Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 6, No. 3, Juni 2025, Page. 1405-1418
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.3.4687

Marketing Analysis of Shoe Products Using Principal Coordinates Analysis and K-Means Clustering Based on the Marketing Mix at Bintang Sepatu Purwokerto MSME

Samuel Sinaga[1], Ridho Ananda*[2], Halim Qista Karima[3], Adrus Mohamad Tazuddin[4]

[1,2,3]Industrial Engineering, Telkom University, Indonesia
[2]Center for Artificial Intelligence Technology, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM, Bangi Selangor, Malaysia
[4]Centre for Pre-University Studies, Universiti Malaysia Sarawak, Sarawak, Malaysia

Email: [1]ridho@ittelkom-pwt.ac.id

Received : May 5, 2025; Revised : Jun 3, 2025; Accepted : Jun 15, 2025; Published : Jun 23, 2025

Abstract

Bintang Sepatu Purwokerto MSME is a micro, small, and medium enterprise engaged in the production of local shoes. Recently, this MSME faced a significant issue in the marketing aspect, namely the low achievement of sales targets. Consequently, inventory will accumulate in the warehouse. Accordingly, this research aimed to formulate targeted marketing strategies by clustering customers based on demographic and marketing mix influencing purchasing behavior. This study applied principal coordinate analysis (PCoA) and k-means clustering to manage categorical and numerical data types within the dataset comprising 179 customers and 16 attributes.. The PCoA algorithm was utilized to derive object configurations that were subsequently employed in k-means. The clustering result produced three clusters with good clustering quality based on the Silhouette score, namely 0.790, indicating accurate and representative segmentation. Each cluster obtained had a different customer characteristic. The first cluster, comprising 68 customers (38%), was oriented towards fundamental needs and tended to shop traditionally, classified as a segment of conventional rational customers. Additionally, the second cluster, with 70 customers (39%), exhibited planned and stable decision-making, categorized as mature rational customers. Furthermore, the third cluster comprises 41 customers (23%) who are digitally aware and combine conventional shopping approaches with technological utilization, identified as rational consumers. The segmentation results provide a data-driven foundation for designing targeted marketing strategies, thereby potentially increasing sales, supporting the sustainability of MSMEs, and encouraging the application of unsupervised learning techniques in decision-making processes.

Keywords : Clustering, K-Means, Marketing, PCoA.

1. INTRODUCTION

Micro, Small, and Medium Enterprises (MSMEs) are business units or economic activities that stand alone, managed by individuals or business entities that are not legal entities. MSMEs can operate in various economic sectors, such as agriculture, manufacturing, trade, and services [1]. MSMEs have a very significant role in Indonesia's economic growth, considering that the number reaches 99% of the total existing business units. In 2023, the number of MSME businesses reached around 66 million. The contribution of MSMEs reached 61% of Indonesia's Gross Domestic Product (GDP), equivalent to IDR 9,580 trillion. MSMEs absorb around 117 million workers (97%) of the total workforce [2]. One of the fastest-growing industries in the MSME sector is fashion, including shoe production, which has high market demand [3].

MSMEs in Banyumas Regency encompass 84,350 business units across various sectors, including the shoe industry [4]. One of the MSMEs in Banyumas engaged in this sector is Bintang Sepatu Purwokerto. This MSME produces leather shoes in various styles for both men and women. In terms of

Vol. 6, No. 3, Juni 2025, Page. 1405-1418
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.3.4687

marketing, it relies on a door-to-door sales strategy. Although the number of salespeople increased from one in May 2023 to three by the end of the year, the main challenge faced was that the products often did not match customer preferences, particularly in size and design. As a result, sales volumes fell short of targets, and the production capacity set at 500 pairs per month was not fully utilized, as illustrated in Figure 1. Consequently, an appropriate marketing approach is essential to address this issue. One of appropriate strategies to be implemented is to comprehend customer characteristics to inform marketing decisions. Prior research demonstrated that analyzing customer characteristics with the components of the marketing mix (product, price, place, promotion, process, and licensing) facilitated the formulation of effective marketing strategies, resulting in increased sales volume [5] - [6]. However, prior studies predominantly employed conventional segmentation methods and tended to analyze either numerical or categorical data in isolation, thereby constraining their capacity to comprehensively capture the complexity of customer profiles.

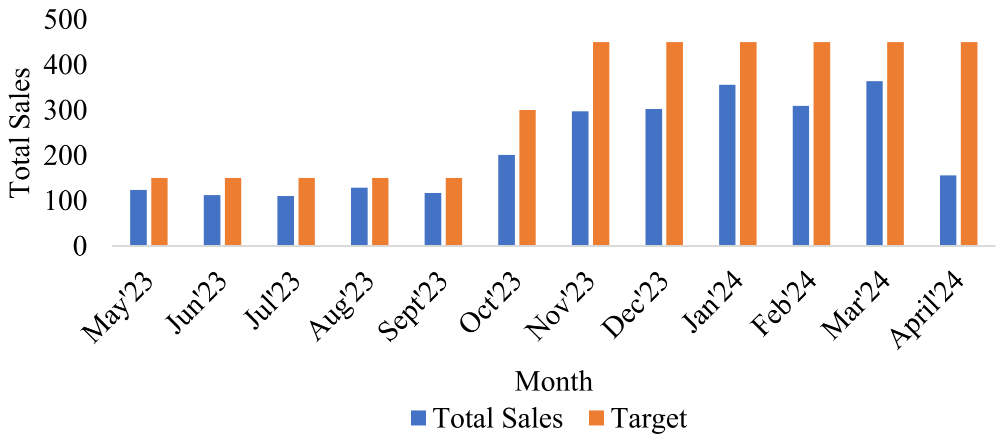


Figure 1. Product Sales Data

One of the methods that can be conducted to analyze customer characteristics is clustering analysis, where k-means is the popular algorithm. This algorithm can be leveraged to cluster objects into several clusters based on observed features, so that the higher similarity of objects in the same cluster is obtained compared to objects from different clusters. The k-means algorithm has been utilized in several fields such as education [7] - [11], environment [12], shape analysis [13], biology [14], feature selection [15], and marketing [16] - [17].

In the clustering process, K-means clustering necessitates the adjustment of each cluster's centroid according to its constituent object. If all features in the dataset are numeric, the data can serve as an object configuration, with the centroid often adjusted using Euclidean distance. Nevertheless, if the dataset comprises both categorical and numeric data, a particular methodology is required to ascertain the configuration of each object inside the dataset. One method that can be employed is the Principal Coordinate Analysis (PCoA) algorithm [15]. This approach can generate the configuration of objects using their proximity matrix. In addition, PCoA facilitates the dimensionality reduction of customer data, enabling highly correlated variables to be depicted in fewer dimensions while preserving essential information[18] - [19].

Although this approach holds significant potential, to date, few studies have integrated the Principal Coordinate Analysis (PCoA) algorithm with k-means clustering for customer segmentation in the MSME sector, particularly within the footwear industry, while considering the full spectrum of mixed attributes in the marketing mix. This situation highlights a research gap where conventional clustering methods have not been able to effectively handle mixed-type marketing data. The novelty of

Vol. 6, No. 3, Juni 2025, Page. 1405-1418
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.3.4687

this study lies in the integration of the Principal Coordinate Analysis (PCoA) algorithm and k-means clustering to analyze customer characteristics involving both categorical and numerical attributes based on the marketing mix, with a specific application in the footwear MSME context. This approach enables the development of a more accurate and comprehensive segmentation model, thereby more effectively reflecting customer preferences compared to conventional methods.

Based on the aforementioned statements, this study aims to analyze the customer characteristics of Bintang Sepatu Purwokerto MSME and categorize them into more specific segments based on their characteristics and the elements of the marketing mix (product, price, place, promotion, process, and licensing) by using clustering analysis. The PCoA algorithm will be utilized to handle the numerical and categorical data simultaneously when constructing the configuration of objects for the k-means algorithm. This research is expected to provide data-driven recommendations to enhance Bintang Sepatu Purwokerto MSME's comprehension of client needs. The results of this study will facilitate the formulation of a more focused marketing plan, enhance sales volume, and optimize manufacturing capacity, thus allowing the MSME to enhance its competitiveness in the intensifying shoe sector.

2. MATERIAL AND METHOD

2.1. The Dataset Used

This study utilized a dataset derived from questionnaire responses of 179 customers, comprising 16 features. The dataset contains two types of data, namely categorical and numeric. Features comprising categorical data encompass address, gender, age, occupation, income, purchase motivation, purchase frequency, preferred model, color, place of purchase, duration of social media use, promotional offers, product information search behavior, and licensing. Meanwhile, the features comprising numeric data include shoe size and price. A straightforward representation of the dataset utilized in this study is presented in Table 1.

Table 1. Research Data

Object	Address	Gender	Age	…	Information	Licensing
N1	Banyumas	Male	36 - 45 Years	…	Facebook	No
N2	Banyumas	Fermale	26 - 35 Years	…	Instagram	Yes
N3	Banyumas	Fermale	36 - 45 Years	…	Shopee	Yes
N4	Banyumas	Male	15 - 25 Years	…	TikTok	Yes
N5	Banyumas	Male	36 - 45 Years	…	Facebook	No
…	…	…	…	…	…	…
N178	Pemalang	Fermale	26 - 35 Years	…	Lazada	Yes
N179	Pemalang	Fermale	36 - 45 Years	…	Friends / Neighbors	Yes

2.2. The Proximity Measure

A proximity measure in this study is utilized to quantify the dissimilarity between two objects. The greater the measure of proximity between two objects, the less likely they are to be categorized in the same cluster. Conversely, the smaller the proximity measure between two objects, the smaller the difference, which means the more likely they are to cluster the two objects in the same group. The Euclidean distance is a widely utilized proximity measure [20] . This distance is extensively utilized in optimization research [21] [22] and data mining [23]. The Euclidean distance formula is provided in equation 1.

$$\boldsymbol{d_E}(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{\sum_{k=1}^{p} \left(x_{ik} - x_{jk}\right)^2} \quad (1)$$

Where $\boldsymbol{d_E}(\mathbf{x}_i, \mathbf{x}_j)$ is the Euclidean distance between the $\boldsymbol{i}$-th and the $\boldsymbol{j}$-th objects. Meanwhile, $\boldsymbol{x_{ik}} \in \mathbf{x}_i, \forall \boldsymbol{k}$.

For categorical data, the proximity measure is inappropriate when computed using Euclidean distance. Therefore, other approaches is required for calculating this measure. One of approaches that can be leveraged to calculate the proximity measure in this type data is the Gower distance[24]. This distance have been implemented in several fields, such as education [25] and imputation data [26]. The Gower distance formula utilized in this study is presented in Equation 2.

$$\boldsymbol{d_G}(\mathbf{x}_i, \mathbf{x}_j) = \sum_{k=1}^{p} \left(1 - S_{ij,k}\right) \quad (2)$$

$$S_{ij,k} = \begin{cases} 1 & \text{if } x_{ik} = x_{jk} \\ 0 & \text{Otherwise} \end{cases} \quad (3)$$

Where $d_G(\mathbf{x}_i, \mathbf{x}_j)$ is the Gower distance between the i-th and the j-the objects. Then, $S_{ij,k}$ is the similarity measure between them.

When the dataset includes both categorical and numeric data, the computation of the proximity measure between objects is conducted independently for each data type. This study used Euclidean distance for numerical data types. Conversely, the Gower distance was employed for categorical data types. Suppose that ${}_n\mathbf{X}_p$ represent an initial dataset of n objects and p variables, where p_1 denotes the count of category variables and p_2 signifies the count of numeric variables, such that $p_1 + p_2 = p$. The procedure for calculating the proximity measure between objects describes as follows:

1. Split the dataset according to its data type, ${}_n\mathbf{X}_p = \left[{}_n\mathbf{Y}_{p_1} \quad {}_n\mathbf{Z}_{p_2}\right]$, where ${}_n\mathbf{Y}_{p_1} = \left(y_{ij}\right)$ and ${}_n\mathbf{Z}_{p_2} = \left(z_{ij}\right)$.
2. Compute the proximity matrix for each partition, yielding $\mathbf{D}_y = \left(d_{(y)ij}\right)$ and $\mathbf{D}_z = \left(d_{(z)ij}\right)$. The $d_{(y)ij} = d_G(\mathbf{y}_i, \mathbf{y}_j)$ and $d_{(z)ij} = d_E(\mathbf{z}_i, \mathbf{z}_j)$ elements are calculated by Gower and Euclidean distance formulas, respectively.
3. Normalize $\mathbf{D}_y$ and $\mathbf{D}_z$.
4. Calculate the proximity matrix of the dataset, namely $\mathrm{D} = \left(d_{\mathrm{ij}}\right)$, using Equation 4. This procedure utilizes weights based on the quantity of categorical or numeric attributes [31].

$$\boldsymbol{d_{\mathrm{ij}}} = \frac{d_{(y)ij}}{p_1} + \frac{d_{(z)ij}}{p_2}, \forall \boldsymbol{i}, \boldsymbol{j} \quad (4)$$

2.3. Principal Coordinates Analysis (PCoA)

Principal coordinates analysis (PCoA) is a method employed to determine the object configurations inside a low-dimensional space, such as two or three dimensions, representing the dissimilarity measure of objects in the proximity matrix [22]. Principal Coordinates Analysis (PCoA) was selected due to its ability to determine object configurations in a low-dimensional space by utilizing a proximity matrix derived from various types of data, including both numerical and categorical attributes. In contrast to Principal Component Analysis (PCA), which only processes numerical data and relies on a covariance matrix, PCoA offers greater flexibility in handling mixed data types a common characteristic in marketing research where customer attributes are diverse. Furthermore, other dimensionality reduction techniques such as t-SNE and Multidimensional Scaling (MDS) tend to be more complex and provide less support for the geometric interpretation required by clustering methods

P-ISSN: 2723-3863
https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871
DOI: https://doi.org/10.52436/1.jutif.6.3.4687

like k-means, which assume a Euclidean space. Therefore, PCoA was chosen because it accommodates mixed data types while producing object configurations in a Euclidean space compatible with distance-based clustering algorithms. The steps of the PCoA algorithm are outlined as follows:

1. Generate matrix $\mathbf{A} = (a_{ij})$ utilizing the proximity matrix ($\mathbf{D}$) using with Equation 5.

$$\boldsymbol{a_{ij}} = -\frac{1}{2}\, \boldsymbol{d_{ij}^2}, \forall \boldsymbol{i,j} \tag{5}$$

2. Calculate $\mathbf{B} = (b_{ij})$ using Equation 6.

$$\boldsymbol{b_{ij}} = \boldsymbol{a_{ij}} - \boldsymbol{\bar{a}_{i.}} - \boldsymbol{\bar{a}_{.j}} + \boldsymbol{\bar{a}_{..}} \tag{6}$$

$$\boldsymbol{\bar{a}_{i.}} = \sum_{j=1}^{n} \frac{\boldsymbol{a_{ij}}}{\boldsymbol{n}}; \boldsymbol{\bar{a}_{.j}} = \sum_{i=1}^{n} \frac{\boldsymbol{a_{ij}}}{\boldsymbol{n}}; \textbf{ and } \boldsymbol{\bar{a}_{..}} = \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\boldsymbol{a_{ij}}}{\boldsymbol{n}}$$

3. Decompose the B matrix leveraging the spectral decomposition process. The result of decomposition is exhibited Equation 7.

$$\mathbf{B} = \mathbf{V\Lambda V'} = \mathbf{V\Lambda^{1/2}\Lambda^{1/2}V'} \tag{7}$$

Where $\mathbf{\Lambda}$ and $\mathbf{V}$ respectively are the ordered matrix of eigenvalues and eigenvalues of $\mathbf{B}$, where $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, .., \lambda_n)$ with $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq 0$.

4. Determine ${}_n\widetilde{\mathbf{X}}_n = \mathbf{V\Lambda}^{1/2}$, namely a matrix representing the configuration of each object in n dimensions.
5. For obtaining this configuration in the low dimensions, for instance two dimensions (k = 2), the first two columns is selected, namely ${}_n\widetilde{\mathbf{X}}_2 = \mathbf{V}_2\mathbf{\Lambda}_2^{1/2}$. The criteria outlined in Equation 8 should be obeyed for choosing the low dimension.

$$\frac{\sum_{i=1}^{k}\lambda_i}{\sum_{j=1}^{q}\lambda_j}\ \boldsymbol{x}\ \mathbf{100}\ \% \ \geq \mathbf{70\%} \tag{8}$$

2.4. Elbow

The Elbow method is a technique for determining the optimal number of clusters by analyzing the percentage of variance explained as the number of clusters (k) increases. The optimal point is identified from a graph where the line forms an 'elbow' or right angle, indicating that the addition of subsequent clusters does not provide significant improvement in clustering results [27]. The optimal k value is determined based on the elbow point in the cluster relationship graph by identifying where the Sum of Squared Errors (SSE) begins to decrease at a slower rate [28]. The formula of SSE is shown in Equation 9.

$$\mathbf{SSE} = \sum_{k=1}^{K} \sum_{\mathbf{x}_i \in S_k} \|\mathbf{x}_i - \mathbf{c}_k\|_2^2 \tag{9}$$

Where, $\boldsymbol{K}$ represents the quantity of clusters. Furthermore, $\mathbf{S}_{\boldsymbol{k}}$ denotes the set of objects contained within cluster $\boldsymbol{k}$. Additionally, $\mathbf{c}_{\boldsymbol{k}}$ represents the centroid of cluster $\boldsymbol{k}$.

2.5. K-Means Clustering

K-Means is a partition-based clustering method that groups data based on similar characteristics by minimizing the distance of each data to the cluster center (centroid) iteratively [29]. This algorithm determines the number of clusters (k) first, then places the data into the nearest cluster using Euclidean

Vol. 6, No. 3, Juni 2025, Page. 1405-1418
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.3.4687

distance or other metrics until it converges [30]. The following are the steps in the basic K-Means algorithm.

1. The first step in the K-Means algorithm is to determine the number of clusters (k).
2. The second step is to determine a random initial value for the centroid..
3. The third step is to assign each data point or object to the nearest centroid. In general, the closest distance is determined using the Euclidean distance formula, as shown in Equation 1.
4. Then, repeat the process until the criteria is satisfied, namely when the new centroid values are identical to the previous centroid values.

2.6. Research Flow

The research flow in this study consists of four primary stages: planning, analysis, interpretation, and conclusion. The planning stage encompasses data collection and preprocessing. During the preprocessing phase, categorical data is converted from its original form into numerical representations. Meanwhile, the numerical data is standardized. Then, the analysis phase involves computing the proximity matrix, object configuration, and object clustering through the PCoA and K-means algorithms. Furthermore, the interpretation stage focuses on identifying clusters and providing marketing recommendations from the clustering results. The last stage entails deriving conclusions based on the interpretations. Figure 2 illustrates an overview of the study process.

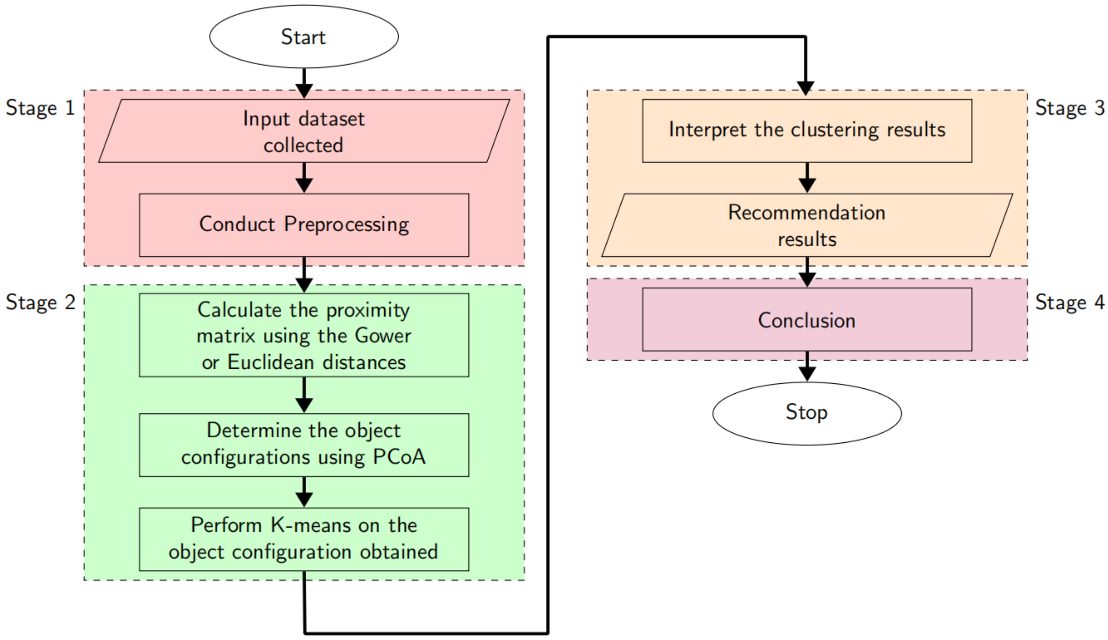


Figure 2. Research Flow in This Study

The software used for data processing includes MATLAB 2021b and Microsoft Excel 365. The questionnaire data were initially compiled in Excel, which was selected for its effectiveness in managing, organizing, and performing preliminary cleaning of survey responses. Principal Coordinates Analysis (PCoA) and K-means clustering were conducted using MATLAB. MATLAB was chosen for its advanced computational capabilities and extensive library of built-in functions, which facilitate complex mathematical procedures and multidimensional data analysis. Its ability to efficiently and accurately handle clustering techniques such as PCoA and K-means makes it particularly well-suited to meet the analytical requirements of this study.

Jurnal Teknik Informatika (JUTIF) Vol. 6, No. 3, Juni 2025, Page. 1405-1418
P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.6.3.4687

3. RESULTS

The questionnaire data obtained, as shown in Table 1, underwent initial processing through the transformation of categorical variables. The categorical attributes include 5 address categories, 2 genders, 4 age ranges, 33 types of occupations, 4 income brackets, 2 purchase motivations, 5 purchase frequencies, 29 shoe models, 5 shoe colors, 2 places of purchase, 5 social media usage durations, 5 types of promotional offers, 9 sources of product information, and 2 licensing types. Meanwhile, numerical data were standardized using min-max normalization to scale the values between 0 and 1, ensuring consistency in magnitude and unit of measurement.

Table 2. Data Transformation And Standardization Results

Object	Adress	…	Model	Size	Color	Price	...	Information	Licensing
N1	1.000	…	4.000	0.714	1.000	0.636	...	2.000	1.000
N2	1.000	…	30.000	0.429	3.000	0.273	...	3.000	2.000
N3	1.000	…	33.000	0.714	5.000	0.273	...	5.000	2.000
N4	1.000	…	26.000	0.571	2.000	0.182	...	7.000	2.000
N5	1.000	…	9.000	0.571	1.000	0.273	...	2.000	0.000
…	…	…	…	…	…	…	...	…	…
N178	5.000	…	5.000	0.286	1.000	0.455	...	4.000	2.000
N179	5.000	…	8.000	0.429	1.000	0.455	...	6.000	2.000

The subsequent stage involves the computation of the proximity matrix. The dataset had both categorical and numeric data types; hence, the method for computing the proximity matrix outlined in Section 2 was used by utilizing Gower or Euclidean distances, depending on the data type. The proximity matrix obtained was illustrated in Table 3.

Table 3. Proximity Matrix

	N1	N2	N3	…	N178	N179
N1	0.000	0.277	0.233	…	0.279	0.200
N2	0.277	0.000	0.179	…	0.156	0.137
N3	0.233	0.179	0.000	…	0.284	0.215
N4	0.289	0.120	0.136	…	0.249	0.215
N5	0.216	0.122	0.117	…	0.205	0.146
…	…	…	…	…	…	…
N178	0.279	0.156	0.284	…	0.000	0.102
N179	0.200	0.137	0.215	…	0.102	0.000

Subsequently, the object configuration was determined by utilizing the proximity matrix obtained. Here, the principal coordinate analysis (PCoA) algorithm was conducted. Figure 3 visualized the configuration of objects in two dimensions, where the 1st and 2nd principal components obtained were 53.79% and 21.42%, respectively. Based on those principal component values, the visualization encompasses 75.21% of the information from the initial proximity matrix. As a result, according to Equation 8, the quality of the configuration in the second dimension is satisfactory.

Vol. 6, No. 3, Juni 2025, Page. 1405-1418
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.3.4687

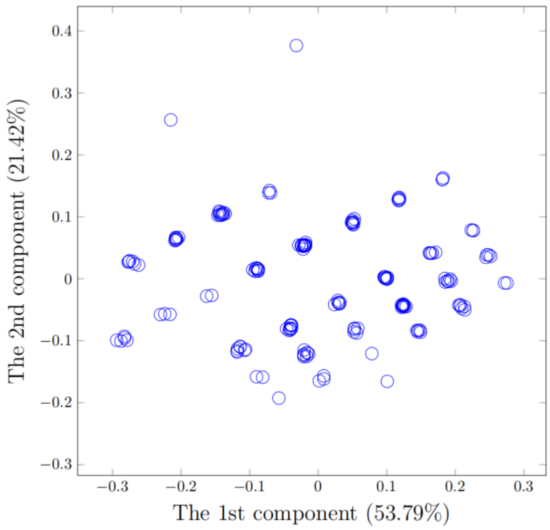
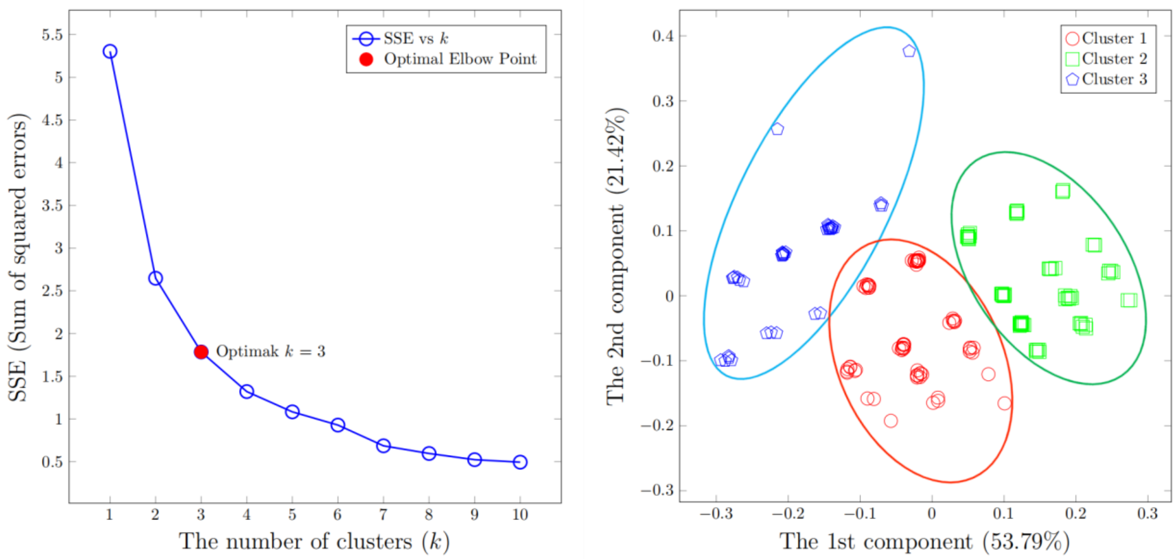


Figure 3. Visualization of the object configuration in the two dimensions

The gained object configuration was subsequently employed to cluster each object utilizing the k-means approach. Before the main clustering process, the optimal number of clusters was determined using the Elbow criterion. The Elbow criterion used indicates that the best number of clusters was three, with a Sum of Squared Errors (SSE) value of 1.784, as shown in Figure 4A. Meanwhile, the k-means results in two dimensions were visualized in Figure 4B. The results of the K-means cluster analysis have been summarized in Table 4.

Table 4. Members Of Each Cluster

Cluster	Number of Cluster Members	Persentage
1	68	38%
2	70	39%
3	41	23%



(a) (b)

Figure 4. The representation of the clustering results in a low dimension. Figure A illustrates the optimal k value determined by the elbow criterion. Additionally, Figure B demonstrates the k-means clustering results.

Vol. 6, No. 3, Juni 2025, Page. 1405-1418
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.3.4687

The results of the clustering analysis using the K-Means method yielded three clusters. Cluster one consists of 68 customers, cluster two consists of 70 customers, and cluster three consists of 41 customers. The variation in the number of members across the three clusters reflects differences in customer characteristics, which are grouped based on the average centroid values [31]. The K-Means method works by grouping data based on the closest distance of each customer to the cluster center. As a result, differences in cluster sizes may indicate uneven data distribution [32]. The imbalance in cluster sizes suggests that customer data is not evenly distributed, with specific patterns emerging in certain groups. This variation provides an early indication that customers exhibit different behaviors and needs, which should be understood in greater depth to better target marketing strategies.

The quality of this cluster is further supported by a silhouette index value of 0.790, A silhouette index value of 0.790 statistically indicates a very strong level of internal validity in the clustering process. This index, which ranges from 0 to 1, measures the extent to which each data point is closely associated with its own cluster (cohesion) compared to the nearest neighboring cluster (separation). A value of 0.790 signifies that the average intra-cluster distance is significantly smaller than the inter-cluster distance, implying that the cluster structures are clearly and consistently formed. Statistically, this value falls within the "very strong" category [33]. indicating that the objects within the dataset are not only compactly grouped but also distinctly separated from other clusters. This demonstrates that the clustering model exhibits a high discriminative capability, minimizes misclassification, and accurately represents the natural structure of the data. As such, the clustering results can be considered highly reliable for further analysis.

A summary table comparing the characteristics of customers within each cluster is presented in Table 5.

Table 5. The Characteristics Of Customers In Each Cluster

Characteristics	Cluster 1	Cluster 2	Cluster 3
Address	Banyumas and Pemalang	Purbalingga	Banjarnegara and Pemalang
Gender	Female	Female	Male
Age	26 – 35 Years	36 - 45 Years	36 – 45 Years
Occupation	Teacher	Teacher	Teacher
Income	IDR 1 – 3 million	IDR 3 – 6 million	IDR 3 – 6 million
Purchase Motivation	Rational	Rational	Rational
Purchase Frequency	One time	One time	One to two times
Place of Purchase	Offline	Offline	Offline
Digital Behavior	Searches for product info via Shopee and Facebook	Searches for product info via Shopee and Instagram	Searches for product info via Shopee and TikTok
Communication Preference	WhatsApp	WhatsApp	WhatsApp

Cluster one represents the 'conventional rational' segment, where purchasing decisions are driven by practical needs, with a focus on price, quality, and functionality rather than emotional factors or trends. This segment is predominantly composed of women aged 26-35 from Banyumas and Pemalang who work as teachers. They are typically financially stable and well-educated, seeking affordable products to meet their daily needs. Their preference for offline shopping reflects a desire for hands-on experiences and trust in traditional retail, although their habit of researching products online beforehand

Vol. 6, No. 3, Juni 2025, Page. 1405-1418
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.3.4687

indicates that they are digitally savvy and value informed choices. They also respond well to impersonal communication through WhatsApp or social media direct messaging.

Cluster two is categorized as part of the 'mature rational customer' segment, characterized by traditional preferences. This cluster consists of women aged 36-45, primarily from Purbalingga, most of whom are teachers with middle incomes. These customers make purchase decisions based on personal needs, avoiding emotional or impulsive motivations. They prefer shoes that combine functionality, comfort, and a classic or elegant design, reflecting good taste and a focus on practicality and suitability for professional activities. Their firm approach to price, preference for neutral colors like black, and attention to general shoe size options indicate the consistency and stability of their preferences. While they typically shop in traditional stores, they are also open to researching products online before making a purchase, demonstrating rational thinking and a desire for a hands-on shopping experience. Their openness to receiving private messages via WhatsApp or direct messages on social media underscores their preference for professional and respectful communication with brands.

Cluster three is categorized as 'digital-aware rational male customers' with conventional shopping behavior. The main characteristics of this group include need-based purchasing, a preference for convenience and functionality, and planned purchasing patterns (evidenced by two purchases that indicate satisfaction and initial loyalty). This cluster consists of financially stable adult males, aged 36-45, primarily living in Banjarnegara and Pemalang, who work as teachers. They tend to prefer formal or semi-formal shoe models in neutral colors, reflecting a professional and practical lifestyle. While they are more comfortable shopping directly at physical stores, their reference-seeking behavior on platforms like Shopee shows that they are not unfamiliar with technology and use it to inform their purchasing decisions. They are also active on social media and open to receiving promotional information delivered personally, as long as it is relevant and unobtrusive. This indicates a willingness to engage in digital communication, making them an ideal segment for marketing strategies that blend a personal touch with an informative digital approach. Based on the clustering results from K-means, this provides great opportunities for Purwokerto Shoe Star MSMEs to design more specific and personalized marketing strategies tailored to each cluster.

4. DISCUSSION

Based on the clustering analysis that has been conducted, an appropriate marketing strategy can be formulated as follows:

1) Cluster One

The ideal marketing strategy for this cluster should focus on an informative, personalized, and balanced approach that integrates both online and traditional shopping. Since these customers make purchase decisions based on practical needs and with caution, marketing messages should emphasize the value of products in terms of convenience, durability, and affordability, while providing easily digestible content such as infographics or authentic customer testimonials. Given their preference for in-person shopping, the in-store experience should be made convenient, informative, and efficient. At the same time, their online information-seeking behavior can be leveraged by offering digital catalogs on platforms such as Shopee or official websites, ensuring they are comprehensive and easy to navigate.

To strengthen customer relationships, MSME can use personalized communication methods via WhatsApp or direct social media messages. These communications can include updates on new products, limited-time promotions, or notifications about the availability of products matching their preferences. Consistent messaging and services across all channels will help build trust and foster long-term loyalty within this segment.

2) Cluster Two

Vol. 6, No. 3, Juni 2025, Page. 1405-1418
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.3.4687

A suitable marketing strategy for this cluster is to combine a convenient offline shopping experience with easily accessible product information online. Marketing should emphasize functionality, comfort, and classic design that aligns with their professional activities, without relying on emotional marketing tactics. Sales interactions should focus on providing an efficient and friendly experience, offering information in a professional manner. For digital communication, it is important to use personalized channels such as WhatsApp or social media DMs to share product updates, exclusive offers, or restock notifications in a polite and unobtrusive way. This approach will help build trust and loyalty by seamlessly integrating offline and online shopping experiences.

3) Cluster Three

The right marketing strategy for this cluster should combine both traditional and digital methods. Since these customers prefer shopping in physical stores but still search for product information online, it's important to optimize both channels. In-store, sales efforts should focus on delivering a convenient and efficient shopping experience. Digitally, detailed product information should be made available through platforms such as Shopee and social media. At the same time, personalized communication via WhatsApp or direct social media messaging can be used to share promotions. Because they are receptive to relevant and unobtrusive messages, personalized email marketing or targeted direct messaging aligned with their interests and needs will be highly effective. These strategies will help build stronger relationships and increase customer loyalty by blending the familiarity of traditional shopping with the convenience of digital engagement.

This study substantiates prior research findings, particularly those of Taufiqurrochman [34], which underscore the critical importance of aligning marketing strategies with the objectives of marketing activities through the appropriate application of the marketing mix. The integration of Principal Coordinates Analysis (PCoA) and K-Means Clustering in this research has proven effective in generating customer segments that are more representative than those derived from conventional demographic segmentation. These results are in line with the assertions of Wang[35], who emphasize that the combination of multivariate analysis and clustering techniques can yield more nuanced and actionable insights into market segmentation.

A distinctive feature of this study lies in its focus on customer segmentation grounded in localized characteristics and actual consumer behavior, particularly with respect to the implementation of the marketing mix. Each cluster identified in the analysis exhibited unique digital behaviors, communication preferences, and purchasing motivations insights that are essential for micro, small, and medium enterprises (MSMEs) seeking to develop marketing strategies that are both targeted and contextually appropriate. For instance, the identification of a specific cluster consisting of female teachers in Banyumas and Pemalang who prefer WhatsApp communication and acquire product information through Shopee and Facebook enables marketers to tailor their outreach strategies accordingly.

Furthermore, the study illustrates the applicability of multidimensional scaling (PCoA) and clustering techniques (K-Means) in advancing MSME business intelligence through customer segmentation. By employing these methodologies on mixed data types, this research contributes to the informatics domain by presenting a replicable framework for translating complex consumer data into strategic insights that are both practical and evidence-based.

Nevertheless, the study has certain limitations that must be acknowledged. The analysis was conducted using a relatively limited sample size and focused exclusively on consumers in the Greater Banyumas region including Pemalang, Purbalingga, and surrounding areas thereby potentially limiting the generalizability of the findings to other geographic or demographic contexts. Additionally, the reliance on self-reported questionnaire data introduces the possibility of response bias, which may influence the accuracy of cluster formation. Future research should consider expanding the dataset, incorporating longitudinal data, and testing the framework across diverse geographic regions to improve

Vol. 6, No. 3, Juni 2025, Page. 1405-1418
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.3.4687

the robustness and external validity of the findings. Acknowledging these limitations is crucial to upholding the scientific integrity of the study and guiding future research efforts

5. CONCLUSIONS

Based on the research results using the K-Means clustering method, three distinct customer groups were identified with a silhouette score of 0.75, indicating a good clustering quality and fairly optimal group separation. Each cluster demonstrates unique characteristics in terms of demographics, consumer behavior, and preferences regarding the marketing mix. Cluster one consists of 68 customers and is categorized as the conventional rational segment individuals who make purchasing decisions based on practical needs, with a focus on price, quality, and functionality. This group, dominated by women aged 26–35, typically negotiates for discounts in the range of 1–5% off the initial price. Cluster two comprises 70 customers and represents the mature rational customer segment, predominantly women aged 36–45, mostly teachers with middle incomes. These consumers tend to make well-planned purchasing decisions, are not influenced by emotional impulses, and generally do not engage in price bargaining—reflecting a more mature consumer behavior. Cluster three includes 41 customers and is categorized as the digital-aware rational male customer segment. It consists of adult males aged 36–45 with stable incomes and purchasing habits based on need. Their pattern of making two purchases suggests emerging loyalty. They prioritize convenience, functionality, and prefer shopping offline, although they actively seek product information online. Within this group, some customers negotiate for minor discounts (1–5%), while others do not. These three clusters highlight the diversity in customer decision-making approaches and provide a solid foundation for developing segmented and more targeted marketing strategies.

This study also contributes to the field of informatics by demonstrating the effective application of Principal Coordinates Analysis (PCoA) and K-Means clustering on mixed-type datasets to support customer segmentation within the context of micro, small, and medium enterprises (MSMEs). This methodological approach offers a replicable model for transforming complex and multidimensional consumer data into clear and actionable strategies—bridging statistical analysis with business intelligence.

However, this research is not without limitations. The sample size is relatively small and geographically concentrated in the regions of Banyumas, Purbalingga, Banjarnegara, and Pemalang, which may limit the generalizability of the findings to broader consumer populations. Furthermore, the data were collected through questionnaires, which may introduce response bias and affect the accuracy of behavioral attributes used in the clustering process. Future studies should address these limitations by expanding the sample to more diverse regions, increasing its size, and incorporating longitudinal or observational data to enhance external validity and ensure broader applicability.

ACKNOWLEDGEMENT

Gratitude is expressed to Allah SWT for His grace and guidance, which enabled the successful completion of this research. The author also extends sincere thanks to UMKM Bintang Sepatu Purwokerto for their valuable contribution as the subject of this research. Appreciation is also given to colleagues who have provided both direct and indirect support throughout the research process. It is hoped that this study will be beneficial, particularly in formulating effective marketing strategies.

REFERENCES

[1] A. Siburian and D. P. Ompusunggu, "Pengaruh Pertumbuhan Usaha Mikro, Kecil dan Menengah Terhadap Pertumbuhan Ekonomi Di Palangka Raya," *JUEB J. Ekon. dan Bisnis*, vol. 2, no. 2, pp. 30–34, 2023, doi: 10.57218/jueb.v2i2.604.

Vol. 6, No. 3, Juni 2025, Page. 1405-1418
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.3.4687

[2] Kominfo, "Transformasi Digital UMKM Jadi Prioritas Penguatan Fondasi Ekonomi," *Kominfo*, 2022.
[3] L. Yulia and W. Setianingsih, "Studi Manajemen Marketing Berbasis Online (Penelitian Pada Umkm Produksi Mebel Di Babakan Muncang Tamansari Kota Tasikmalaya)," *J. Maneksi*, vol. 9, no. 1, pp. 346–354, 2020, doi: 10.31959/jm.v9i1.397.
[4] A. Septiana and Zulkifli, "Jurnal Riset Akuntansi dan Bisnis Jurnal Riset Akuntansi dan Bisnis," *Pengaruh Literasi Keuangan, Sikap Keuangan, dan Perilaku Pengelolaan Keuang. terhadap Kinerja UMKM*, vol. 18, no. 2, pp. 140–152, 2024, doi: 10.17509/jrak.v11i3.52018.
[5] A. Affandi *et al.*, "Optimization of MSMEs empowerment in facing competition in the global market during the COVID-19 pandemic time," *Syst. Rev. Pharm.*, vol. 11, no. 11, pp. 1506–1515, 2020, doi: 10.31838/srp.2020.11.213.
[6] C. Sianturi and J. M. Simanjuntak, "Pengaruh Strategi Pemasaran Terhadap Peningkatan Volume Penjualan Lahan Di Pt. Kawasan Industri Modern (Persero) Medan," *J. Ilmu Sos. Dan Polit.*, vol. 1, no. 2, pp. 205–224, 2021, doi: 10.51622/jispol.v1i2.418.
[7] R. Ananda, "Silhouette Density Canopy K-Means for Mapping the Quality of Education Based on the Results of the 2019 National Exam in Banyumas Regency," *Khazanah Inform. J. Ilmu Komput. dan Inform.*, vol. 5, no. 2, pp. 158–168, 2019, doi: 10.23917/khif.v5i2.8375.
[8] A. Sulistiyawati and E. Supriyanto, "Implementasi Algoritma K-means Clustring dalam Penetuan Siswa Kelas Unggulan," *J. Tekno Kompak*, vol. 15, no. 2, p. 25, 2021, doi: 10.33365/jtk.v15i2.1162.
[9] S. N. Br Sembiring, H. Winata, and S. Kusnasari, "Pengelompokan Prestasi Siswa Menggunakan Algoritma K-Means," *J. Sist. Inf. Triguna Dharma (JURSI TGD)*, vol. 1, no. 1, p. 31, 2022, doi: 10.53513/jursi.v1i1.4784.
[10] S. K. Dirjen, P. Riset, D. Pengembangan, R. Dikti, R. Ananda, and A. Z. Yamani, "Penentuan Centroid Awal K-means pada proses Clustering Data Evaluasi Pengajaran Dosen," *Masa Berlaku Mulai*, vol. 1, no. 3, pp. 544–550, 2017.
[11] C. S. D. B. Sembiring, L. Hanum, and S. P. Tamba, "Penerapan Data Mining Menggunakan Algoritma K-Means Untuk Menentukan Judul Skripsi Dan Jurnal Penelitian (Studi Kasus Ftik Unpri)," *J. Sist. Inf. dan Ilmu Komput. Prima(JUSIKOM PRIMA)*, vol. 5, no. 2, pp. 80–85, 2022, doi: 10.34012/jurnalsisteminformasidanilmukomputer.v5i2.2393.
[12] R. Ananda, A. R. Dewi, and N. Nurlaili, "a Comparison of Clustering By Imputation and Special Clustering Algorithms on the Real Incomplete Data," *J. Ilmu Komput. dan Inf.*, vol. 13, no. 2, pp. 65–75, 2020, doi: 10.21609/jiki.v13i2.818.
[13] R. Ananda and A. Prasetiadi, "Hierarchical and K-means Clustering in the Line Drawing Data Shape Using Procrustes Analysis," *Int. J. Informatics Vis.*, vol. 5, no. 3, pp. 306–312, 2021, doi: 10.30630/JOIV.5.3.532.
[14] K. M. Daud, R. Ananda, S. Zainudin, C. W. Howe, K. Moorthy, and N. I. B. M. Saleh, "Optimizing the Production of Valuable Metabolites using a Hybrid of Constraint-based Model and Machine Learning Algorithms: A Review," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 10, pp. 1091–1105, 2023, doi: 10.14569/IJACSA.2023.01410115.
[15] X. F. Song, Y. Zhang, D. W. Gong, and X. Z. Gao, "A Fast Hybrid Feature Selection Based on Correlation-Guided Clustering and Particle Swarm Optimization for High-Dimensional Data," *IEEE Trans. Cybern.*, vol. 52, no. 9, pp. 9573–9586, 2022, doi: 10.1109/TCYB.2021.3061152.
[16] E. F. L. Awalina and W. I. Rahayu, "Optimalisasi Strategi Pemasaran dengan Segmentasi Pelanggan Menggunakan Penerapan K-Means Clustering pada Transaksi Online Retail," *J. Teknol. dan Inf.*, vol. 13, no. 2, pp. 122–137, 2023, doi: 10.34010/jati.v13i2.10090.
[17] M. F. Nugraha and U. Hayati, "Clustering Data Indonesian Food Delivery Menggunakan Metode K-Means Pada Gofood Product List," vol. 8, no. 3, pp. 3484–3492, 2024, doi: 10.36040/jati.v8i3.9727.
[18] A. Koroluk, E. Paczos-Grzęda, S. Sowa, M. Boczkowska, and J. Toporowska, "Diversity of Polish Oat Cultivars with a Glance at Breeding History and Perspectives," *Agronomy*, vol. 12, no. 10, 2022, doi: 10.3390/agronomy12102423.
[19] D. Hediyati and I. M. Suartana, "Penerapan Principal Component Analysis (PCA) Untuk Reduksi Dimensi Pada Proses Clustering Data Produksi Pertanian Di Kabupaten Bojonegoro,"

J. Inf. Eng. Educ. Technol., vol. 5, no. 2, pp. 49–54, 2021, doi: 10.26740/jieet.v5n2.p49-54.

[20] S. Saniah and Edy Victor Haryanto, "Rancang Bangun Aplikasi Pencarian Lokasi Vaksin Pada Puskesmas Di Kota Medan Menggunakan Metode Euclidean Distance Berbasis Android," *J. Ilm. Sist. Inf. dan Ilmu Komput.*, vol. 3, no. 1, pp. 01–11, 2023, doi: 10.55606/juisik.v3i1.365.

[21] R. Ananda, K. M. Daud, and S. Zainudin, "A review of advances in integrating gene regulatory networks and metabolic networks for designing strain optimization," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 36, no. 6, p. 102120, 2024, doi: 10.1016/j.jksuci.2024.102120.

[22] S. Dhau, R. Ananda, I. Nurisusilawati, N. Afrina, H. Zulkarnain, and J. Tengah, "Optimasi Rute Distribusi Frozen Food Menggunakan Metode Principal Coordinate Analysis Dan Nearest Neighbor," vol. 1, pp. 162–170, 2024, doi: 10.61124/sinta.v1i4.34.

[23] S. Sucipto, D. Dwi Prasetya, and T. Widiyaningtyas, "Educational Data Mining: Multiple Choice Question Classification in Vocational School," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 23, no. 2, pp. 379–388, 2024, doi: 10.30812/matrik.v23i2.3499.

[24] C. Gower, J, "A general coefficient of similarity and some of its properties.," *Biometrics*, vol. 27, no. 4, pp. 857–871, 1971, [Online]. Available: https://www.jstor.org/stable/2528823

[25] I. Fadil, R. Goejantoro, and S. Prangga, "Aplikasi K-Nearest Neighbor Dengan Fungsi Jarak Gower Dalam Klasifikasi Kelulusan Mahasiswa (Studi Kasus : Mahasiswa Program Studi Statistika, Jurusan Matematika, Fakultas Matematika Dan Ilmu Pengetahuan Alam, Universitas Mulawarman)," *J. EKSPONENSIAL*, vol. 13, no. 1, pp. 57–62, 2022, doi: 10.30872/eksponensial.v13i1.881.

[26] I. Izzati, D. Yusuf, B. Susetyo, and L. O. Abdul, "Perbandingan Metode Hot-deck , Regression dan K-Nearest," vol. 12, no. 1, pp. 22–35, 2023, doi: 10.29244/xplore.v12i1.1056.

[27] A. Winarta and W. J. Kurniawan, "Optimasi Cluster K-Means Menggunakan Metode Elbow Pada Data Pengguna Narkoba Dengan Pemrograman Python," *JTIK (Jurnal Tek. Inform. Kaputama)*, vol. 5, no. 1, pp. 113–119, 2021, doi: 10.59697/jtik.v5i1.593.

[28] A. Yudhistira and R. Andika, "Pengelompokan Data Nilai Siswa Menggunakan Metode K-Means Clustering," *J. Artif. Intell. Technol. Inf.*, vol. 1, no. 1, pp. 20–28, 2023, doi: 10.58602/jaiti.v1i1.22.

[29] L. G. Rady Putra and A. Anggrawan, "Pengelompokan Penerima Bantuan Sosial Masyarakat dengan Metode K-Means," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 1, pp. 205–214, 2021, doi: 10.30812/matrik.v21i1.1554.

[30] A. P. Riani, A. Voutama, and T. Ridwan, "Penerapan K-Means Clustering Dalam Pengelompokan Hasil Belajar Peserta Didik Dengan Metode Elbow," *J-SISKO TECH (Jurnal Teknol. Sist. Inf. dan Sist. Komput. TGD)*, vol. 6, no. 1, p. 164, 2023, doi: 10.53513/jsk.v6i1.7351.

[31] R. Bayu Prasetyo, Y. Agus Pranoto, and R. Primaswara Prasetya, "Implementasi Data Mining Menggunakan Algoritma K-Means Clustering Penyakit Pasien Rawat Jalan Pada Klinik Dr. Atirah Desa Sioyong, Sulteng," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 4, pp. 2144–2151, 2023, doi: 10.36040/jati.v7i4.7419.

[32] M. S. Pangestu and M. A. Fitriani, "Perbandingan Perhitungan Jarak Euclidean Distance, Manhattan Distance, dan Cosine Similarity dalam Pengelompokan Data Bibit Padi Menggunakan Algoritma K-Means," *Sainteks*, vol. 19, no. 2, p. 141, 2022, doi: 10.30595/sainteks.v19i2.14495.

[33] M. S. Abidin *et al.*, "Application of the DBSCAN Algorithm in MSME Clustering using the Silhouette Coefficient Method," vol. 07, no. 03, pp. 260–270, 2024, doi: https://doi.org/10.37396/jsc.v7i3.472.

[34] M. Taufiqurrochman, W. Khristianto, and E. Wahyudi, "Implementasi Bauran Pemasaran Pada Kopixel Jember," *J. Ilm. Manaj. dan Bisnis*, vol. 3, no. 1, pp. 1–12, 2024, doi: 10.24034/jimbis.v3i1.6075.

[35] S. Wang, L. Sun, and Y. Yu, "A dynamic customer segmentation approach by combining LRFMS and multivariate time series clustering," *Sci. Rep.*, vol. 14, no. 1, pp. 1–18, 2024, doi: 10.1038/s41598-024-68621-2.