

Cluster Mapping of Waste Exposure Using DBSCAN Approach: Study of Spatial Patterns and Potential Distribution in Bantul Regency

Achmad Fauzan ^{a,*}, Ganjar Fadillah ^b, Annisa Fitria ^c, Hannura Adriana ^a, Muhammad Bariklana ^a

^a Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Islam Indonesia, Sleman, Sleman, Indonesia

^b Department of Chemistry, Faculty of Mathematics and Natural Sciences, Universitas Islam Indonesia, Sleman, Sleman, Indonesia

^c Department of Pharmacy, Faculty of Mathematics and Natural Sciences, Universitas Islam Indonesia, Sleman, Sleman, Indonesia

Corresponding author: *achmadfauzan@uii.ac.id

Abstract— High level of plastic waste production is a common issue in various places, including the Special Region of Yogyakarta province. It is proven that the Piyungan Integrated Waste Disposal Site (TPST) or Final Disposal Site (TPA) in Bantul Regency has been closed several times due to capacity exceeding the quota and some blockages from residents around the TPA. Issues related to microplastic contamination resulting from discarded plastic waste are fascinating to study, considering the long-term impact of microplastic contamination on human health. This research aims to map the distribution of locations with the potential for waste accumulation to reduce the negative consequences of microplastic contamination. The population used included TPS and markets in Bantul District, with the sample being the distribution of TPS and market points in Bantul District in 2023. The results of checking the point distribution pattern using the quadrant and nearest neighbor method showed that the distribution of waste accumulation points had a clustered design; thus, it could be continued with cluster analysis using the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) method. Meanwhile, the number of clusters was determined iteratively from a combination of DBSCAN parameters (MinPts and Epsilon). The best method was evaluated based on the Silhouette Coefficient (SC) value, which in this case was 0.78 (MinPts 7 and Eps 1500) and included in the strong category. Subsequently, exploration was carried out by reducing the MinPts value and the lowest limit value of strong SC.

Keywords—DBSCAN; spatial patterns; waste exposure.

Manuscript received 10 Jan. 2024; revised 15 Feb. 2024; accepted 20 Mar. 2024. Date of publication 31 May 2024.
International Journal on Informatics Visualization is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.



I. INTRODUCTION

Increasing population and industrialization have led to several environmental problems in recent decades, including plastic and microplastic contamination. This problem has been going on for generations and has always been a topic of social polemic. It is stated to have been going on for generations as it has existed from the past until now and has not even been able to be appropriately resolved. It is stated as a social polemic since it always causes restlessness in everyday life. The Piyungan Final Disposal Site (TPA) or Integrated Waste Disposal Site (TPST) is the largest landfill in the Special Region of Yogyakarta (DIY) province. Located in Ngablak and Watugender Hamlet, Sitimulyo Village, Piyungan District, Bantul Regency, it is the final processing of waste originating from Yogyakarta City, Bantul Regency, and Sleman Regency. Until now, it is still actively functioning, and the waste volume continuously increased yearly [1]. The volume of waste and the various problems it

causes increase yearly; unfortunately, it does not align with the public's perception of the waste problem. Meanwhile, the waste produced by everyone is not a matter for the community.

Recently, impact and various problems, especially health, can be caused by the waste content itself. Plastic waste, for example, is always present from time to time. Plastic is a material that cannot be naturally decomposed (non-biodegradable). After being used, fabrics made from plastic will become waste that is difficult to decompose by soil microbes, polluting the environment [2]. Environmental pollution by plastic is now widely recognized as a significant burden of environmental pollution [3], especially in aquatic environments where it has prolonged biophysical damage [4-7], detrimental adverse effects on wildlife [8-12], and limited plastic removal options [13-16]. From a health perspective, the content of plastic waste also impacts human health. Several current issues include plastic waste contamination, which can be a source of hazardous

microplastic contamination. Microplastics are tiny plastic particles that are not more than 5 millimeters long and can accumulate in various environments, including sea, air, and soil. Microplastic pollution can harm the environment, human health, and biodiversity [17], [18]. Previous research reported that microplastic contamination can bioaccumulate and biomagnify in aquatic food, impacting the environment and human health [19]. In addition, the level of microplastic pollution in terrestrial ecosystems is 20 times higher than in the ocean.

Based on these problems above, the initiation of this research focuses on clustering the distribution of waste collection centers (TPS/TPA) to determine the group of these TPAs. Plastic waste mapping becomes critical to anticipate microplastic pollution in the environment. It is carried out to identify locations as the primary sources of microplastic pollution, such as illegal waste dumps or densely populated urban areas. Mapping can also help in planning more effective and efficient waste management. Density-based spatial clustering of applications with noise (DBSCAN), an algorithm where the cluster formation process is carried out based on the density level of distance between objects in the dataset, is used in spatial cluster analysis. It has the advantage of being able to detect noise/things that are not included in any clusters [20].

II. MATERIAL AND METHOD

A. Research Population and Sample

The population of this study refers to a distribution of TPS points and locations that have the potential to have a lot of waste, such as markets in Bantul District. At the same time, the samples used in the research included TPS and markets in Bantul Regency in 2023. The sample above was used because

these centers have potential waste piles. Sample data was obtained using Google Earth Pro, Google Earth Engine, and Botsol Google Maps Crawler in 2023.

B. Research Variable

The variables in this research included the location of coordinate points from the distribution of polling stations and markets in the Bantul District. The coordinate system used was the Universal Transverse Mercator (UTM) coordinate projection system. The data used is illustrated in Table 1.

C. Research Stage

The initial stage aimed to search for point distribution data and continued with data input. The next step was data preprocessing. Some of the treatments carried out included eliminating points outside the Bantul Regency and eliminating data redundancies. The third stage was Exploratory Data Analysis (EDA) in which point distribution data and initial visualization of point density were carried out. The next stage was checking the distribution of points and whether the points were spread in groups. The Quadrant and nearest neighbor methods were used to check the spatial distribution of the points. If the points were spread out in groups, then the DBSCAN clustering method could be continued. However, other methods, such as zoning mapping for each sub-district, can be used if it does not spread in groups. The further stage was to calculate the DBSCAN parameters, i.e., MinPts and Epsilon, to determine the number of clusters formed. The MinPts and Epsilon values were obtained from the second iteration and evaluated using the silhouette coefficient according to Equation 4. The sixth stage explored the parameter values and interpretation of the results through visualization. Overall, the research stages are presented in Fig. 1.

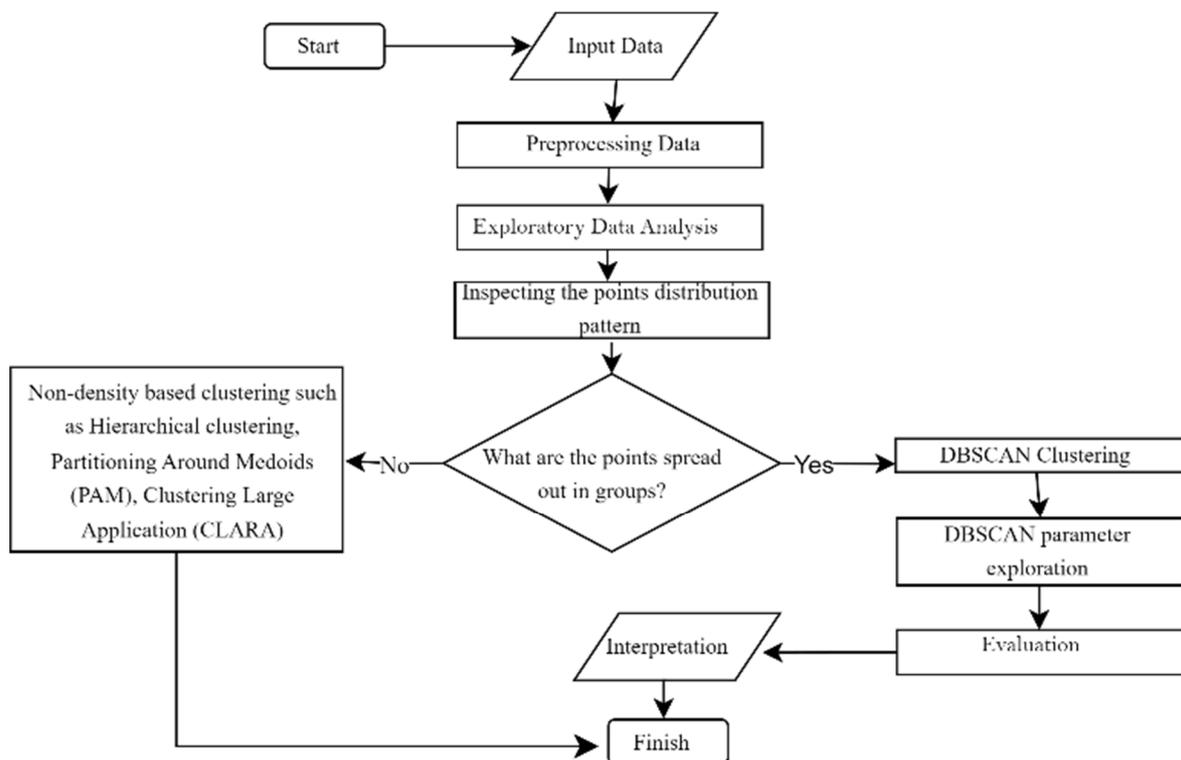


Fig. 1 Research flow diagram.

D. Toxicity of microplastics

It is reported that microplastic contamination exists in various biophysical environments such as water, sediment, soil, and air. Microplastics have a toxic impact on both the environment and human health. Some of their toxic effects on the environment include: (1) Accumulation in organisms: microplastics can be consumed by marine organisms, such as plankton and fish, and accumulate in the tissues of these marine organisms. This impacts the health of the organism and potentially damages the food chain. (2) Environmental pollution: microplastics that are thrown into the environment can cause water and soil pollution. This can threaten the survival of organisms and disrupt the ecosystem. (3) Increased exposure to toxic chemicals: microplastics can absorb toxic chemicals found in the environment, such as

pesticides and heavy metals. Organisms that consume microplastics contaminated with these chemicals can experience exposure to toxic chemicals that impact their health and metabolic systems. Inflammation of organs, cancer, and other metabolic diseases can result from the accumulation of these substances in the body. Lastly, (4) Disturbances to the development and reproduction of organisms: microplastics can affect the growth and reproduction of marine organisms. Several studies in fish showed that exposure to microplastics can disrupt embryo development and reduce fertility [21]. However, microplastic contamination in the environment can be toxic to human health through the food chain cycle. Plastic waste mapping, as one of the primary sources of microplastic contamination, is an essential step in reducing the effects of microplastics.

TABLE I
ILLUSTRATION OF RESEARCH DATA

No.	Name	Type	City	mE	mS	Zone
1.	TPST 3R Palbapang Bantul	Laystall	Bantul Regency	425596.4775	9126483.913	49 S
2.	TPST 3R Himamimmo Grugugan Bantul	Laystall	Bantul Regency	425701.4952	9130488.03	49 S
⋮	⋮	⋮	⋮	⋮	⋮	⋮
143	Celep Market	Traditional market	Bantul Regency	421316.8551	9119021.753	49 S

E. Configuration Points in The Space

The basic idea behind point pattern investigations is to compare theoretical assumptions about their behavior with empirical evidence from data. Spatial point pattern methods were developed in biology and ecology as early as the mid-20th century, and they became most popular in economics. Ripley's K is a popular approach for determining point distribution clustering. Some of the studies completed using the Ripley's K approach are Vidanapathirana [22], Self [23], [24], and Tzai-Hung [25].

A Poisson point process (also known as complete spatial randomness [CSR] or complete spatial randomness and

independence [CSRI] for multitype point patterns) is a classic example of a point pattern that reveals random behavior, stationarity, and isotropy (which means it retains statistical properties under any shifts or rotations). Such patterns are uncommon, but they serve as a useful benchmark, particularly when early analysis does not indicate CSR as a starting point. Point pattern analysis is beneficial for determining whether an underlying data generation process is random or non-random and for investigating probable links between points of one or more categories. Verifying a CSR/CSRI hypothesis can result in either non-rejection, in which case the studied pattern is Poisson, or rejection, in which case one of three possibilities is possible. Fig. 2 depicts all of them.

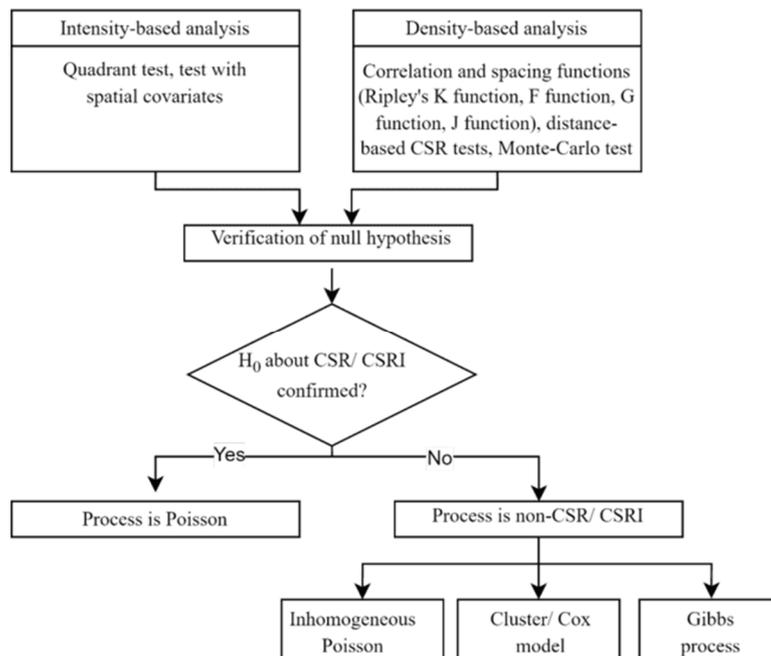


Fig. 2 Algorithm of univariate pattern analysis [26]

In approaching the point distribution pattern, the quadrant method, distance-based CRS, and Nearest-neighbor test function were utilized, as shown in Fig. 2. Nearest-neighbor analysis is a method for measuring patterns in two or three dimensions in a data point. The average point distance between all points and their nearest neighbors was calculated using this approach [27]. The nearest-neighbor index was calculated by dividing the observed distance by the expected distance. The distance utilized in Equation (1) is the Euclidean distance. R represents the nearest-neighbor index, and the calculation is described in Equation (2) [28].

$$d(i, j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{ip} - x_{jp})^2} \quad (1)$$

$$R = \frac{\bar{d}_{obs}}{\bar{d}_{ran}} \quad (2)$$

\bar{d}_{obs} is the observed mean closest-neighbor distance or the average distance between each point and the nearest neighbor. \bar{d}_{ran} is the predicted mean nearest-neighbor distance or the expected average distance for the provided random points. Equation (3) shows the calculation of \bar{d}_{ran} and \bar{d}_{obs} .

$$\bar{d}_{obs} = \frac{\sum_{i=1}^n d_i}{n} \text{ and } \bar{d}_{ran} = \frac{1}{2\sqrt{p}} = \frac{1}{2\sqrt{n/A}} \quad (3)$$

The distance between i and its nearest neighbor is denoted by d_i , the number of points is denoted by n , and A is the minimal rectangular area around all points, or a set area value. c is the test statistics, and $SE_{\bar{d}}$ is the standard error of the mean nearest neighbor, using the calculation algorithm specified in Equations (4) and (5).

$$Z_{value} = \frac{\bar{d}_{obs} - \bar{d}_{ran}}{SE_{\bar{d}}} \quad (4)$$

$$SE_{\bar{d}} = \frac{0.26136}{\sqrt{np}} = \frac{0.26136}{\sqrt{n^2/A}} \quad (5)$$

The nearest-neighbor index ranges from 0 to 2.15. A nearest-neighbor index of 0 implies that the pattern is completely clustered. While, the nearest-neighbor index (2.15) exhibits a fully dispersed pattern. The nearest-neighbor index value of 1 indicates a random pattern. The point configuration is stated to be categorized as a cluster if the calculated $Z_{value} < -1.96$ (significance level $\alpha = 5\%$) [29]. In addition to the quadrant approach, the quadrant method is utilized to analyze the point distribution of the spatial structure. The quadrant approach begins by dividing the area into m parts of roughly equal size. Then, suppose n , count the total number of occurrences in that region. The next step is to compute the average number of events per cell (\bar{x}), the variance of the number of events per cell (s^2), and the Variance-Mean Ratio (VMR), as shown in Equation 6.

$$VMR = \frac{s^2}{\bar{x}} \quad (6)$$

$$\bar{x} = \frac{n}{m} \text{ and } s^2 = \frac{\sum_{i=1}^m (x_i - \bar{x})^2}{m-1}$$

m : the number of cells per grid, n : the total number of occurrences, and x_i : the number of events in cell i . The VMR value is the variance-to-mean ratio. If the value of $VMR = 0$, the point arrangement in space is uniform or perfectly regular.

If VMR is equal to one, the organization of points in space is random. Similarly, a value of VMR 1 (the variance value less than the average value) implies that the point arrangement in space is more inclined to the regular shape. Meanwhile, $VMR > 1$ indicates that locations in space are more clustered than random. If the number of quadrants is less than 30, then the $(m - 1)$ VMR is distributed according to the Chi-Square distribution with degrees of freedom $m - 1$. Equation 7 displays the Chi-Square test statistics.

$$\chi^2 = \frac{\sum_{i=1}^m (x_i - \bar{x})^2}{\bar{x}} \quad (7)$$

The null hypothesis (H_0) is that the arrangement of points in space is random with a statistical calculation of $(m - 1)VMR$. If the value $\chi^2 > \chi_{table}^2$, reject H_0 . Meanwhile, if m is more than 30, $(m - 1)VMR$ spreads according to the normal distribution $(m - 1, 2(m - 1))$ [30]. If the distribution pattern of these points creates a cluster, the DBSCAN Algorithm is used to continue spatial clustering. The DBSCAN technique can detect outliers/noise without requiring any inputs in the form of the number of clusters (k) like the K-Means or K-Medoid methods do. This approach also has the advantage of being able to detect any problematic/irregular clusters [30 - 31]

F. Density-based Spatial Clustering of Applications with Noise (DBSCAN)

Density-based spatial clustering of applications with noise (DBSCAN) requires two parameters: minimum points (MinPts) and epsilon (Eps), the values of which are specified by the researcher. MinPts is the minimal number of items in a cluster, and Eps is the value for the distance (radius) between items that serves as the foundation for establishing an item point's neighborhood. Neighborhoods inside the radius (ϵ) of the data item are referred to as ϵ -neighborhoods [33]. An item is considered a core object if its ϵ -neighborhood includes at least as many objects as MinPts. Core objects are density pillars. In a DBSCAN clustering, there are two sorts of points: core points and border points with the neighborhood of border points containing much fewer items compared to the neighborhood of core points [34]. A boundary point may be associated with more than one cluster. Given a collection of objects D and the parameters and MinPts, it is feasible to identify all core objects. It reduces clustering by leveraging core objects and neighbors to build a density, where density areas are clusters [35].

DBSCAN searches for clusters by examining the ϵ -neighborhood (Eps-neighborhood) of each point in the database. If the ϵ -neighborhood of point p contains more than MinPts, a new cluster is generated with p as the core object. Then, DBSCAN repeatedly gathers density-reachable items directly from the core object which may include merging numerous density-reachable clusters. The DBSCAN algorithm works as follows: (1) randomly choosing an initial point p , (2) determining Eps and MinPts to retrieve all points that are density reachable to point p , (3) if p is a core point, a cluster is formed, (4) if p is a border point, there is no density-reachable relationship from p and DBSCAN will visit the next point from the database, (5) continuing the process until all points have been processed. The acquired findings are the independent of the order of the retrieved point process [33].

Iteratively, the choice of MinPts and Eps is made. The best Eps determination is made by measuring the distance around the elbow. DBSCAN clustering is performed once obtaining an ideal epsilon value, such that clusters from each site are created based on their density, meet MinPts and Eps, and then displayed.

The first phase involves calculating the inter-TPS distances once spatial data comprising geospatial information about the TPS has been acquired. Next, the identification of both parameters (Minpts and epsilon value) in the DBSCAN technique makes it easier to use the obtained results to identify TPS clusters or groups based on geographic density. For instance, let us examine a situation in Kabupaten Bantul where TPS are dispersed over the urban area, but at varying densities. In this case, DBSCAN is used to identify denser-aggregated TPS clusters that are closer to metropolitan centers or areas with denser populations. On the other hand, TPS located in rural areas with lower population density might be considered noise or irrelevant to major cluster forms.

G. Evaluation of Cluster Model

The average silhouette width is the cluster's evaluation stage. Equation 8 depicts the Silhouette method's estimate [35 - 37]. Assume that for data point $i \in C_i$ (data point i in the cluster C_i).

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (8)$$

$$a(i) = \frac{1}{|C_i| - 1} \sum_{j \in C_i, i \neq j} d(i, j) \quad (9)$$

$$b(i) = \min_{k \neq i} \frac{1}{|C_k|} \sum_{j \in C_k} d(i, j)$$

The silhouette (value) at point i in a cluster is denoted by $s(i)$. $b(i)$ is the average distance between the i^{th} point and all data not in the same cluster as the i^{th} point. $a(i)$ is the average distance between the i^{th} point and all data in a cluster. $|C_i|$ denotes the number of points in cluster i . The distance between data objects i and j in the cluster C_i is given by $d(i, j)$. After calculating the average silhouette width value, the average silhouette width is stated to be strong if it is higher than 0.71 [39].

III. RESULT AND DISCUSSION

A. Exploratory Data Analysis

The first step was to prepare the data. It started with data profiling, which was conducted by defining and transforming the data used (in the context of this research, the distribution of points with the potential for plastic waste accumulation). Furthermore, data preprocessing was carried out by eliminating the distribution of points not included in the

points under study, in this case, the Bantul Regency, and elimination if data were duplicated. The results of the first stage were then visualized in the density map as presented in Fig. 3.

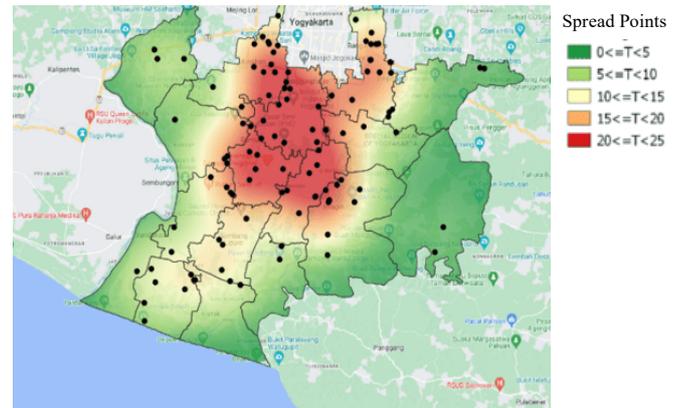


Fig. 3 Point spread density map.

As shown in Fig. 4, the black dots refer to the distribution of the observed objects (markets and landfills) in the Bantul Regency. However, the area of each point was not included in this study due to limited data obtained. From the distribution of these points, density gradation was carried out with five different levels. Green color shows the lowest density of points (also stated to have the least potential in microplastic contamination), while red areas are the most densely distributed areas (as seen in the legend ranging from 20 to 25 points). When observed further, green areas are those located on the edges of Bantul Regency, while in the center of the regency are the areas that fall into the dense category. At first glance, the distribution of points had a group pattern, located around the center of Bantul Regency. However, it still needed to be confirmed by using statistical tests to confirm it by using the quadrant method and nearest neighbor analysis such as Equation (5) dan (7).

B. Spatial Point Pattern

The location variable was used to determine the point distribution pattern. The projection coordinate system was used in this calculation. If the coordinates used were still in the geographic coordinate system, conversion could be done first. Some software can do this, such as Quantum Geographic Coordinate System (QGIS) [40] or ArcGIS [41]. This research used the Universal Transverse Mercator in zone 49 S (because the Bantul regency is in that zone). In the analysis stage of the quadrant method, based on the equation, simulations were carried out by dividing the area with $m = 3$ (3x3) by $m = 20$ (20x20). Table 2 presents the results of the quadrant method.

TABLE II
RESULT OF THE QUADRANT METHOD

No.	Quadrant Size	Number Of Quadrants	Mean	Variance	Chi-Square/ Z Count	Chi-Square/ Z Table	Point Distribution Pattern
1.	3X3	9	12.67	182.00	64.23	16.91	CLUSTER
2.	4X4	16	7.13	54.52	78.34	26.29	CLUSTER
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
18.	20X20	400	0.44	0.66	8.38	1.96	CLUSTER

The table shows that the calculated chi-square value was higher than the chi-square table value for a large m between 3 and 5. For a matter of m more than 5, the computed z value was also higher when compared to the z table value using the calculation in Equation 2. Based on the above review, it can be concluded that the research object points had a cluster pattern. In addition, it is strengthened by the simulation of the VMR value in Fig 4. The VMR value with the calculation in Equation 1 was consistently worth more than one, even though the simulation was carried out by increasing the number of quadrants (reducing the size of each grid). Based

on this, the points could be declared to spread in groups when approached by the quadrant method.

Furthermore, a nearest-neighbor approach was also carried out to determine the point distribution pattern. Initiated from the nearest-neighbor index (R) value using Equation 1 of 0.04 (close to the value of 0) or tended to lead to a clustered distribution pattern. Based on Equation 1, the calculated z value was $-26.95 < -1.96$, so the points spread in groups. Based on these two methods, the point distribution pattern in this study tended to be clustered so that it could be continued in the DBSCAN clustering analysis.

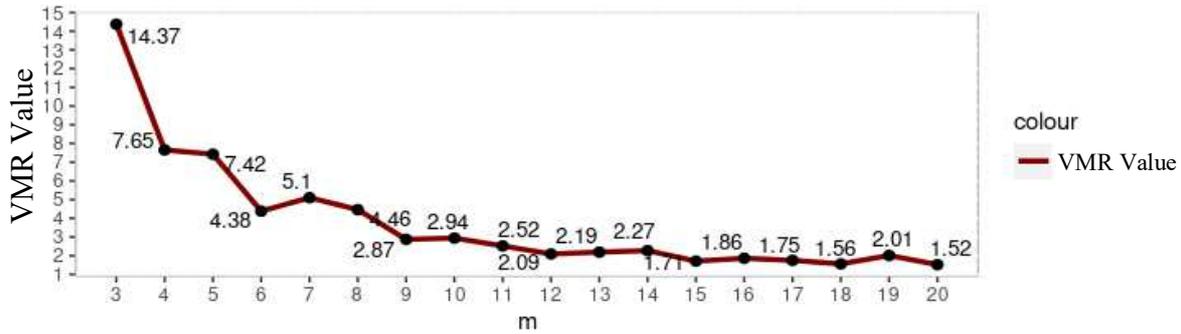


Fig. 4 VMR value

TABLE III
SILHOUETTE COEFFICIENT EVALUATION RESULTS

No	Minpts	Eps	Silhouette Coefficient	Number Of Cluster	Multiple Points On Noise	Multiple Points In Cluster
1	3	1500	0.503	12	32	82
⋮	⋮	⋮	⋮	⋮	⋮	⋮
101	3	2500	0.381	3	8	106
102	4	1500	0.583	9	44	70
⋮	⋮	⋮	⋮	⋮	⋮	⋮
202	4	2500	0.641	5	61	63
⋮	⋮	⋮	⋮	⋮	⋮	⋮
505	7	2500	0.355	3	26	88

C. DBSCAN Clustering and Parameter Exploration

The DBSCAN method has two primary parameters: MinPts and Eps. Before determining the two best parameters, the k-Nearest Neighbor Distance (kNNdist) is determined, which is described as a rapid computation of the k-nearest neighbor distance in the points matrix. The findings of kNNdist are utilized to select the suitable eps and MinPts areas. The kNNdist plot will eventually form a knee that can be used to anticipate the ideal MinPts and Eps values. An illustration of the elbow in this algorithm is presented in Fig. 5. There is a blue dotted line, which is the iteration limit used. Iteration was performed in the area whose MinPts lay on the elbow 3 to 7, and the Eps value was between 1500 and 2500 (in increments of 10 meters). Based on the silhouette coefficient, the optimal cluster was chosen by combining these two characteristics. Table 3 displays the results of the combination of the two parameters.

The maximum silhouette coefficient value was 0.782, namely at MinPts=7 and Eps=1500 with 3 clusters formed, 81

noises, and the number of points in the cluster = 33 (shown at label “a” in Fig. 6). A silhouette coefficient value of 0.782 meant that the cluster had a strong structure [38]. The average silhouette values of the simulated MinPts and Eps values are presented in Fig. 6. Based on the figure, mainly after the Eps value of 1700, the Average Silhouette value tended to decrease compared to the previous value. This meant that although the radius was vast, the strength of the cluster formed tended to weaken. In addition to clustering based on the highest silhouette value, data exploration was carried out with the aim of disaster mitigation (especially garbage pollution). Disaster mitigation is a study purposely to minimize the impact of losses due to disaster events, both material losses and moral losses [42]. Exploration was conducted by reducing the MinPts value but still taking the most optimum Eps. Fig. 6 presents labels “b” and “c”. Furthermore, exploration was conducted using the Average Silhouette value, which approached the smallest threshold of the Average Silhouette value, which was strong (0.71).

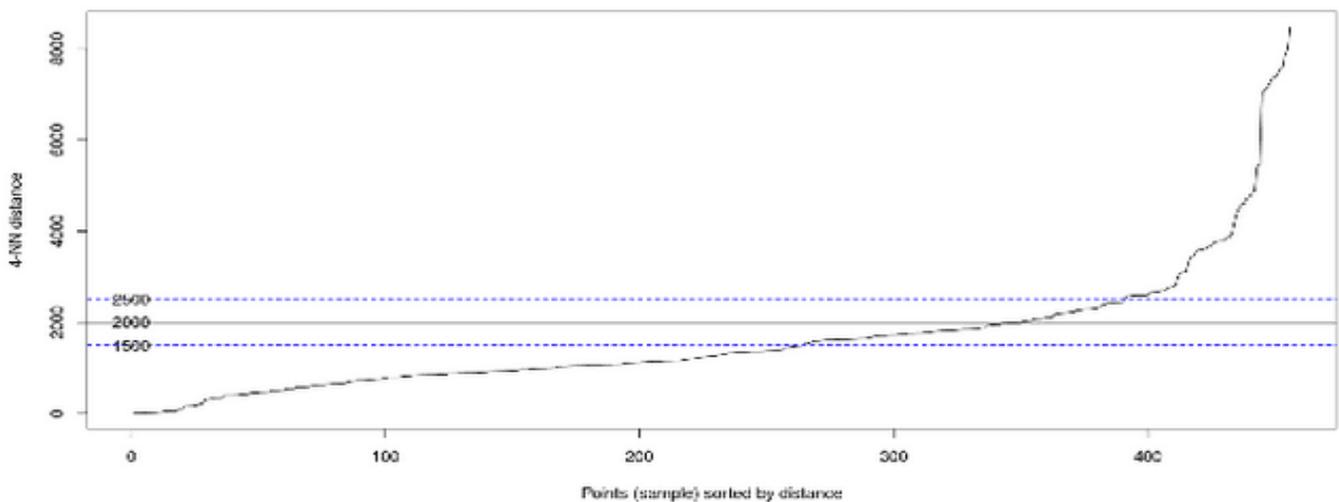


Fig. 5 The iteration limit for Eps parameter

D. Clustering by Lowering the MinPts Value

As shown in Fig. 6, with the criteria of average silhouette value closest to strong, two possible parameter combinations were obtained, namely 1. MinPts = 5 and Eps = 1500 with an Average Silhouette value of 0.641 (given in label “c”). Then the second exploration was MinPts = 6 and Eps = 1500 with Average Silhouette value = 0.678 (presented in label “b”). From the first exploration results, the number of clusters was obtained to be 4 groups with a noise count of 72 (while the points in the cluster were 42). This result increased the number of clusters compared to the initial clustering and decreased the noise. Likewise, the results obtained in the second exploration showed 5 clusters with 61 noises and 53 points in the cluster. The illustrations of the results of the first and second explorations are presented in Fig. 7.

E. Clusterization with a Value Close to the Minimum Average Ailhouette Limit

The MinPts and Eps values that met the Average Silhouette closest to the minor threshold were identified as vital in the third exploration. From the simulation results, a MinPts value 7 was obtained with an Eps length of 1760 m. The Average Silhouette value was 0.778. In this third exploration, there were 3 clusters and 78 noise locations. However, in this third exploration, the number of points in the noise was more significant than the number of points in the cluster (the number of points in the noise was 78, while the number of points in the cluster was 36 locations).

Generally, the spatial distribution of Waste Collection Points (TPS) within Bantul Regency exhibits a pronounced clustering pattern, mainly concentrated in urban cores. This spatial phenomenon is attributed to the accessibility factor influencing community waste disposal practices, indicating a notable association between TPS locations and the public service demands pertinent to waste management. This observation assumes significance within the backdrop of escalating urbanization trends in Bantul Regency, where

providing sanitation services and waste management infrastructure is paramount for urban dwellers. Nevertheless, the aggregation of TPS within urban settings also engenders adverse repercussions, including the potential for environmental degradation stemming from waste accumulation and the heightened risk of disease transmission, underscoring the imperative for meticulous consideration within sustainable urban development frameworks.

Conversely, spatial analysis elucidates that specific locale within the southeastern and southern territories of Bantul Regency, notably encompassing the districts of Dlingo, Pundong, and Kretek, exhibit a deficiency in TPS infrastructure. These areas, characterized by diminished population densities and distinct geographical features, present themselves as prospective candidates for future centralized TPS deployments, subject to the caveat of pertinent environmental impact mitigation measures. Furthermore, their strategic geographical positioning, such as proximity to elevated terrain or coastal regions, presents opportunities to adopt innovative waste management strategies, including waste-to-energy conversion technologies or alternative waste disposal methodologies. However, it is imperative to acknowledge that determining TPS locations constitutes merely the nascent phase within a holistic waste management endeavor. Effective and sustainable waste management necessitates comprehensive planning, encompassing the integration of processing facilities and recycling initiatives.

Consequently, development strategies amalgamating judicious TPS siting with efficacious waste management frameworks emerge as indispensable requisites in pursuing sustainable, eco-centric urban development paradigms. Hence, the fortification of collaborative partnerships among governmental bodies, academic institutions, and local stakeholders assumes heightened significance in devising tailored solutions to surmount waste management challenges in Bantul Regency, thereby advancing the broader agenda of sustainable development objectives.

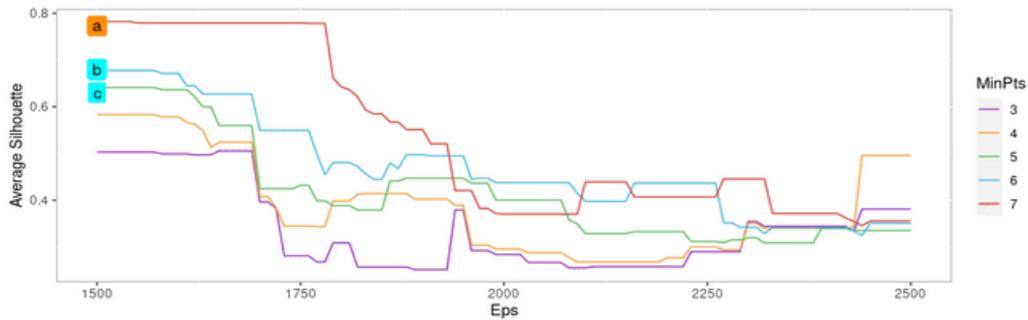


Fig. 6 Average Silhouette Width of MinPts and Eps simulations

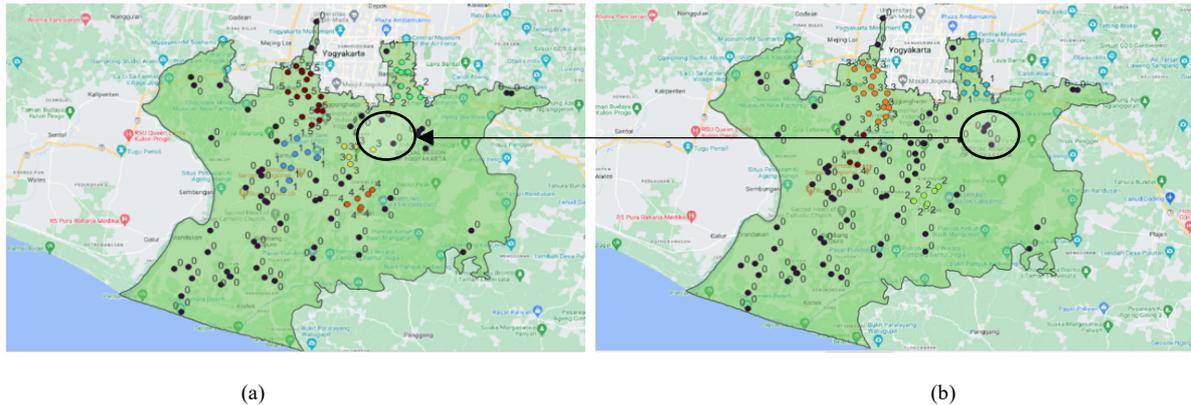


Fig. 7 (a) The results of first exploration of DBSCAN, (b) The results of second exploration of DBSCAN

IV. CONCLUSIONS

Based on the results of point density exploration, most of the potential waste centers were concentrated around the central government of the Bantul District. Then, based on the quadrant and nearest neighbor methods, the results showed that the distribution pattern of potential waste contamination areas was clustered. A simulation was conducted using a combination of MinPts with an interval of 3 to 7 and Eps from 1500 to 2500 based on the area around the elbow. From the simulation results, the DBSCAN clustering results obtained an optimum average silhouette value of 0.783, which is in the strong category. These results showed that the number of clusters was 3 with parameter values MinPts = 7 and Eps = 1500. Then, to mitigate the disaster population of waste pollution, exploration was carried out with 3 scenarios: decreasing the MinPts value (5 and 6), and DBSCAN using the value Eps and MinPts. From the results of this exploration, it was obtained that the group was formed and remained in the center of government.

ACKNOWLEDGMENT

The authors thank the Faculty of Mathematics and Natural Sciences at Universitas Islam Indonesia for their valuable support through the Collaboration Grant program.

REFERENCES

- [1] S. D. Bantul, "Volume Sampah Yang Terangkut Ke TPA," *Satu Data Bantul*. 2020.
- [2] J. Wahyudi, H. T. Prayitno, A. Dwi, A. B. Perencanaan, P. Daerah, and K. Pati, "The Utilization of Plastic Waste as Raw Material For

- Producing Alternative Fuel," *Jurnal Litbang*, vol. XIV, no. 1, pp. 58–67, 2018.
- [3] R. Kumar *et al.*, "Impacts of Plastic Pollution on Ecosystem Services, Sustainable Development Goals, and Need to Focus on Circular Economy and Policy Interventions," *Sustainability*, vol. 13, no. 17, p. 9963, Sep. 2021, doi: 10.3390/su13179963.
- [4] J. G. B. Derraik, "The pollution of the marine environment by plastic debris: a review," *Marine Pollution Bulletin*, vol. 44, no. 9, pp. 842–852, Sep. 2002, doi: 10.1016/s0025-326x(02)00220-5.
- [5] L. Peng, D. Fu, H. Qi, C. Q. Lan, H. Yu, and C. Ge, "Micro- and nano-plastics in marine environment: Source, distribution and threats — A review," *Science of The Total Environment*, vol. 698, p. 134254, Jan. 2020, doi: 10.1016/j.scitotenv.2019.134254.
- [6] P. Alam, D. Mamalis, C. Robert, C. Floreani, and C. M. Ó Brádaigh, "The fatigue of carbon fibre reinforced plastics - A review," *Compos B Eng*, vol. 166, pp. 555–579, Jun. 2019, doi:10.1016/j.compositesb.2019.02.016.
- [7] K. P. Isaacson *et al.*, "Drinking water contamination from the thermal degradation of plastics: implications for wildfire and structure fire response," *Environ Sci (Camb)*, vol. 7, no. 2, pp. 274–284, 2021, doi:10.1039/D0EW00836B.
- [8] M. Thiel *et al.*, "Impacts of Marine Plastic Pollution From Continental Coasts to Subtropical Gyres—Fish, Seabirds, and Other Vertebrates in the SE Pacific," *Front Mar Sci*, vol. 5, 2018, doi:10.3389/fmars.2018.00238.
- [9] S. Avery-Gomm, S. B. Borrelle, and J. F. Provencher, "Linking plastic ingestion research with marine wildlife conservation," *Science of The Total Environment*, vol. 637–638, pp. 1492–1495, 2018, doi:10.1016/j.scitotenv.2018.04.409.
- [10] C. G. Awuchi and C. G. Awuchi, "Impacts of plastic pollution on the sustainability of seafood value chain and human health," *International Journal of Advanced Academic Research*, vol. 5, no. 11, pp. 46–138, 2019.
- [11] N. A. Welden, "Chapter 8 - The environmental impacts of plastic pollution," in *Plastic Waste and Recycling*, T. M. Letcher, Ed., Academic Press, 2020, pp. 195–222. doi: 10.1016/B978-0-12-817880-5.00008-6.
- [12] J. Senko, S. Nelms, J. Reavis, B. Witherington, B. Godley, and B. Wallace, "Understanding individual and population-level effects of

- plastic pollution on marine megafauna," *Endanger Species Res*, vol. 43, pp. 234–252, Oct. 2020, doi: 10.3354/esr01064.
- [13] C. Wilcox, E. Van Sebille, B. D. Hardesty, and J. A. Estes, "Threat of plastic pollution to seabirds is global, pervasive, and increasing," *Proc Natl Acad Sci U S A*, vol. 112, no. 38, pp. 11899–11904, Sep. 2015, doi: 10.1073/pnas.1502108112.
- [14] R. Qi, D. L. Jones, Z. Li, Q. Liu, and C. Yan, "Behavior of microplastics and plastic film residues in the soil environment: A critical review," *Science of The Total Environment*, vol. 703, p. 134722, Feb. 2020, doi: 10.1016/j.scitotenv.2019.134722.
- [15] L. Hou, D. Kumar, C. G. Yoo, I. Gitsov, and E. L. W. Majumder, "Conversion and removal strategies for microplastics in wastewater treatment plants and landfills," *Chemical Engineering Journal*, vol. 406, p. 126715, Feb. 2021, doi: 10.1016/j.cej.2020.126715.
- [16] S. Ügdüler, K. M. Van Geem, M. Roosen, E. I. P. Delbeke, and S. De Meester, "Challenges and opportunities of solvent-based additive extraction methods for plastic recycling," *Waste Management*, vol. 104, pp. 148–182, Mar. 2020, doi: 10.1016/j.wasman.2020.01.003.
- [17] B. Chen *et al.*, "Global distribution of Plastic Wastes Disposal: A Review," *J Hazard Mater*, vol. 451, p. 131198, 2023, doi: 10.1016/j.jhazmat.2023.131198.
- [18] A. O. A., O. K. I., A. Oluwaseun, and A. O. E., "Public and Environmental Health Effects of Plastic Wastes Disposal: A Review," *Journal of Toxicology and Risk Assessment*, vol. 5, no. 2, Nov. 2019, doi: 10.23937/2572-4061.1510021.
- [19] K. Blackburn and D. Green, "The potential effects of microplastics on human health: What is known and what is unknown," *Ambio*, vol. 51, no. 3, Springer Science and Business Media B.V., pp. 518–530, Nov. 2022, doi: 10.1007/s13280-021-01589-9.
- [20] M. M. R. Khan, Md. A. B. Siddique, R. B. Arif, and M. R. Oishe, "ADBSCAN: Adaptive Density-Based Spatial Clustering of Applications with Noise for Identifying Clusters with Varying Densities," in *2018 4th International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT)*, 2018, pp. 107–111, doi: 10.1109/ICEEICT.2018.8628138.
- [21] L. Huang *et al.*, "Microplastic contamination in coral reef fishes and its potential risks in the remote Xisha areas of the South China Sea," *Mar Pollut Bull*, vol. 186, p. 114399, Nov. 2023, doi:10.1016/j.marpolbul.2022.114399.
- [22] N. Vidanapathirana, Y. Wang, A. C. McLain, and S. Self, "Cluster Detection Capabilities of the Average Nearest Neighbor Ratio and Ripley's K Function on Areal Data: an Empirical Assessment." 2022.
- [23] Self S, Overby A, Zgodic A, White D, McLain A, Dyckman C. A Generalization of Ripley's K Function for the Detection of Spatial Clustering in Areal Data. arXiv preprint arXiv:2204.10852. 2022 Apr 22.
- [24] G. B. Vincent, A. P. Proudian, and J. D. Zimmerman, "Three dimensional cluster analysis for atom probe tomography using Ripley's K-function and machine learning," *Ultramicroscopy*, vol. 220, p. 113151, Jan. 2021, doi: 10.1016/j.ultramic.2020.113151.
- [25] F.-Y. Wen Tzai-Hung and Kuo, "Point Pattern Analysis for Identifying Spatial Clustering Tendency," in *Methodological Approaches in Integrated Geography*, F. B. Mustafa, Ed., Cham: Springer International Publishing, 2023, pp. 45–59, doi: 10.1007/978-3-031-28784-8_4.
- [26] K. Koczczyńska, *Applied Spatial Statistics and Econometrics*. Routledge, 2020, doi: 10.4324/9781003033219.
- [27] D. Ebdon, *Statistics in Geography*. Blackwell Publishing, 1985.
- [28] J. Han, M. Kamber, and J. Pei, *Data mining: Data Mining Concepts and Techniques*. Morgan Kaufman, 2014.
- [29] M. N. Aidi, *Konfigurasi Titik dalam Ruang*. Bogor: Institut Pertanian Bogor, 2013. [Online]. Available: https://repository.ipb.ac.id/jspui/bitstream/123456789/68973/1/Gabungan_konfigurasi_titik_ruang.pdf
- [30] H. Schabenberger, "Spatial count regression Repository." John Wiley & Sons, 2009.
- [31] J. Alexandra and K. P. Sinaga, "Machine Learning Approaches for Marketing Campaign in Portuguese Banks," in *2021 3rd International Conference on Cybernetics and Intelligent System (ICORIS)*, 2021, pp. 1–6, doi: 10.1109/icoris52787.2021.9649623.
- [32] A. Fauzan, A. Novianti, R. R. M. A. Ramadhani, and M. A. S. Adhiwibawa, "Analysis of Hotels Spatial Clustering in Bali: Density-Based Spatial Clustering of Application Noise (DBSCAN) Algorithm Approach," *EKSAKTA: Journal of Sciences and Data Analysis*, pp. 25–38, Mar. 2022, doi: 10.20885/eksakta.vol3.iss1.art4.
- [33] I. M. S. Putra, "Algoritma DBSCAN (Density-Based Spatial Clustering of Application with Noise) dan contoh perhitungannya," Bali, 2018.
- [34] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," in *Proceeding on the 2nd International Conference on Knowledge Discovery*, 1996, doi: 10.1016/B978-044452701-1.00067-3.
- [35] J. Han, M. Kamber, and J. Pei, *Data mining: Data Mining Concepts and Techniques*. USA: Morgan Kaufman, 2014.
- [36] K. R. Shahapure and C. Nicholas, "Cluster Quality Analysis Using Silhouette Score," in *2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)*, 2020, pp. 747–748, doi:10.1109/DSAA49011.2020.00096.
- [37] 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.
- [38] F. Batool and C. Hennig, "Clustering with the Average Silhouette Width," *Comput Stat Data Anal*, vol. 158, p. 107190, Jun. 2021, doi:10.1016/J.CSDA.2021.107190.
- [39] P. J. Rousseeuw, "Silhouettes: A graphical Aid to the Interpretation and Validation of Cluster Analysis," *J Comput Appl Math*, vol. 20, no. C, pp. 53–65, 1987.
- [40] P. Holloway, *Understanding GIS through Sustainable Development Goals Case Studies with QGIS*. CRC Press, 2023.
- [41] M. Law and A. Collins, *Getting to Know ArcGIS Pro 2.8*. Esri Press, 2021.
- [42] D. Hermon, *Geografi Bencana Alam - Rajawali Pers*. Jakarta: PT Raja Grafindo Persada, 2015.