BAREKENG: Journal of Mathematics and Its Applications
June 2025 Volume 19 Issue 2 Page 0961–0972
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol19iss2pp0961-0972

# APPLICATION OF DBSCAN FOR CLUSTERING SOCIETY BASED ON WASTE MANAGEMENT BEHAVIOR

**Hafizh Syihabuddin Al Jauhar [1*], Solimun[2], Rahma Fitriani[3]**

[1,2,3]*Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Brawijaya*
*Jln. Veteran No.10-11, Ketawanggede, Kecamatan Lowokwaru, Malang, Jawa Timur, 65145, Indonesia*

*Corresponding author's e-mail: * **aljauhar.hafizh@student.ub.ic.id***

*ABSTRACT*

***Article History:***

*Received: 17th August 2024*
*Revised: 20th January 2025*
*Accepted: 28th February 2025*
*Published: 1st April 2025*

***Keywords:***

*DBSCAN Clustering;*
*Environmental Quality;*
*Noise Detection;*
*Waste Management Behavior;*
*3R-Based Waste Management.*

*This research aims to answer the challenge of identifying the characteristics of the Batu City community in waste management, where traditional clustering techniques are often suboptimal due to the presence of noise or objects that do not fit the general pattern. As a solution, the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm is applied, which allows the clustering of objects based on local density and detects the presence of noise or outliers in the data. DBSCAN is considered more flexible than other clustering methods, especially in clustering data that is not linear or has a non-uniform distribution. This study successfully identified three clusters of waste management behavior with a silhouette index of 0.875, indicating good cluster quality. The first cluster consists of communities with good environmental quality, active participation in the use of waste banks, and a deep understanding of 3R-based waste management. The second cluster has adequate infrastructure quality and high awareness of the potential economic benefits of waste, while the third cluster displays a pretty good level of understanding of the 3Rs and relatively good environmental quality. The results of this study provide important insights into the differences in waste management characteristics between clusters, with environmental quality proving to be a significant factor in cluster formation.*



***How to cite this article:***

H. S. A. Jauhar, Solimun and R. Fitriani., "APPLICATION OF DBSCAN FOR CLUSTERING SOCIETY BASED ON WASTE MANAGEMENT BEHAVIOR," *BAREKENG: J. Math. & App.*, vol. 19, iss. 2, pp. 0961-0972, June, 2025.


Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

**Research Article · Open Access**

# 1. INTRODUCTION

Clustering technique is the process of grouping a set of data objects into several groups or clusters so that objects in a cluster have high similarity but are very different from objects in other clusters [1]. Not all objects in the study have characters that can be classified in a particular cluster. Some objects with unique characters are much different from other objects. These objects are noise that, if included in a cluster, can affect character identification and reduce the accuracy of the cluster results formed [2]. Objects identified as noise can provide information that cannot be provided by other objects in the data, for example, because noise arises from a combination of unusual circumstances that may be very important and need to be investigated [3].

Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm is able to detect the presence of noise in data that is effective for clustering data based on data density [2]. This algorithm system groups data by identifying points that are not included in the cluster as noise and grouping points that have high similarity.

DBSCAN was introduced by Ester, this algorithm is a non-parametric clustering technique [4]. This algorithm was awarded the test of the time award in 2014 by ACM (Association for Computing Machinery) during the data mining conference. DBSCAN is a type of partition-based cluster where denser regions are considered as clusters and low-density regions are called noise [5]. In DBSCAN, there is no need to provide an input value for the number of clusters. Therefore, clusters are formed based on density, not from any form of assumption [6]. The main advantage of this algorithm is that it will not include data that is considered noise in any cluster. Of course, this is an advantage in itself considering that noise data may have characteristics that are much different from the general data set.

Previous research on DBSCAN cluster analysis has been conducted by [7] entitled "Adaptive marine traffic behavior pattern recognition based on multidimensional dynamic time warping and DBSCAN algorithm". This research identifies marine traffic behavior patterns using the DBSCAN algorithm. The results show that the DBSCAN approach is more accurate and adaptive in terms of recognizing hidden traffic behavior patterns than other methods.

Another study using DBSCAN cluster analysis was conducted by [8] in a study entitled "Cluster Mapping of Waste Exposure Using DBSCAN Approach: Study of Spatial Patterns and Potential Distribution in Bantul Regency". This study aims to map locations that have the potential to experience waste accumulation in order to reduce the negative impact of microplastic pollution. The results showed that the optimal value was obtained with MinPts of 7 and Eps ($\varepsilon$) of 1500, resulting in a Silhouette Coefficient (SC) value of 0.78, which is included in the strong category.

Cluster analysis can be applied in various fields, one of which is the field of community grouping based on waste management behavior patterns. Waste management is one of the biggest challenges faced by the Indonesian Government, especially Batu City, with waste production reaching more than 65 million tons per year, according to data from the Ministry of Environment and Forestry (MoEF) [9]. Increasing population and rapid urbanization exacerbate the situation, while public awareness of the importance of good waste management is still low, and waste treatment facilities are limited. This situation underscores the urgency of this study, given the negative impacts resulting from ineffective waste management. Poorly managed waste can cause environmental pollution and health problems and harm the local economy. In addition, with the increasing volume of waste, Batu City faces challenges in providing adequate waste management infrastructure. Therefore, this study aims to obtain the characteristics of the Batu City community based on waste management behavior patterns and provide a comprehensive understanding of the dynamics of waste management behavior in Batu City. The results of this study are not only relevant for the development of more effective waste management strategies but also make a significant contribution to the scientific literature in the field of waste management and statistical analysis.

## 2. RESEARCH METHODS

### 2.1 Hopkins Statistical Test

According to [10], clustering imposes a classification on an evenly distributed random data set even if there are no meaningful groups in it. Therefore, a clustering propensity assessment method should be used to evaluate the validity of the clustering analysis. That is, whether a given data set contains meaningful groups. The Hopkins statistic is used to assess the clustering tendency of a set of objects by measuring the probability that a given set of objects is generated by a uniform distribution of objects. In other words, this method is used to test the spatial randomness of the data. Hopkins statistics can be obtained using Equation (1).

$$\boldsymbol{H} = \frac{\sum_{i=1}^{n} \boldsymbol{W_i}}{\sum_{i=1}^{n} \boldsymbol{U_i} + \sum_{i=1}^{n} \boldsymbol{W_i}} \tag{1}$$

Description:

$H$ : Hopkins Statistical Test, where $0 \leq H \leq 1$

$U_i$ : The nearest neighbor distance between objects $X_i$ and $X_i'$ in the original data

$W_i$ : Nearest neighbor distance between objects $X_i$ and $X_i'$ sampled from probability sampling of the original data.

Hopkins statistics are not deterministic, so the boundary points that can be considered part of a cluster can vary depending on the order in which the data are processed [11]. The interpretation used until now still follows the guidelines of [12], which states that if the Hopkins statistical value is close to 0, it indicates that the data spreads in the form of a grid, while if the Hopkins statistic is equal to 0.5, it indicates that the data spreads randomly so it is less suitable for clustering, on the other hand, if the Hopkins statistical value is close to 1, it indicates that the data tends to form a grouped pattern so that it is feasible to do the clustering process. One way to overcome the problem of no tendency for groups to form is by increasing the data dimension.

### 2.2 Principal Component Analysis (PCA)

According to [13], principal component analysis is a statistical technique that aims to transform most of the original variables used and correlated with each other into a new set of smaller, uncorrelated variables.

Principal component analysis has two types of input matrices, namely, the variance matrix and the correlation matrix. The variance matrix is used if the units in the variables are the same, while the correlation matrix is used if the units in the variables are different [13]. The matrix is used to construct the characteristic root and characteristic vector. The variance matrix can be seen in the following matrix.

$$\boldsymbol{S_{p\times p}} = \begin{bmatrix} \boldsymbol{S_1^2} & \boldsymbol{S_{12}} & \cdots & \boldsymbol{S_{1p}} \\ \boldsymbol{S_{21}} & \boldsymbol{S_2^2} & \cdots & \boldsymbol{S_{2p}} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{S_{p1}} & \boldsymbol{S_{p2}} & \cdots & \boldsymbol{S_p^2} \end{bmatrix}$$

Elements in the variance matrix are obtained using the following formula

$$\boldsymbol{S_j^2} = \frac{\sum_{i=1}^{n}(\boldsymbol{x_{ji}} - \boldsymbol{x_{j.}})^2}{\boldsymbol{n-1}} \tag{2}$$

$$\boldsymbol{S_{j,j'}} = \frac{\sum_{i=1}^{n}(\boldsymbol{x_{ji}} - \boldsymbol{x_{j.}})(\boldsymbol{x_{j'i}} - \boldsymbol{x_{j'.}})}{\boldsymbol{n-1}} \tag{3}$$

After forming the variance matrix, we will then look for the value of the characteristic root $(\boldsymbol{\lambda_1} \geq \boldsymbol{\lambda_2} \geq \cdots \geq \boldsymbol{\lambda_m} \geq \boldsymbol{0})$. The characteristic root is obtained by solving Equation (4).

$$|\boldsymbol{S} - \boldsymbol{\lambda I}| = \boldsymbol{0} \tag{4}$$

The following is the formula for calculating component weights in Equation (5).

$$(\boldsymbol{S} - \boldsymbol{\lambda I})\boldsymbol{a_j} = \boldsymbol{0} \tag{5}$$

After obtaining the component weights, a linear principal component equation can then be formed. Suppose there are $\boldsymbol{p}$ variables, namely $\boldsymbol{X_1, X_2, \dots, X_p}$, transformed into a new variable vector $\boldsymbol{KU' = (KU_1, KU_2, \dots, KU_p)}$. The linear combination of the main components formed is presented in **Equation (6)**.

$$\begin{aligned} KU_1 &= a'_1X = a_{11}X_1 + a_{12}X_2 + \dots + a_{1p}X_p \\ KU_2 &= a'_2X = a_{21}X_1 + a_{22}X_2 + \dots + a_{2p}X_p \\ &\vdots \\ KU_p &= a'_pX = a_{p1}X_1 + a_{p2}X_2 + \dots + a_{pp}X_p \end{aligned} \tag{6}$$

Description:

| | | |
|---|---|---|
| $\boldsymbol{KU_p}$ | : | The number of new variables formed as many as $\boldsymbol{p}$ |
| $\boldsymbol{a_{pp}}$ | : | Component weight |
| $\boldsymbol{X_p}$ | : | The number of initial variables $\boldsymbol{p, j = 1, 2, \dots, p}$ |

If the main component has been obtained, then proceed to calculate the component score of each individual, which will be used for further analysis. The principal component score of the $\boldsymbol{i}$-th individual can be found using **Equation (7)**.

$$SK_{ij} = z_i a_j \tag{7}$$

Description:

| | | |
|---|---|---|
| $\boldsymbol{SK_{ij}}$ | : | The $\boldsymbol{j}$-th component score of the $\boldsymbol{i}$-th individual |
| $\boldsymbol{z_i}$ | : | Data vector of the $\boldsymbol{i}$-th individual |
| $\boldsymbol{a_j}$ | : | Component weight vector |

### 2.3 Density-Based Spatial Clustering of Applications with Noise

DBSCAN is a clustering algorithm based on data density. The concept of density in DBSCAN is the amount of data within the Epsilon or Eps radius ($\boldsymbol{\varepsilon}$) of each data. If the amount of data in the radius $\boldsymbol{\varepsilon}$ is more than or equal to MinPts (minimum amount of data in radius $\boldsymbol{\varepsilon}$), the data falls into the desired density category, and the amount of data in the radius includes the data itself **[14]**. In this algorithm, several terms are known as follows **[7]**:

1. Core is the center point in the cluster based on the density where there are a number of points that must be within the user-defined Eps (radius or threshold value) and MinPts (minimum points in the cluster).
2. Border is a point that becomes a boundary within the center point area (core).
3. Noise is a point that cannot be reached by the core and is not a border.
4. Direct reachable density is a point said to be a direct reachable point if the point is directly connected to the centre point (core).
5. Reachable density is a point said to be a reachable point if the point is indirectly connected to the center point (core).
6. Connected density is a point said to be connected to each other by another point.

The DBSCAN algorithm requires two important parameters, namely Eps ($\boldsymbol{\varepsilon}$) and the minimum number of neighboring data, to form a group (MinPts). The algorithm of DBSCAN is as follows **[15]**:

1. Select the starting point p randomly.
2. Initialize the input parameters MinPts and Eps.
3. Calculate Eps or all affordable density distances to $\boldsymbol{p}$ using Euclidean distance.

$$d_{ij} = \sqrt{\sum_a^p (x_{ia} - x_{ja})^2} \tag{8}$$

Where $\boldsymbol{x_{ia}}$ is the $\boldsymbol{a}$-th variable of the object $\boldsymbol{i}$ ($\boldsymbol{i = 1, \dots, n}$; $\boldsymbol{a = 1, \dots, p}$) and $\boldsymbol{d_{ij}}$ is the Euclidean Distance value.

4. If the number of points that satisfy Eps is more than MinPts, then point $\boldsymbol{p}$ is the center point (core), and the cluster is formed.

5. Repeat steps 3 - 4 until all points are processed. If $\boldsymbol{p}$ is a border point and there is no point whose density is affordable to $\boldsymbol{p}$, then the process continues to another point.

There are several advantages possessed by the DBSCAN algorithm, including [16].

1. Recognition of non-convex clusters.
2. Automatic determination of the number of clusters.
3. There is no need to use an index to determine the appropriate number of clusters in a partition.

### 2.4 K-Nearest Neighbour Algorithm

Clustering data using the DBSCAN algorithm requires two main parameters: epsilon and MinPts [17]. Determination of epsilon and MinPts can be done with the help of the K-Nearest Neighbour (KNN) algorithm. The KNN algorithm is an algorithm for classifying objects against learning data that is closest to the object [18]. To determine epsilon and MinPts, we can use a $k$-dist graph.

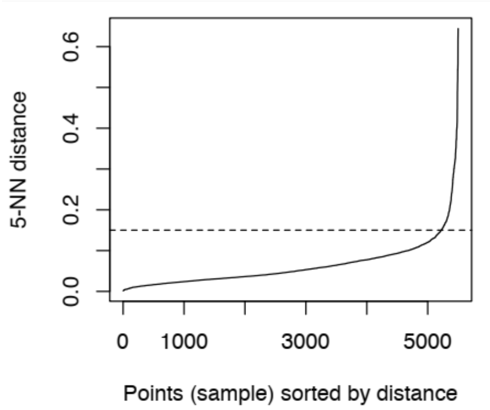


**Figure 1. Example of $k$-dist Graph**
(Source: [10])

The illustration above is a $k$-dist graph with a $k$ value of 5. Sharp changes (elbow-shaped) that occur on the graph will be considered as epsilon and the $k$ value is used as MinPts. The vertical line on the graph is a threshold that is determined based on a significant increase in the overall $k$-dist graph. The threshold line aims to prevent errors in determining the epsilon value where if the eps value is too large, it results in outliers joining the group [19]. The steps in determining epsilon and MinPts are as follows.

1. Calculate the $k$-dist value: For each point, compute the average Euclidean distance to its $k$-nearest neighbors to obtain the $k$-dist value.
2. Sort the $k$-dist values in increasing order from least to greatest average.
3. Observe the point with the sharpest change in each $k$-dist graph and mark it as the threshold.
4. The change in the $k$-dist value or the point forming the elbow will be used as the epsilon value and the $k$ value as the corresponding MinPts.
5. If the value of $k$ is too large, the group formed is smaller, and many points will be identified as noise.

### 2.5 Silhouette Index

Silhouette Index (SI) is a cluster validation method that can be applied to measure the validity of data in a single cluster (one cluster out of all clusters) [20]. In calculating the SI value of the $\boldsymbol{i}$-th data, two coefficients are required, namely $\boldsymbol{a}$ and $\boldsymbol{b}$. The coefficient $\boldsymbol{a}$ is used to measure the extent to which the $\boldsymbol{i}$-th data is not similar to the cluster followed. Meanwhile, the $\boldsymbol{b}$ value aims to evaluate the extent to which the data is incompatible with other clusters. The higher the $\boldsymbol{b}$ value, the more incompatible the $\boldsymbol{i}$-th data is to other clusters. The formulas representing the coefficients $\boldsymbol{a}$ and $\boldsymbol{b}$ can be found in Equation (9) and Equation (10) as explained by [20].

$$a_i^j = \frac{1}{m_j - 1} \sum_{\substack{r=1 \\ r\neq 1}}^{m_j} d\left(x_i^j, x_r^j\right), \qquad i = 1, 2, \dots, m_j \tag{9}$$

$$b_i^j = min\left\{\frac{1}{m_n}\sum_{\substack{r=1\\r\neq 1}}^{m_n} d\left(x_i^j, x_r^n\right)\right\}, i = 1, 2, \dots, m_n \quad (10)$$

where:

| | | |
|---|---|---|
| $d(x_i^j, x_r^j)$ | : | Distance of $i$-th data with $r$-th data in one cluster $j$ |
| $m_j$ | : | Number of data in the $j$-th cluster |

The SI value has a range of $[-1, 1]$ which means:

1) When the SI value is close to 1, it indicates that the data is more appropriate to be in the cluster.
2) If the SI value is negative ($a_i > b_i$), it indicates that the data does not fit in the cluster but is closer to another cluster.
3) If the SI value = 0 or close to 0, it indicates that the data lies on the boundary between two adjacent clusters.

The SI value can be found in several conditions, namely, the SI value of the $i$-th data, the cluster SI value, and the Global SI value. The three SI values are used to assess the proximity of data or clusters to the number 1. The formula for the SI value of the $i$-th data is given in **Equation (11)**.

$$SI_i^j = \frac{b_i^j - a_i^j}{\max\{a_i^j, b_i^j\}} \quad (11)$$

Cluster SI is obtained by calculating the average SI value of all data in the cluster and by using the formula in **Equation (12)**.

$$SI_j = \frac{1}{m_j}\sum_{i=1}^{m_j} SI_i^j \quad (12)$$

Global SI is obtained by calculating the average SI value of all clusters; thus, the formula in **Equation (13)** is obtained.

$$SI = \frac{1}{k}\sum_{j=1}^{k} SI_j \quad (13)$$

Where k is all the clusters.

### 2.6 Data Source

Research data in the form of primary data obtained from questionnaires. Respondents of this research are the people of Batu City. The unit of analysis and sample unit in this study is the individual. The population in this study was all people in Batu City. The sample of this study was the community in Batu City, with a total of 100 respondents referring to the theory put forward by **[21]**, which states that models using $\leq 7$ variables require a sample of at least 100. The sampling method used is the purposive sampling method. The research was conducted from July to August 2024. In summary, the data structure of this study can be seen in **Table 1**.

**Table 1. Data Structure**

| respondents | Variable | | | | |
|---|---|---|---|---|---|
| | Environmental Quality ($X_{1i}$) | Quality of Facilities and Infrastructure ($X_{2i}$) | Use of Waste Bank ($X_{3i}$) | Understanding of 3R-Based Waste Management ($X_{4i}$) | Economic Benefits of Waste ($X_{5i}$) |
| 1 | $x_{11}$ | $x_{21}$ | $x_{31}$ | $x_{41}$ | $x_{51}$ |
| 2 | $x_{12}$ | $x_{22}$ | $x_{32}$ | $x_{42}$ | $x_{52}$ |
| 3 | $x_{13}$ | $x_{23}$ | $x_{33}$ | $x_{43}$ | $x_{53}$ |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 100 | $x_{1100}$ | $x_{2100}$ | $x_{3100}$ | $x_{4100}$ | $x_{5100}$ |

### 2.7 Research Steps

The following are the steps of the research conducted:

1. Principal Component Analysis (PCA): Reduce variables to form component scores, retaining principal components (PCs) with a cumulative proportion of variance exceeding 80%.
2. View the tendency of the object set with Hopkins' statistics.
3. Calculating the Euclidean distance with **Equation (8)**.
4. Determining epsilon and MinPts values.
5. Randomly determine the starting point 'P' for cluster analysis with the DBSCAN algorithm.
6. Perform point checking in the form of core points, border points, and noise by comparing epsilon values and Euclidean distances between objects.
7. Repeat checking all points.
8. Group results with the DBSCAN algorithm are formed.
9. Perform group validation to find the best cluster result using the silhouette index according to **Equations (11)** until **Equation (13)**.
10. Comparison of the goodness of the two methods.
11. Interpretation of character clustering results.

## 3. RESULTS AND DISCUSSION

### 3.1 Principal Component Analysis Result

The data used in this study has the same units of each variable, so it can form a covariance matrix that will be used in determining eigenvalues and eigenvectors. The covariance variance matrix is obtained as follows.

$$\boldsymbol{S}_X = \begin{bmatrix} 0.238 & 0.134 & 0.141 & 0.148 \\ 0.134 & 0.158 & 0.102 & 0.115 \\ 0.141 & 0.102 & 0.244 & 0.117 \\ 0.148 & 0.115 & 0.117 & 0.208 \end{bmatrix}$$

The covariance matrix can be used to obtain eigenvalues and eigenvectors. The eigenvalue can be used to determine the number of PCs formed, while the normalized eigenvector can be the PC coefficient and the determinant of a variable to be included in the PC formed. The determination of the number of PCs can be seen from the magnitude of the diversity of each PC. In this study, two PCs were taken that had a total proportion of diversity of more than 80%, which was used to reduce correlated variables and extract values into simpler components.

**Table 2. PC Proportion of Variance**

| PC | Proportion of Variance | Cumulative Proportion of Variance |
|---|---|---|
| 1 | 0.703 | 0.703 |
| 2 | 0.130 | 0.833 |
| 3 | 0.093 | 0.926 |
| 4 | 0.074 | 1.000 |

Based on **Table 2**, it can be seen that PC1 and PC2 have a total variance proportion of more than 80%, which makes PC1 and PC2 representative of information from the four variables used. This is in accordance with the purpose of PCA, which is to reduce variables and overcome the correlation between variables. **Table 3** presents the PC coefficients and PCA results of the variables in the data used in this study.

**Table 3. Component Score Coefficient of PCA Result**

| No | PC1 | PC2 | PC3 | PC4 |
|---|---|---|---|---|
| 1 | 0.660 | 1.128 | -0.623 | 0.968 |
| 2 | 0.326 | 0.019 | 1.914 | -0.062 |

| No | PC1 | PC2 | PC3 | PC4 |
|---|---|---|---|---|
| 3 | -0.209 | 0.609 | -0.627 | 0.326 |
| 4 | -1.100 | 0.014 | -0.229 | 0.574 |
| 5 | -0.181 | 0.523 | 0.158 | 0.357 |
| 6 | -0.208 | 0.291 | -0.175 | -0.430 |
| 7 | -0.158 | -0.272 | -0.802 | -0.895 |
| 8 | -0.078 | 0.700 | 0.768 | -0.136 |
| 9 | -0.213 | -2.342 | -1.813 | -1.581 |
| 10 | 0.003 | -0.079 | 1.014 | -0.637 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 99 | -0.979 | 1.488 | 0.143 | 0.718 |
| 100 | 0.978 | -0.815 | -0.505 | 0.138 |

### 3.2 Optimal Epsilon and MinPts Values

Optimal epsilon and MinPts values are obtained from the identification of temporary epsilon and MinPts value options. The best epsilon and MinPts values are determined in the elbow formation results and from the sharpest changes in the $k$-dist graph. The following is a visualization of the $k$-dist graph with $k = 3$, $k = 5$, and $k = 7$.

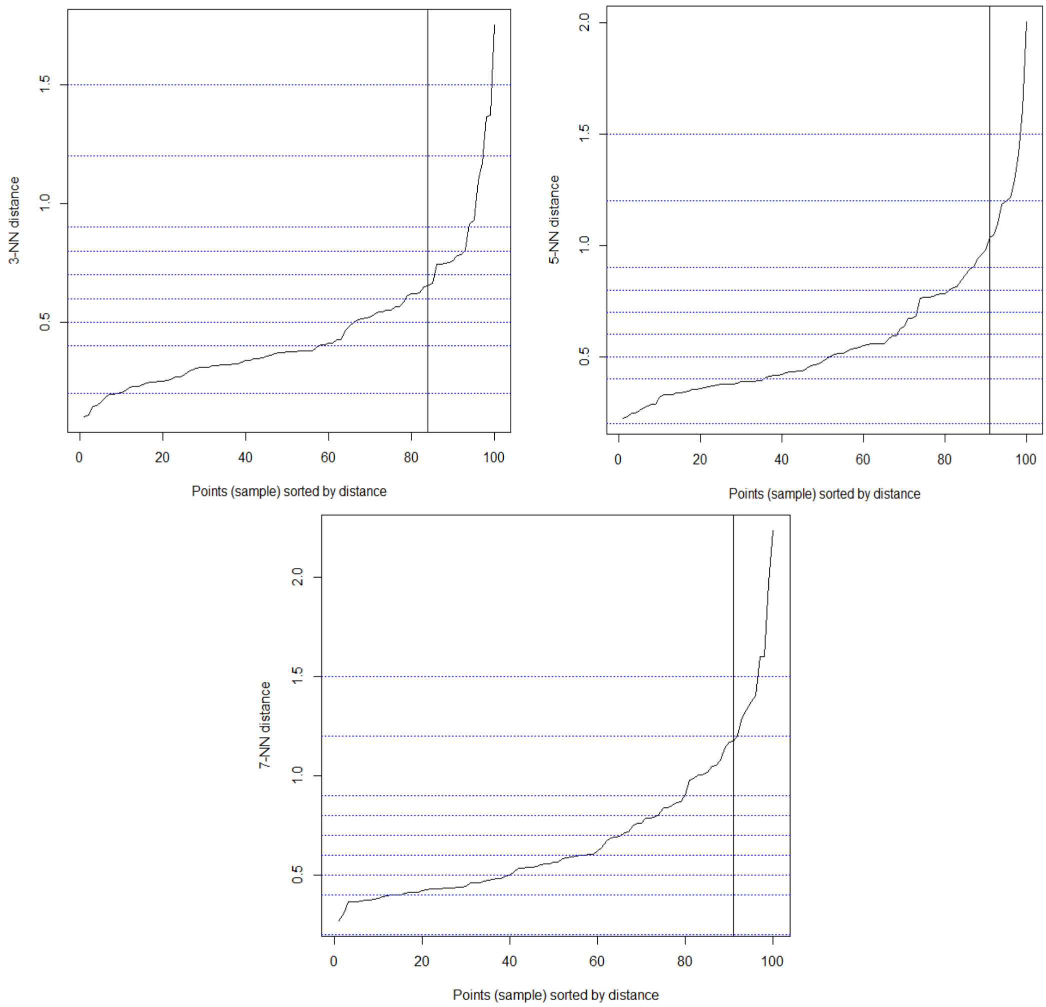


**Figure 2. Visualization of $k$-dist Graph**

From the $k$-dist graphs combined in **Figure 2**, with $k$ values of 3, 5, and 7, respectively, it can be seen that the threshold points are at the 85th, 91st, and 90th points on the X-axis. The observed transient epsilon values for $k = 3$ are 0.2, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.2, and 1.5; for $k = 5$ are 0. 1, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.2, and 1.5; and for $k = 7$ are 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.2, and 1.5. These epsilon values are indicated by the dashed blue line on the graph. The determination of epsilon and MinPts values based on the elbow points on the $k$-dist graph has the disadvantage of being subjective. Therefore, the best epsilon and MinPts values are identified by evaluating the silhouette values of the clusters formed. The clustering results using the DBSCAN method with the best combination of MinPts and epsilon are presented in **Table 4**.

**Table 4. Comparison of Best Parameter Values on DBSCAN**

| MinPts | Epsilon | Cluster Result | Number of points | Silhouette Index |
|---|---|---|---|---|
| 2 | 0.68 | Noise | 5 | 0.768 |
| | | 1 | 87 | |
| | | 2 | 6 | |
| | | 3 | 2 | |
| 3 | 0.50 | Noise | 15 | 0.875 |
| | | 1 | 74 | |
| | | 2 | 4 | |
| | | 3 | 7 | |
| 5 | 0.68 | Noise | 9 | 0.678 |
| | | 1 | 80 | |
| | | 2 | 5 | |
| | | 3 | 6 | |
| 7 | 0.68 | Noise | 56 | 0.277 |
| | | 1 | 22 | |
| | | 2 | 12 | |
| | | 3 | 10 | |

The reason for using the pair of minimum points and epsilon values in **Table 4**. Because only these parameters meet the requirements for group formation with DBSCAN, the epsilon value to be used is 0.875, and the minimum points value is 3. Based on the highest silhouette index value, it can be concluded that the DBSCAN method is more optimal if using MinPts 3 and epsilon 0.5.

### 3.3 DBSCAN Cluster Results

The clustering results use the DBSCAN method, which uses two main parameters, namely epsilon and MinPts. In this study, the optimal values for these parameters were determined to be 0.5 for epsilon and 3 for MinPts. The clustering results consist of three groups, comprising 74 communities, 4 communities, and 7 communities, respectively, with 15 communities identified as noise. The clustering achieved a Silhouette Index exceeding 0.5, indicating acceptable cluster quality. The following is a visualization of the cluster results using DBSCAN.

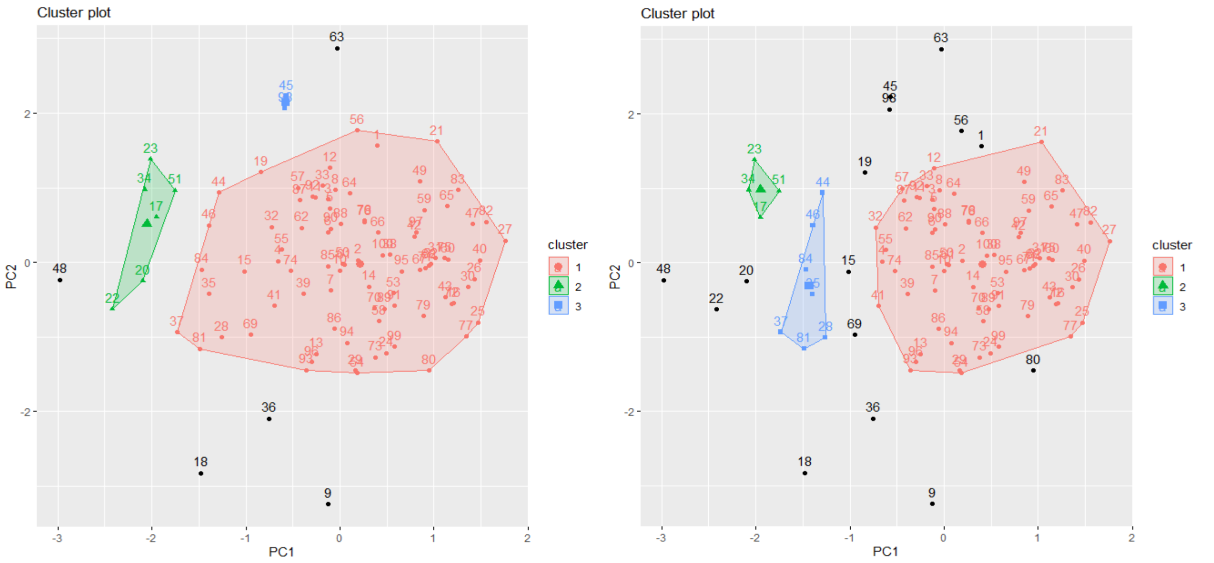


**Figure 3. Visualization of Cluster Results with DBSCAN**

The image on the left side is a visualization with epsilon 0.68 and MinPts as much as 2. While the image on the right is a visualization with epsilon 0.50 and MinPts as much as 3. It can be seen that the distribution of noise makes it easier to identify its characteristics through the nearest cluster using the image on the left. However, the optimal cluster chosen is when MinPts is 3 with epsilon 0.50 because it considers a much higher cluster validation value.

### 3.4 Interpretation of Cluster Results

The silhouette index value in the DBSCAN method is 0.875. this indicates that the DBSCAN method is appropriate for interpreting the cluster results in this case. The three groups have their own characteristics. the characteristics of each are as follows:

1. Cluster 1

Cluster 1 has the highest average value compared to the other two clusters, so it can be considered a high cluster. The Batu City community in this cluster has characteristics such as good environmental quality, good use of waste banks, and a good understanding of 3R-based waste management (Reduce, Reuse, Recycle). This shows that community awareness and participation in waste management in cluster 1 is relatively high.

However, there are some aspects that need further attention. Despite the good environmental quality and the use of waste banks, respondents in cluster 1 still have low-quality facilities and infrastructure. Inadequate facilities can be an obstacle to implementing more effective and efficient waste management practices.

In addition, the level of economic usefulness of waste is also low in cluster 1, which means that the economic potential of waste management has not been maximized. This factor could be caused by the lack of access to markets or the lack of knowledge and skills to process waste into products of economic value. Therefore, there needs to be efforts to improve the quality of facilities and infrastructure, as well as economic empowerment programs that focus on waste utilization. This intervention is expected to improve community welfare while supporting more sustainable waste management efforts.

2. Cluster 2

Cluster 2 has an average value between Cluster 1 and Cluster 3, so it can be considered a medium cluster. Respondents included in this cluster have characteristics such as good quality of facilities and infrastructure and a good understanding of the economic benefits of waste. This indicates that there is better economic potential in waste utilization in cluster 2, which can have a positive impact on the welfare of the community.

However, there are some critical aspects that need to be considered. Communities in cluster 2 have poor environmental quality. This may be due to suboptimal waste management or the presence of other environmental factors that affect quality of life, such as air or water pollution.

In addition, the use of waste banks in cluster 2 is low, indicating a lack of community participation in organized waste management programs. This could indicate that despite an understanding of the economic benefits of waste, participation in community programs such as waste banks is lacking. This low participation could limit the overall effectiveness of waste management.

The understanding of 3Rs-based waste management (Reduce, Reuse, Recycle) is also poor in cluster 2. This suggests that although respondents are aware of the economic benefits of waste, awareness and practice of reducing, reusing, and recycling waste is still low. The absence of ongoing education programs could be one of the reasons for this low understanding.

To improve this condition, there need to be interventions that focus on education and awareness raising regarding 3R-based waste management as well as increased participation in the waste bank program. Training programs and environmental awareness campaigns can help improve environmental quality and increase more sustainable waste management practices in cluster 2. In addition, increased collaboration between the government and the community can also accelerate these positive changes.

3. Cluster 3

Cluster 3 has a lower average value than Cluster 1 and Cluster 2, so it can be considered a low cluster. Respondents in this cluster have characteristics such as an understanding of 3R-based waste management (Reduce, Reuse, Recycle) and fairly good environmental quality. However, there are some important aspects that need attention in this cluster.

Although the understanding of waste management and environmental quality is quite good, respondents in Cluster 3 have a low level of quality facilities and infrastructure. Inadequate facilities can hinder the implementation of effective waste management. In addition, the use of waste banks in this cluster is minimal, indicating low community participation in the waste bank program.

The level of economic usefulness of waste is also low in Cluster 3, which means that the economic potential of waste management has not been maximized. This could be due to the lack of access to markets or low awareness of the economic value of waste. Therefore, there needs to be more focused interventions on improving the quality of facilities and infrastructure, as well as economic empowerment from waste management to optimize the benefits that can be obtained from waste management efforts that are already quite good in terms of understanding and environmental quality.

In addition to the three clusters formed, there were 15 objects classified as noise. Further analysis was conducted to identify the characteristics of these objects based on their proximity to the average (centroid) of each cluster. The closeness was measured using the Euclidean distance between each noise point and the centroid of the nearest cluster. Objects that have a closer distance to the centroid of a cluster indicate the presence of certain patterns or trends similar to that cluster, even if they do not meet the inclusion criteria, such as local density or the minimum number of neighbors (minPts) required in the DBSCAN algorithm. The identification of noise point characteristics with the nearest cluster is presented in the following **Table 5.**

**Table 5. Identification of Noise Points with the Closest Cluster**

| Noise | Characteristics |
|---|---|
| **1, 15, 19, 36, 56, 69, and 80** | They have almost the same characteristics as Cluster 1, such as good environmental quality, good use of waste banks, and a good understanding of 3R-based waste management (Reduce, Reuse, Recycle), but their scores deviate slightly or are not fully consistent with the main clustering pattern of Cluster 1. |
| **20** | They have almost the same characteristics as Cluster 2, such as good quality facilities and infrastructure and a good understanding of the economic benefits of waste, but their scores deviate slightly or are not fully consistent with the main clustering pattern of Cluster 2. |

## 4. CONCLUSIONS

The conclusions obtained based on the results of the analysis are as follows.

1. The DBSCAN method using Euclidean distance can be applied to the grouping of Batu City community characters based on waste management behavior with a Silhouette Index value of 0.875.
2. Cluster 1 consists of people with good environmental quality, good use of waste banks, and a good understanding of 3R-based waste management (Reduce, Reuse, Recycle).
3. Cluster 2 is characterized by communities with good quality facilities and infrastructure and a good understanding of the economic benefits of waste.
4. Cluster 3 includes groups with a good understanding of 3R-based waste management (Reduce, Reuse, Recycle) and good environmental quality.

## REFERENCES

[1] J. F. Hair, W. C. Black, and R. E. Anderson, MULTIVARIATE DATA ANALYSIS *(5th ed.)*. 1998.

[2] I. D. Id, A. Astrid, and E. Mahdiyah, "MODIFIKASI DBSCAN (DENSITY-BASED SPATIAL CLUSTERING WITH NOISE) PADA OBJEK 3 DIMENSI," *Jurnal Komputer Terapan*, vol. 3, no. 1, pp. 41–52, 2017, [Online]. Available: http://jurnal.pcr.ac.id

[3] E. T. K. Dewi, A. Agoestanto, and S. Sunarmi, "METODE LEAST TRIMMED SQUARE (LTS) DAN MM-ESTIMATION UNTUK MENGESTIMASI PARAMETER REGRESI KETIKA TERDAPAT OUTLIER," *UNNES Journal of Mathematics*, vol. 5, no. 1, pp. 47–54, 2016.

[4] T. N. Tran, K. Drab, and M. Daszykowski, "REVISED DBSCAN ALGORITHM TO CLUSTER DATA WITH DENSE ADJACENT CLUSTERS," *Chemometrics and Intelligent Laboratory Systems*, vol. 120, pp. 92–96, Jan. 2013, doi: 10.1016/j.chemolab.2012.11.006.

[5] P. B. Nagpal and P. A. Mann, "COMPARATIVE STUDY OF DENSITY BASED CLUSTERING ALGORITHMS," *Int J Comput Appl*, vol. 27, no. 11, pp. 44–77, Dec. 2011, doi: 10.5120/3341-4600.

[6] S. U. Rehman, S. Asghar, S. Fong, and S. Sarasvady, "DBSCAN: PAST, PRESENT AND FUTURE," in *The Fifth International Conference on the Applications of Digital Information and Web Technologies (ICADIWT 2014)*, IEEE, Feb. 2014, pp. 232–238. doi: 10.1109/ICADIWT.2014.6814687.

[7] Z. Wei, Y. Gao, X. Zhang, X. Li, and Z. Han, "ADAPTIVE MARINE TRAFFIC BEHAVIOUR PATTERN RECOGNITION BASED ON MULTIDIMENSIONAL DYNAMIC TIME WARPING AND DBSCAN ALGORITHM," *Expert Syst Appl*, vol. 238, p. 122229, Mar. 2024, doi: 10.1016/j.eswa.2023.122229.

[8] A. Fauzan, G. Fadillah, A. Fitria, H. Adriana, and M. Bariklana, "CLUSTER MAPPING OF WASTE EXPOSURE USING DBSCAN APPROACH: STUDY OF SPATIAL PATTERNS AND POTENTIAL DISTRIBUTION IN BANTUL REGENCY," *JOIV : International Journal on Informatics Visualization*, vol. 8, no. 2, p. 751, May 2024, doi: 10.62527/joiv.8.2.2475.

[9] "LAPORAN KINERJA 2023," 2023.

[10] A. Kassambara, *PRACTICAL GUIDE TO CLUSTER ANALYSIS IN R: UNSUPERVISED MACHINE LEARNING*, vol. 1. Sthda, 2017.

[11] K. Wright, "WILL THE REAL HOPKINS STATISTIC PLEASE STAND UP?," *R Journal*, vol. 14, no. 3, 2022.

[12] R. G. Lawson and P. C. Jurs, "NEW INDEX FOR CLUSTERING TENDENCY AND ITS APPLICATION TO CHEMICAL PROBLEMS," *J Chem Inf Comput Sci*, vol. 30, no. 1, pp. 36–41, Feb. 1990, doi: 10.1021/ci00065a010.

[13] R. A. Johnson and D. W. Wichern, *APPLIED MULTIVARIATE STATISTICAL ANALYSIS (6th ed.)*. New Jersey: Pearson education, Inc, 2007.

[14] E. Prasetyo, DATA MINING MENGOLAH DATA MENJADI INFORMASI MENGGUNAKAN MATLAB. 2014.

[15] A. S. Devi, I. K. G. D. Putra, and I. M. Sukarsa, "IMPLEMENTASI METODE CLUSTERING DBSCAN PADA PROSES PENGAMBILAN KEPUTUSAN," *J. Ilm. Teknol. Inf*, vol. 6, no. 3, p. 185, 2015.

[16] S. Scitovski, "A DENSITY-BASED CLUSTERING ALGORITHM FOR EARTHQUAKE ZONING," *Comput Geosci*, vol. 110, pp. 90–95, Jan. 2018, doi: 10.1016/j.cageo.2017.08.014.

[17] P. N. Tan, M. Steinbach, and V. Kumar, *DATA MINING CLUSTER ANALYSIS: BASIC CONCEPTS AND ALGORITHMS*. 2013.

[18] W. Yustanti, "ALGORITMA K-NEAREST NEIGHBOUR UNTUK MEMPREDIKSI HARGA JUAL TANAH," vol. 9, no. 1, pp. 57–68, 2012.

[19] B. Mu, M. Dai, and S. Yuan, "DBSCAN-KNN-GA: A MULTI DENSITY-LEVEL PARAMETER-FREE CLUSTERING ALGORITHM," *IOP Conf Ser Mater Sci Eng*, vol. 715, no. 1, p. 012023, Jan. 2020, doi: 10.1088/1757-899X/715/1/012023.

[20] D. Jollyta, M. Siddik, H. Mawengkang, and S. Efendi, *TEKNIK EVALUASI KLASTER SOLUSI MENGGUNAKAN PYTHON DAN RAPIDMINER*. Deepublish, 2021.

[21] J. F. J. Hair, W. C. Black, B. J. Babin, and R. E. Anderson, MULTIVARIATE DATA ANALYSIS *(7th ed.)*. Prentice Hall, 2009.