multidimensional scaling for big data,” *Adv Data Anal Classif*, vol. 19, no. 3, pp. 649–670, Sep. 2025, doi: [10.1007/s11634-024-00591-9](https://doi.org/10.1007/s11634-024-00591-9).
- [61] M. Shutaywi and N. N. Kachouie, “Silhouette Analysis for Performance Evaluation in Machine Learning with Applications to Clustering,” *Entropy*, vol. 23, no. 6, p. 759, Jun. 2021, doi: [10.3390/e23060759](https://doi.org/10.3390/e23060759).
- [62] P. Mishra, C. Pandey, U. Singh, A. Gupta, C. Sahu, and A. Keshri, “Descriptive statistics and normality tests for statistical data,” *Ann Card Anaesth*, vol. 22, no. 1, p. 67, 2019, doi: [10.4103/aca.ACA_157_18](https://doi.org/10.4103/aca.ACA_157_18).
- [63] G. Keklik and M. Şahin, “Multidimensional Scaling Analysis and an Application,” *Black Sea Journal of Engineering and Science*, vol. 5, no. 2, pp. 71–78, Apr. 2022, doi: [10.34248/bsengineering.1077162](https://doi.org/10.34248/bsengineering.1077162).
- [64] J. Del Ser et al., “More is not Always Better: Insights from a Massive Comparison of Meta-heuristic Algorithms over Real-Parameter Optimization Problems,” in *2021 IEEE Symposium Series on Computational Intelligence (SSCI)*, Orlando, FL, USA: IEEE, Dec. 2021, pp. 1–7. doi: [10.1109/SSCI50451.2021.9660030](https://doi.org/10.1109/SSCI50451.2021.9660030).
- [65] R. Ananda, K. Mohd Daud, and S. Zainudin, “Enhanced integrated reliability-based algorithm for identifying the optimal gene knockout in strain optimization,” *Syst Microbiol and Biomanuf*, vol. 6, no. 2, p. 37, Apr. 2026, doi: [10.1007/s43393-025-00431-7](https://doi.org/10.1007/s43393-025-00431-7).
- [66] F. L. Gewers et al., “Principal Component Analysis: A Natural Approach to Data Exploration,” *ACM Comput. Surv.*, vol. 54, no. 4, pp. 1–34, May 2022, doi: [10.1145/3447755](https://doi.org/10.1145/3447755).