Terbit *online* pada laman : https://ejournal.iainbukittinggi.ac.id/index.php/ijokid/

**Knowbase : International Journal of Knowledge in Database**

| ISSN (Print) 2798-0758 | ISSN (Online) 2797-7501 |

# The Mapping of Waste Management Facilities in Bogor Regency Using a K-Means

*Nadhira Faza Fatiha[1,] , Irmayansyah[2*]*

*[1, 2] Binaniaga Indonesia University, Bogor, Indonesia*

**Article Information**

Submitted: November 7, 2025
Revision: December 24, 2025
Accepted: December 25, 2025
Published: December 31, 2025

**Keywords**

Mapping
Waste Management Facility
K-Means

**Correspondence**

E-mail: Irma@unbin.ac.id *

**A B S T R A C T**

Waste is the solid residue of human activities or natural processes that is considered useless. If not managed properly, it can have a negative impact on the environment and health. When waste management facilities such as transport fleets, waste banks, and TPS3R (3R waste management sites) are insufficient to handle the volume of waste effectively, waste accumulation will occur, which can pollute the air, soil, and water, and increase the risk of disease spread. Therefore, new data-driven thinking is needed to improve more targeted and efficient waste management. The application of the K-Means *clustering* technique in waste management can be done, as demonstrated by the results of regional clustering in Bogor Regency with recommendations for appropriate facilities. The use of variables such as uncollected waste *volume*, distance to landfill sites, number of villages and population, as well as the stages of determining the number of clusters, initial *centroid* point determination, calculation of data distance to *centroid* points, and grouping of data according to minimum distance to *centroid* points were carried out to produce the clustering. The variables and stages were then applied to a prototype decision support system to assist the Bogor District Environmental Agency in placing waste management facilities more effectively. The prototype system developed has undergone feasibility testing by users using the *Post-Study System Usability Questionnaire* (PSSUQ) instrument and produced a feasibility level of 86.5% and feasibility testing by experts using the ISO 9126 instrument and produced a feasibility level of 100%. A *cluster* validity test was also conducted using *the Silhouette Coefficient* on the K-means algorithm applied, with a value of 0.354.

This is an open access article under the CC–BY-SA license

## 1. *Introduction*

Waste management is a global issue that affects various aspects of life, from the environment and public health to the economy. In Southeast Asia, waste management poses a major challenge due to limited infrastructure, low public awareness, and weak regulations. Improperly managed waste often ends up in the sea, polluting ecosystems and threatening biodiversity. In addition, methane emissions from landfills contribute to the increase in greenhouse gases that accelerate climate change.

Several countries, such as Japan and South Korea, have successfully overcome waste problems through the implementation of strict policies and active public participation. Japan implements a strict waste sorting system at the household level, thereby reducing the volume of waste sent to landfills. On the other hand, South Korea implements a *Pay as You Throw* policy that encourages the public to reduce waste production. The success of these two countries shows that a combination of regulations, technology, and public awareness can create an efficient and sustainable waste management system.

Indonesia faces major challenges in waste management, particularly as it is one of the world's largest contributors to plastic waste. Based on data from SIPSN KLHK in 2023, as of July 2024, the total national waste accumulation reached 31.9 million tons, with 35.7% of it not being managed properly[1] . The increasing population and limited disposal facilities are the main obstacles in the implementation of waste reduction programs. In addition, low public participation worsens the effectiveness of policies that have been implemented by the government.

In Bogor Regency in particular, the population of 5.56 million people has exacerbated environmental problems, including waste[2] . Population density has led to an increase in waste volume, while management facilities such as temporary storage sites, fleets, and waste banks are still limited. Disparities between subdistricts in terms of infrastructure and access to management have caused some areas to experience unmanaged waste accumulation, as shown in Table 1. This condition demonstrates the need for a data- and region-based approach so that waste management strategies become more efficient, adaptive, and sustainable.

Data mining offers a solution to this waste management problem with one relevant technique: the K- Means clustering algorithm, which groups data into several *clusters* based on similar characteristics. The use of K-Means in the context of waste management is an innovation that utilizes technology to design a more efficient system and provide targeted service recommendations. This data-driven approach helps to understand waste issues in a more structured and evidence-based manner, rather than relying solely on conventional methods. Thus, the K-Means approach has the potential to support more accurate decision-making, improve environmental quality, and strengthen sustainable waste management efforts.

Several previous studies have applied K-Means in the field of waste management. Research conducted by Gusta Nugraha, M.Maimunah, and P. Sukmasetya[3] used the K-Means algorithm to map areas based on waste generation levels. The clustering results divided the area into three groups: areas with low waste levels (200.8 kg), medium levels (498 kg), and high levels (835 kg). The clustering evaluation results, using *a silhouette score* of 0.79, showed good clustering quality. Meanwhile, research conducted by Prasetyo, M. Maimunah, and P. Sukmasetya [4] discusses the efficiency of waste transportation routes to the Banyuurip landfill in Magelang City. The study utilizes the K-Means algorithm to divide the area based on subdistricts (North, Central, and South) with clustering evaluation results with *silhouette scores* ranging from 0.63 to 0.64. However, both studies have limitations in linking the clustering results with concrete steps that can be taken to improve waste management services. This is an important gap that needs to be addressed, especially in the context of strategic decision-making for effective resource allocation.

**Table 1.** Waste Management Data in Bogor Regency in 2024

| No | Subdistrict Name | Volume of Uncollected Waste (Tons/Month) | Distance to Landfill (KM) | Number of Villages | Number of Residents | Service Coverage Percentage |
|---|---|---|---|---|---|---|
| 1 | Cibinong Subdistrict | 1,860 | 33 | 13 | 371,380 | 77 |
| 2 | Babakan Madang Subdistrict | 961 | 36 | 9 | 124,745 | 63 |
| 3 | Sukaraja Subdistrict | 2,604 | 30 | 13 | 205,421 | 18 |
| … | … | … | … | … | … | … |
| 38 | Rumpin Subdistrict | 2,015 | 16 | 14 | 159,802 | 12 |
| 39 | Cigudeg Subdistrict | 1,829 | 39 | 15 | 145,014 | 14 |
| 40 | Sukajaya Subdistrict | 961 | 23 | 11 | 73,405 | 9 |

Based on the data in Table 1, it can be seen that there are still significant gaps in waste management, including differences in waste transportation levels, limited infrastructure facilities, and uneven allocation of infrastructure facilities in each region, which are the main problems in waste management in Bogor Regency. Some areas experience very poor service, where the transportation level is far less than the daily amount of waste accumulated. As a result, the disparity in the level of waste management services is becoming increasingly apparent.

This study is designed to fill the existing gap by combining the results of regional clustering with recommendations for appropriate facilities. Using more diverse data, this study will produce clusters for the main needs of service improvement. These cluster-based recommendations not only improve operational efficiency but also support sustainable long-term planning . This integrated approach provides added value not found in previous studies, making it a useful reference for policymakers in waste management across various regions.

## 2. *Method*

### 2.1 *Research Procedure*

Research procedures are a series of steps from the initial to the final stages of research to answer research questions. This procedure includes the preparation stage (covering observation, data collection, and data understanding), the implementation stage (covering data selection and data normalization, modeling, prototype design), the evaluation stage (covering product testing, result testing, and revision), and the final stage (reporting the final prototype product results)[5].The research procedure carried out can be seen in Figure 1.

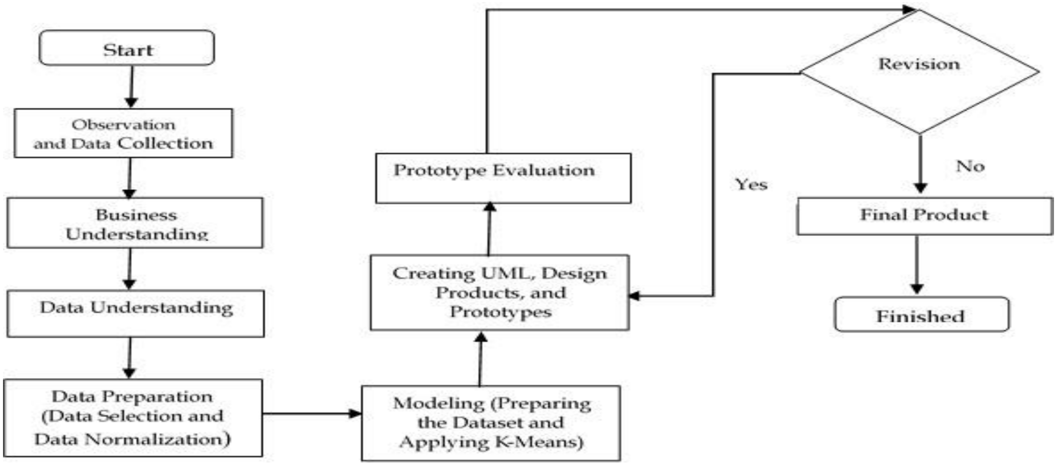


**Figure1 .** Research Procedure

### 2.2 *Data Mining*

This research uses data mining techniques, which is a process for discovering patterns, trends, and valuable information from large and complex data sets[6] . One of the standard data handling processes in data mining is *the Cross-Industry Standard Process for Data Mining* (CRISP-DM) with the following data management stages[7-9] .

### 2.3 *K-Means*

One *data mining* technique that can be used for mapping waste management facilities by region in Bogor Regency is *clustering*. *Clustering* is a technique in data mining that is used to group similar data into groups or *clusters* based on similarities in certain attributes[10 -11] . The main purpose of this technique is to group and divide data patterns into several data sets so that they form similar patterns and are grouped into the same cluster and separated by forming different patterns into different *clusters*[12 -13] . One of the *clustering* algorithms is K-Means.

K-means is a group analysis method that partitions N observation objects into K groups (*clusters*) where each observation object belongs to a group with the closest mean[14 -15] . The K-Means algorithm works as follows[16 -17] :

1. determine the value of k;
2. selecting *cluster* centers (*centroids*) randomly;
3. measure the distance of each data point using the *Euclidean* distance;

$$d_{ij} = \sqrt{\sum_{j=1}^{m}(x_{kj} - c_{ij})^2} = \sqrt{(x_k - c_i)^2} + \sqrt{(x_j - c_j)^2} \quad (1)$$

4. assign each point to the nearest cluster;

$$\min \sum_{k=1}^{k} d_{ik} = \sqrt{\sum_{j=1}^{m}(x_{kj} - c_{ij})^2} \quad (2)$$

5. Calculate the average of each *cluster* as the new centroid using the following formula;

$$C_{ij} = \frac{\sum_{k=1}^{p} x_{kj}}{p} \quad (3)$$

$x_{kj}$ ∈ cluster ke- k

p : number of memberscluster ke- k

m : data dimension

6. Repeat steps 3-5 with the new *cluster* centers, repeat until convergence (no further changes).

### 2.4 *Silhouette Coefficient*

The Silhouette Coefficient is a metric that indicates the extent to which data in a group is similar, and is calculated individually for each object in the group. The closer the Silhouette Coefficient is to 1, the higher the quality of clustering in that group. Conversely, the closer the Silhouette Coefficient is to -1, the poorer the clustering quality of that cluster. It plays a role in determining the quality and strength of the clusters formed[17]. The following are the steps for testing the results using the Silhouette Coefficient[19 - 20]:

1. For each object I, calculate the average distance from object I to all objects within the same *cluster,* denoted as a;
2. For each object I, calculate the average distance from object I to objects in other *clusters*. From all these average distances, take the smallest value, which is called bj (small square);
3. Then, the *Silhouette Coefficient* value for object i is;

$$S_{(i)} = \frac{b_i - a_i}{max\{a_i, b_i\}} \quad (4)$$

$a_i$ : the average distance between members within *a cluster*

$b_i$ : the minimum value of the average distance between object i and objects in other *clusters*

### 2.5 *Prototype*

*Prototype* is a method in software development where *a prototype* or initial model of the system is created to provide a visual representation of how the proposed system will work[21-22] . The main purpose of *a prototype* is to better understand user needs and to ensure that the proposed solution meets their expectations and requirements. *Prototypes* are used to bridge users lack of understanding of technical matters and to clarify the specifications of user requirements to software developers[22] .

The *prototype* method begins with a process of listening to the needs and input of users. Developers and users meet to determine the overall objectives and identify the various software requirements needed. The process then continues with the developers creating an overview of the application, which is then presented to the customer. The overview focuses on representing the aspects of the application that will be visible to the customer/user[24-25] .

### 2.6 ISO 9126 Quality Model

One *prototype* testing technique that can be used to map the conditions of waste management facilities in Bogor Regency is ISO 9126. ISO 9126 testing is used to define the quality of software products, models, quality characteristics, and related metrics used to evaluate and determine the quality of a *software* product[25] . According to ISO 9126, quality factors include the following six quality characteristics[26-28] :

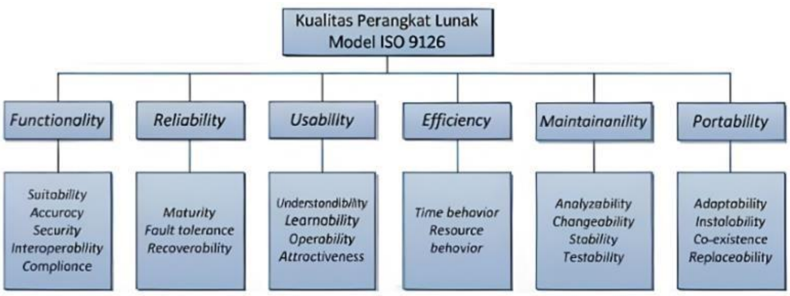


**Figure 2.** ISO-9126 Testing [25]

### 2.7 *Usability* Measurement Using PSSUQ (*Post Study System Questionnaire*)

*Prototype* testing with users can use PSSUQ. PSSUQ is a measurement used to observe user interaction with the system and measure the level of user satisfaction with the system. All questions in PSSUQ are written in a positive manner[27] . PSSUQ has 16 questionnaire items. The usability aspects assessed in the questionnaire include[29-30] :

a. *System Usefulness*, a measure of user satisfaction with the system's ability to function well in terms of functionality;
b. *Information Quality*, a measure of user satisfaction with the system's ability to help users complete tasks by reading information;
c. *Interface Quality*, a measure of user satisfaction with the system's ability to work well in providing an easy-to-use interface for users;
d. *The Overall Satisfaction Score*, a measure *of* user satisfaction assessed from the overall experience.

## 3. *Results and Discussion*

This research stage was adjusted based on the data mining concept with the CRISP-DM stages as follows:

### 3.1 *Business Understanding*

In the previous system, the mapping of waste management facilities per region was not accurate and the process of mapping waste management facilities was not effective. Given these conditions, this study aims to map waste management facilities per region in Bogor Regency through a decision-making system using the K-Means approach.

### 3.2 *Data Understanding*

Through interviews with sources, namely the Head of Waste Management, the documents were successfully collected. Data on waste management in 2024 was obtained from the Bogor Regency Environment Agency, while data on the population based on the 2024 regional administration was obtained from the Bogor Regency Central Statistics Agency. At this stage, an initial check of the data was carried out. Data selection was based on its relevance to the actual conditions of waste management in each region. Of the 40 data points, an imbalance in the variables was noted. This imbalance is an important consideration because it can affect the performance of the K-Means algorithm, particularly in the grouping of regions.

### 3.3 *Data Preparation*

In the data preparation stage, 40 waste management data points from 2024 were used, as shown in Table 1. The waste management data was selected to ensure that the data analyzed was valid. The variables used consisted of the volume of unserved waste, distance to the landfill, number of villages, and population. The data will be normalized using *Decimal Scaling* to ensure a balanced scale, making the calculation process more accurate and fair.

$$new\ data = \frac{data}{i} \quad (5)$$

New data = normalized
data = data used
i = maximum value of the criteria

**Table 2.** Waste Management Data Normalization

| No | Volume of Uncollected Waste (Tons/Month) | Distance to Landfill (KM) | Number of Villages | Population |
|---|---|---|---|---|
| 1 | 0.50 | 0.36 | 0.81 | 1,000 |
| 2 | 0.26 | 0.39 | 0.56 | 0.336 |
| 3 | 0.70 | 0.33 | 0.81 | 0.553 |
| … | … | … | … | … |
| 38 | 0.54 | 0.17 | 0.88 | 0.430 |
| 39 | 0.49 | 0.42 | 0.94 | 0.390 |
| 40 | 0.26 | 0.25 | 0.69 | 0.198 |

### 3.4 *Modeling*

At this stage, the K-Means algorithm was applied. The data used in this calculation included variables such as uncollected waste volume, distance to the landfill, number of villages, and population based on 2024 waste management data obtained from the Bogor Regency Environment Agency, as well as 2024 population and regional administration data obtained from the Bogor Regency Central Statistics Agency. The waste management dataset is as follows:

**Table 3.** Waste Management Dataset

| No | Subdistrict Name | Volume of Uncollected Waste (Tons/Month) | Distance to Landfill (KM) | Number of Villages | Population |
|---|---|---|---|---|---|
| 1 | Cibinong Subdistrict | 0.5 | 0.36 | 0.81 | 1,000 |
| 2 | Babakan Madang Subdistrict | 0.26 | 0.39 | 0.56 | 0.336 |
| 3 | Sukaraja Subdistrict | 0.70 | 0.33 | 0.81 | 0.553 |
| … | … | … | … | … | … |
| 38 | Rumpin Subdistrict | 0.54 | 0.17 | 0.88 | 0.430 |
| 39 | Cigudeg Subdistrict | 0.49 | 0.42 | 0.94 | 0.390 |
| 40 | Sukajaya District | 0.26 | 0.25 | 0.69 | 0.198 |

From the dataset in Table 3, the number of clusters was determined to be 3 (three). The determination of the number of clusters was based on the needs of the research conducted with the aim of final labeling according to requirements. The names of the labels or clusters are cluster 1 TPS3R, cluster 2 Bank Sampah, and cluster 3 Armada. After determining the clusters, the centroid values of each cluster were determined as follows:

$$\frac{n\ Data}{n\ Cluster+1} \quad (6)$$

The centroid points for Cluster 1, Cluster 2, and Cluster 3 are as follows:

C1 = Total data / C1+1 = 40 / 1+1 = 20 (20th data point)
C2 = Total data / C2+1 = 40 / 2+1 = 13 (13th data point) C3 = Number of data / C3+1 = 40 / 3+1 = 10 (10th data)

The calculation of the distance between data and *the centroid* is done to determine the shortest distance that will be used to determine cluster grouping. *Euclidean Distance* is used to calculate the distance between data and *the centroid*. *Euclidean Distance* is the calculation of the distance between two points and is used to study the relationship between angles and distances. The *Euclidean Distance* formula is as follows:

$$d_{ij} = \sqrt{\sum_{j=1}^{m}(x_{kj} - c_{ij})^2} = \sqrt{(x_k - c_i)^2} + \sqrt{(x_j - c_j)^2} \quad (7)$$

$Dij$ = *Euclidean Distance*
$(kj, ij)$ = Number of Objects
$x$ = Object Coordinates
$c$ = *Centroid* Coordinates

Calculation of data distance to *the centroid* using *Euclidean Distance* for Iteration 1:

**Table 4.** Iteration 1 Data

| No | C1 | C2 | C3 | Closest Distance | Cluster |
|---|---|---|---|---|---|
| 1 | 0.599 | 1.106 | 0.334 | 0.334 | C3 |
| 2 | 0.550 | 0.639 | 0.577 | 0.550 | C1 |
| 3 | 0.264 | 0.938 | 0.348 | 0.264 | C1 |
| … | … | … | … | … | … |

| | | | | | |
|---|---|---|---|---|---|
| 38 | 0.142 | 0.967 | 0.531 | 0.142 | C1 |
| 39 | 0.388 | 0.752 | 0.443 | 0.388 | C1 |
| 40 | 0.499 | 0.755 | 0.711 | 0.499 | C1 |

After the data grouping process, new *centroid* points were determined using the formula:

$$\mu_k = \frac{1}{N_k}\sum_{t=1}^{N_k} x_i \quad (8)$$

$\mu_k$ : *Centroid* point of *cluster* K
$N_k$ : Number of data points in *cluster* K
$x_i$ : Data point i in *cluster* K

Repeat the calculation of the distance between data points and *the* new *centroid* until the data clustering or *centroid* values no longer change. The calculation stops at iteration 4 because the new *centroid* values have not changed or are the same as *the centroid* in the previous iteration. The following are the results of the distance calculation with *the centroid* in Iteration 4:

**Table 5.** Iteration 4 Data

| No | C1 | C2 | C3 | Closest Distance | Cluster |
|---|---|---|---|---|---|
| 1 | 0.701 | 0.844 | 0.354 | 0.354 | C3 |
| 2 | 0.215 | 0.281 | 0.647 | 0.215 | C1 |
| 3 | 0.425 | 0.645 | 0.216 | 0.216 | C3 |
| ... | ... | ... | ... | ... | ... |
| 38 | 0.291 | 0.631 | 0.454 | 0.291 | C1 |
| 39 | 0.338 | 0.464 | 0.472 | 0.338 | C1 |
| 40 | 0.174 | 0.409 | 0.737 | 0.174 | C1 |

After the calculations were performed, the results of the data grouping into clusters were obtained. These results show how *K-Means* clusters waste management facilities per region in Bogor Regency based on the variables of uncollected waste volume, distance to the landfill, number of villages, and population. The following is a table of data grouping per cluster

**Table 6. Cluster 1 TPS3R Data**

| No | Subdistrict Name | Volume of Unserviced Waste (Tons/Month) | Distance to Landfill (KM) | Number of Villages | Population |
|---|---|---|---|---|---|
| 1 | Babakan Madang Subdistrict | 961 | 36 | 9 | 124,745 |
| 2 | Tajur Halang Subdistrict | 1,302 | 25 | 7 | 130,958 |
| 3 | Ciawi Subdistrict | 1,333 | 32 | 13 | 119,691 |
| 4 | Megamendung Subdistrict | 992 | 38 | 12 | 115,955 |
| 5 | Caringin Subdistrict | 1,860 | 40 | 12 | 141,326 |
| 6 | Cijeruk Subdistrict | 1,116 | 33 | 9 | 97,895 |
| 7 | Ciampea District | 2,308 | 8 | 13 | 177,866 |
| 8 | Ciomas Subdistrict | 1,612 | 19 | 11 | 163,471 |
| 9 | Dramaga Subdistrict | 1,457 | 15 | 10 | 115,108 |
| 10 | Tamansari Subdistrict | 899 | 25 | 8 | 114,569 |
| 11 | Tenjolaya Subdistrict | 899 | 18 | 7 | 69,519 |
| 12 | Parung Subdistrict | 1,488 | 30 | 9 | 131,775 |
| 13 | Kemang Subdistrict | 1,178 | 22 | 9 | 134,193 |
| 14 | Rancabungur Subdistrict | 496 | 14 | 7 | 64,949 |
| 15 | Ciseeng Subdistrict | 1,271 | 26 | 10 | 123,719 |
| 16 | Gunung Sindur Subdistrict | 1,581 | 33 | 10 | 124,610 |
| 17 | Leuwiliang District | 1,488 | 2 | 11 | 135,298 |
| 18 | Leuwisadeng Subdistrict | 930 | 4 | 8 | 84,186 |
| 19 | Cibungbulang Subdistrict | 1,767 | 3 | 15 | 155,190 |

| | | | | | |
|---|---|---|---|---|---|
| 20 | Pamijahan Subdistrict | 2,108 | 11 | 15 | 169,046 |
| .... | ...................................... | .................. | .......... | .......... | .......... |
| .... | ...................................... | .................. | .......... | .......... | .......... |
| 25 | Cigudeg Subdistrict | 1,829 | 39 | 15 | 145,014 |
| 26 | Sukajaya District | 961 | 23 | 11 | 73,405 |

**Table 7. Data on Cluster 2 of the Waste Bank**

| No | Subdistrict Name | Volume of Uncollected Waste (Tons/Month) | Distance to Landfill (KM) | Number of Villages | Population |
|---|---|---|---|---|---|
| 1 | Jonggol Subdistrict | 279 | 66 | 14 | 151,086 |
| 2 | Klapanunggal Subdistrict | 1,612 | 50 | 9 | 112,424 |
| 3 | Cariu Subdistrict | 651 | 78 | 10 | 55,031 |
| 4 | Sukamakmur Subdistrict | 1,178 | 56 | 10 | 93,780 |
| 5 | Tanjungsari Subdistrict | 744 | 92 | 10 | 62,286 |
| 6 | Cisarua District | 992 | 44 | 10 | 134,234 |
| 7 | Cigombong Subdistrict | 1,147 | 42 | 9 | 102,979 |
| 8 | Tenjo Subdistrict | 775 | 49 | 9 | 79,795 |

**Table 8. Data on Fleet Cluster 3**

| No | Subdistrict Name | Volume of Uncollected Waste (Tons/Month) | Distance to Landfill (KM) | Number of Villages | Population |
|---|---|---|---|---|---|
| 1 | Cibinong Subdistrict | 1,860 | 33 | 13 | 371,380 |
| 2 | Sukaraja Subdistrict | 2,604 | 30 | 13 | 205,421 |
| 3 | Bojong Gede Subdistrict | 3,565 | 27 | 9 | 275,239 |
| 4 | Citereup Subdistrict | 2,511 | 37 | 14 | 224,692 |
| 5 | Gunung Putri Subdistrict | 3,720 | 43 | 10 | 299,046 |
| 6 | Cileungsi Subdistrict | 2,077 | 52 | 12 | 280,084 |

### 3.5 *Evaluation*

Next, an evaluation of the K-Means modeling was conducted to see the extent to which the algorithm was able to cluster waste management data. This evaluation became the basis for determining whether K-Means was good enough and suitable for use in clustering waste management data, particularly as a decision support system for recommending the allocation of waste management facilities that were more targeted in each region.

The evaluation was carried out using *the Silhouette Coefficient* to test the quality of each *cluster* produced by combining the *cohesion* and *separation* methods. The following are the steps for testing the evaluation results with *the silhouette coefficient*:

1. Calculate the average distance of a data point from all other data points within the same cluster, for example, data point 1 from all data points in that cluster, by calculating a(i) (*average dissimilarity*) using the following formula:

$$a(i) = |A| - 1 \sum j \in A, j \neq 1 \, d(i,j) \quad (9)$$

$a(i)$ : The average distance of a data point from all other data points within a cluster A : The number of data points in cluster A
$d(i,j)$ : The distance between data point i and j

a. Data point 1 relative to all data points in cluster 1:

$$a(i)_1 = \sqrt{\frac{\begin{array}{c}(0{,}26 - 0{,}26)^2 + (0{,}39 - 0{,}39)^2 + (0{,}56 - 0{,}56)^2 \\ + (0{,}336 - 0{,}336)^2 + (0{,}26 - 0{,}35)^2 + (0{,}39 - 0{,}27)^2 + \\ (0{,}56 - 0{,}44)^2 + (0{,}336 - 0{,}353)^2 + \ldots + \\ (0{,}26 - 0{,}26)^2 + (0{,}39 - 0{,}25)^2 + (0{,}56 - 0{,}69)^2 \\ + (0{,}336 - 0{,}198)^2\end{array}}{}}$$

$$a(i)_1 = \sqrt{\frac{7{,}597393994}{26}} = 0{,}292$$

b. Data point 1 relative to all data points in cluster 2:

$$a(i)_2 = \sqrt{\begin{array}{c}(0{,}08 - 0{,}08)^2 + (0{,}72 - 0{,}72)^2 + (0{,}88 - 0{,}88)^2 \\ + (0{,}407 - 0{,}407)^2 + (0{,}08 - 0{,}43)^2 + (0{,}72 - 0{,}54)^2 + \\ (0{,}88 - 0{,}56)^2 + (0{,}407 - 0{,}303)^2 + \ldots + \\ (0{,}08 - 0{,}21)^2 + (0{,}72 - 0{,}53)^2 + (0{,}88 - 0{,}56)^2 \\ + (0{,}407 - 0{,}215)^2\end{array}}$$

$$a(i)_2 = \sqrt{\frac{3{,}095191915}{8}} = 0.387$$

c. Data point 1 relative to all data points in cluster 3:

$$a(i)_3 = \sqrt{\begin{array}{c}(0{,}50 - 0{,}50)^2 + (0{,}36 - 0{,}36)^2 + (0{,}81 - 0{,}81)^2 \\ + (0{,}407 - 0{,}407)^2 + (0{,}50 - 0{,}70)^2 + (0{,}36 - 0{,}33)^2 + \\ (0{,}81 - 0{,}553)^2 + (0{,}407 - 0{,}333)^2 + \ldots + \\ (0{,}50 - 0{,}56)^2 + (0{,}36 - 0{,}57)^2 + (0{,}81 - 0{,}75)^2 \\ + (0{,}407 - 0{,}754)^2\end{array}}$$

$$a(i)_3 = \sqrt{\frac{2{,}431449446}{6}} = 0{,}405$$

The results of the a(i) calculation for one cluster are shown in Table 9

**Table 9. Calculation Results of a(i)**

| Data to | a(i) |
|---|---|
| 1 | 0.405 |
| 2 | 0.292 |
| 3 | 0.302 |
| ……………. . | ………. |
| 39 | 0.392 |
| 40 | 0.282 |

2. Calculate the average distance of a data point from all other data points in other clusters, for example, data point 1 from all data points in other clusters, by calculating d(i,C) using the formula

$$d(i,C) = \frac{1}{|C|\sum} j \in C\, d(i,j) \tag{10}$$

Explanation:
$d(i,C)$: Average distance of data point i from all data points in different clusters
$C$ : Number of data points in cluster C
$d(i,j)$ : Distance between data i and j

a. Cluster 1 to Cluster 2:

$$d(1,2) = \sqrt{\begin{array}{c}(0,08-0,26)^2+(0,72-0,39)^2+(0,88-0,56)^2 \\ (0,407-0,336)^2+\ (0,43-0,26)^2+(0,54-0,39)^2+ \\ (0,56-0,56)^2+(0,303-0,336)^2+\ \ldots + \\ (0,21-0,26)^2+(0,53-0,39)^2+(0,56-0,56)^2 \\ +(0,215-0,336)^2\end{array}}$$

$$d(1,2) = \sqrt{\frac{2,541}{8}} = 0,318$$

b. Cluster 1 ke Cluster 3:

$$d(1,3) = \sqrt{\begin{array}{c}(0,5-0,26)^2+(0,36-0,39)^2+(0,81-0,56)^2 \\ (1-0,336)^2+\ (0,7-0,26)^2+(0,33-0,39)^2+ \\ (0,81-0,56)^2+(0,553-0,336)^2+\ \ldots + \\ (0,56-0,26)^2+(0,57-0,39)^2+(0,75-0,56)^2 \\ +(0,754-0,336)^2\end{array}}$$

$$d(1{,}3) = \sqrt{\frac{4{,}171}{6}} = 0{,}695$$

The results of calculating the average distance from data i to all data in other clusters can be seen in Table 10

**Table 10 . Calculation Results d(i,C)**

| Data to | d(i,C) | | |
|---|---|---|---|
| | d(i,1) | d(i,2) | d(i,3) |
| 1 | 0.736 | 0.870 | - |
| 2 | - | 0.318 | 0.695 |
| 3 | 0.474 | 0.673 | - |
| …….. | ………… | ……….. | ……….. |
| 39 | - | 0.512 | 0.519 |
| 40 | - | 0.445 | 0.777 |

3. After calculating the$d\ (i\ ,C\ )$ for all clusters C ≠ A, take the smallest or minimum value as the$b\ (i\ )$, using the following formula:

$$b(i) = \frac{min}{C\#A}\ d(i,j) \tag{11}$$

**Table 11. Calculation Results for b(i)**

| Data to | b(i) |
|---|---|
| 1 | 0.736 |
| 2 | 0.318 |
| 3 | 0.474 |
| ……… | ……… |
| 39 | 0.512 |
| 40 | 0.445 |

4. After obtaining the values of a(i) and b(i), the *silhouette coefficient* value is then calculated for each data point in its cluster, using the following formula:

$$S_{(i)} = \frac{b_i - a_i}{max\{a_i, b_i\}} \tag{12}$$ Explanation:

$a_i$ : average distance between members within *a cluster*
$b_i$ : minimum value of the average distance between object i and objects in other *clusters*

Data ke-1

$$s(i) = \frac{(0{,}736-0{,}405)}{\max(0{,}405{,}0{,}736)} = 0{,}450$$

Data ke-2

$$s(i) = \frac{(0{,}318-0{,}292)}{\max(0{,}292{,}0{,}318)} = 0{,}080$$

Data ke-3

$$s(i) = \frac{(0{,}474-0{,}302)}{\max(0{,}302{,}0{,}474)} = 0{,}362$$

…

Data ke-40

$$s(i) = \frac{(0{,}445-0{,}282)}{\max(0{,}282{,}0{,}445)} = 0{,}367$$

From the *silhouette* calculation results above, *the silhouette coefficient* of all data based on data sequence can be seen in Table 12.

**Table 12.** ***Silhouette Coefficient*** **Calculation Results**

| Data | s(i) |
|---|---|

| | |
|---|---|
| 1 | 0.450 |
| 2 | 0.080 |
| 3 | 0.362 |
| ……… | ……….. |
| 39 | 0.235 |
| 40 | 0.367 |

The *silhouette coefficient* value is the sum of s(i) obtained by combining a(i) and b(i). Based on the *silhouette coefficient* table, it can be concluded that the closer the *silhouette* value is to 1, the better the data grouping into a *cluster*. Conversely, if the *silhouette* value is closer to 0, the less effective the data grouping into a *cluster*. The results of testing using *the silhouette coefficient* on three *clusters* obtained an average value of 0.354, which means it falls into the weak structure category. This factor could be caused by data with high variation in the variable of uncollected waste volume between regions, so that *the clusters* are not completely separated (the distance between clusters is too close).

### 3.6 *Deployment*

The final result of this research is a prototype decision support system for mapping waste management facilities per region in Bogor Regency, which can assist the Bogor Regency DLH in mapping the allocation of waste management facilities to be more targeted by forming clusters based on their level of importance. The following is a view of the prototype of the waste management facility mapping system per region in Bogor Regency that has been created. the prototype *file upload* page has a *browse files* button, which allows users to select the csv file to be uploaded. After selecting the CSV file, the file will be uploaded and saved to the database. The data is immediately normalized on *the file upload* page. Then, the user selects variables, determines the number *of clusters*, and enters the initial *centroid* value.


**Figure 3.** Prototype *File Upload* Page


**Figure 4.** Prototype *File Upload* Page


**Figure 5.** Prototype *File Upload* Page

the prototype of the calculation results page, users can view the clustering calculation results data by clicking the "Run Clustering" button.

## Nilai Centroid Akhir

| Cluster | Volume Sampah Tidak Terlayani | Jarak ke TPA | Jumlah Desa | Jumlah Penduduk |
|---|---|---|---|---|
| 1 | 0.3081 | 0.2471 | 0.6803 | 0.3332 |
| 2 | 0.2479 | 0.6481 | 0.6328 | 0.2664 |
| 3 | 0.7319 | 0.4322 | 0.7396 | 0.7431 |

© DLH Kabupaten Bogor 2025

**Figure 6.** Calculation Results Page Prototype

the prototype of the cluster result diagram page, users can see the cluster result diagram in the form of a scatter plot to see the data distribution results.

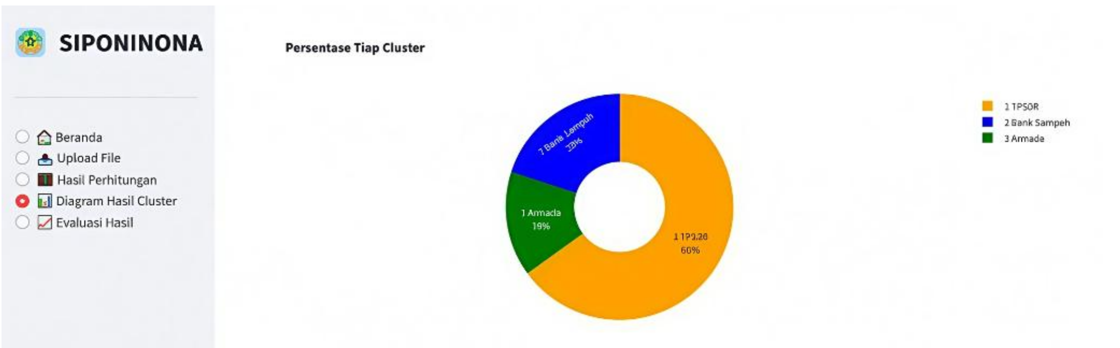


**Figure 7.** Cluster Result Diagram Page Prototype

the prototype of the results evaluation page, users can see the Silhouette Score.

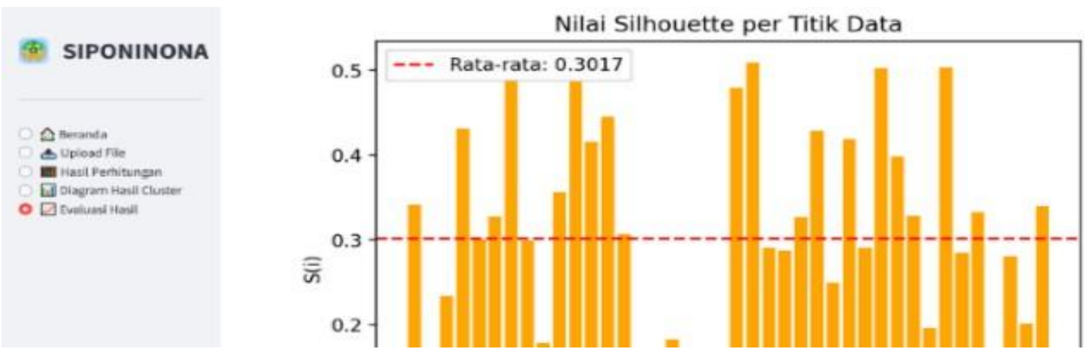


**Figure 8.** Prototype of the Results Evaluation Page

### 3.7 Quality Evaluation Results Based on ISO 9126

Testing of the system prototype by information system experts was conducted using a questionnaire instrument based on the indicators contained in ISO-9126. The prototype was tested starting from the input stage and output checking. The questionnaire results obtained were assessments by system experts of the quality of the Facility Management Mapping system. The questions used in the questionnaire for system experts included functionality, reliability, *usability*, *efficiency*, *maintainability*, and portability of *the prototype system*. The assessment in the expert system test questionnaire used the Guttman Scale, where "Yes = 1" and "No = 0". The following is the formula for quality evaluation based on ISO-9126:

$$Feasibility\ Percentage(\%) = \frac{the\ observed\ score}{expected\ score}\ x\ 100 \quad (13)$$

**Table 2 .** ISO-9126 Test Results

| Respondents | ISO 9126 Characteristics | | | | | |
|---|---|---|---|---|---|---|
| | Functionality | Reliability | Usability | Efficiency | Maintainability | Portability |
| R1 | 4 | 3 | 4 | 2 | 4 | 4 |
| R2 | 4 | 3 | 4 | 2 | 4 | 4 |
| R3 | 4 | 3 | 4 | 2 | 4 | 4 |

(a) Functionality

$$Feasibility\ Percentage\ (\%) = \frac{4+4+4}{4x3}\ x\ 100\% = 100\%$$

(b) Reliability

$$Feasibility\ Percentage\ (\%) = \frac{3+3+3}{3x3}\ x\ 100\% = 100\%$$

(c) Usability

$$Feasibility\ Percentage\ (\%) = \frac{4+4+4}{4x3}\ x\ 100\% = 100\%$$

(d) Efficiency

$$Feasibility\ Percentage\ (\%) = \frac{2+2+2}{2x3}\ x\ 100\% = 100\%$$

(e) Maintainbility

$$Feasibility\ Percentage(\%) = \frac{4+4+4}{4x3}\ x\ 100\% = 100\%$$

(f) Portability

$$Feasibility\ Percentage\ (\%) = \frac{4+4+4}{4x3}\ x\ 100 = 100\%$$

Based on the overall feasibility percentage results, the assessment results obtained from all respondents based on the feasibility percentage are declared "Highly Feasible"

### *3.8 Usability Evaluation Results Using PSSUQ*

Prototype system testing with users was conducted using a questionnaire instrument based on the indicators contained in the PSSUQ. The prototype was tested starting from the input process stage and output result checking. The questionnaire results obtained were an assessment of user satisfaction with the waste management facility mapping system in Bogor Regency. The questions used in the questionnaire included System Functionality, *Information Quality, Interface* Quality*, and* overall user satisfaction with the prototype system. The assessment in the user test questionnaire used a Likert scale consisting of seven response levels, ranging from "Strongly Disagree = 1" to "Strongly Agree = 7". The following is the formula for PSSUQ testing:

$$Feasibility\ Percentage(\%) = \frac{the\ observed\ score}{expected\ score}\ x\ 100 \quad (14)$$

**Table 3.** PSSUQ Test Results

| Respondents | Category Name | | | |
|---|---|---|---|---|
| | Overall | SysUse | InfoQual | IntQual |
| R1 | 96 | 36 | 36 | 24 |
| R2 | 93 | 35 | 34 | 24 |
| R3 | 101 | 37 | 38 | 26 |

(a) Keseluruhan (*Overall*)

$$Feasibility\ Percentage\ (\%) = \frac{96 + 93 + 101}{112x3}\ x\ 100\% = 86{,}3\%$$

(b) Kegunaan Sistem (*SysUse*)

$$Feasibility\ Percentage\ (\%) = \frac{36 + 35 + 37}{42x3}\ x\ 100\% = 85{,}7\%$$

(c) Kualitas Informasi (*InfoQual*)

$$Persentase\ Kelayakan\ (\%) = \frac{36 + 34 + 38}{42x3}\ x\ 100\% = 85{,}7\%$$

(d) Kualitas Antarmuka (*IntQual*)

$$Feasibility\ Percentage\ (\%) = \frac{24 + 24 + 26}{28x3}\ x\ 100 = 88\%$$

Based on the overall feasibility percentage results, the assessment results obtained from all respondents based on the feasibility percentage are declared "Highly Feasible".

## 4. *Conclusion*

The results of research on mapping waste management facilities per region in Bogor Regency using the *K- Means* approach produced three *clusters*, namely *cluster* 1 (TPS3R) with 26 sub-districts, **cluster** 2 (Waste Bank) with 8 sub-districts, and cluster 3 (Fleet) with 6 sub-districts. The mapping process through a K-Means-based decision support system proved to be more effective than the previous method. The system prototype was developed with Excel/CSV data input, and the results were evaluated using *the silhouette coefficient*. The feasibility test results showed that the system prototype was rated "Highly Feasible" by system experts with a score of 100% and by users with a score of 86.5%, while the K-means method application test using *the silhouette coefficient* yielded a score of 0.354.

## References

[1] S. Annatasya, O. Febrina, E. Riaudan, Q. Aini, dan V. Ivana, “Hidup Lebih Ramah Lingkungan dengan Mengurangi Kantong Plastik di Pesisir Teluk Lampung,” *Jurnal Masyarakat Wirausaha*, vol. 3, no. 1, pp. 1–6, 2024.

[2] N. Suryana, M. Baharuddin, dan Z. Tjenreng, “Pengelolaan Sampah di Kabupaten Bogor: Tantangan dan Solusi,” *Jurnal Lingkungan dan Pembangunan*, vol. 5, no. 1, pp. 339–353, 2025.

[3] Y. G. Nugraha, M. Maimunah, dan P. Sukmasetya, “Penerapan Algoritma K-Means Clustering untuk Daerah Penyebaran Sampah Kelurahan,” *Building Informatics, Technology and Science*, vol. 5, no. 2, pp. 1–10, 2023.

[4] A. N. B. Prasetyo, M. Maimunah, dan P. Sukmasetya, “K-Means Clustering Method for Determining Waste Transportation Routes to Landfill,” *Jurnal Riset Informatika*, vol. 5, no. 3, pp. 547–556, 2023.

[5] A. Data, D. A. N. Penyimpulan, dan E. Penelitian, “Langkah Penelitian Manajemen Pendidikan: Penemuan Masalah, Telaah Pustaka, Persiapan Penelitian, dan Pengumpulan Data,” *Jurnal Manajemen Pendidikan*, vol. 2, pp. 509–523, 2025.

[6] S. Syam et al. (2024). *Data Mining: Teori dan Penerapannya dalam Berbagai Bidang*. PT Sonpedia Publishing Indonesia. [On-

line]. Available: https://books.google.co.id/books?id=hTAxEQAAQBAJ [Diakses: 20 Juni 2025].

[7] N. Hapsari et al. (2024). *Transformasi Digital: Pengantar Literasi Digital untuk Mahasiswa*. Seval Literindo Kreasi. [On-line]. Available: https://books.google.co.id/books?id=3k8KEQAAQBAJ [Diakses: 20 Juni 2025].

[8] Z. Niqotaini, B. S. Yulistiawan, E. Krisnanik, dan R. D. Amalia. (2023). *Analisa dan Perancangan Sistem Informasi dengan Unified Modeling Language*. Bandung: Indie Press.

[9] Z. Setiawan et al. (2023). *Buku Ajar Data Mining*. PT Sonpedia Publishing Indonesia. [On-line]. Available: https://books.google.co.id/books?id=1nLVEAAAQBAJ [Diakses: 20 Juni 2025].

[10] N. Hendrastuty, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering dalam Evaluasi Hasil Pembelajaran Siswa," *Jurnal Pendidikan dan Teknologi Informasi*, vol. 3, pp. 46–56, 2024.

[11] R. Fitriana, A. N. Habyba, dan E. Febriani. (2022). *Data Mining dan Aplikasinya: Contoh Kasus di Industri Manufaktur dan Jasa*. Wawasan Ilmu. [On-line]. Available: https://books.google.co.id/books?id=9XGwEAAAQBAJ [Diakses: 20 Juni 2025].

[12] R. A. Indraputra dan R. Fitriana, "K-Means Clustering Data COVID-19," *Jurnal Informatika*, vol. 10, no. 3, pp. 275–282, 2020.

[13] P. M. Hasugian et al. (2023). *Teknologi Big Data untuk Klasifikasi Status Desa Berdasarkan Indeks Desa Membangun (IDM) Menggunakan Algoritma K-Means*. Cattleya Darmaya Fortuna. [On-line]. Available: https://books.google.co.id/books?id=WTRcEAAAQBAJ [Diakses: 20 Juni 2025].

[14] A. Fitri, N. Masruriyah, C. E. Sukmawati, dan B. A. Dermawan, "Analisis Penerapan Sistem Informasi Berbasis Data," *Jurnal Sistem Informasi*, 2024.

[15] R. F. Putra et al. (2024). *Algoritma Pembelajaran Mesin: Dasar, Teknik, dan Aplikasi*. PT Sonpedia Publishing Indonesia. [On-line]. Available: https://books.google.co.id/books?id=MX7-EAAAQBAJ [Diakses: 20 Juni 2025].

[16] R. Silaban, A. Fauzi, L. Arliana, dan N. Kadim, "Penerapan K-Means Clustering untuk Menentukan Lokasi Promosi Penerimaan Mahasiswa Baru," *Jurnal Teknologi Informasi*, vol. 2, no. 5, pp. 310–318, 2024.

[17] B. Indikator dan I. P. M. Dengan, "Analisis Perbandingan Silhouette Coefficient dan Metode Elbow pada Pengelompokan Provinsi di Indonesia Berdasarkan Indikator IPM," *Jurnal Gaussian*, vol. 13, no. 1, pp. 13–24, 2024.

[18] Y. Januzaj, E. Beqiri, dan A. Luma, "Determining the Optimal Number of Clusters Using Clustering Validation Methods," *International Journal of Computer Science*, pp. 174–182, 2023.

[19] Y. M. Rangkuti et al. (2022). *Visualisasi Penyebaran COVID-19 dengan SIG dan K-Means*. Jejak Pustaka. [On-line]. Available: https://books.google.co.id/books?id=jP2yEAAAQBAJ [Diakses: 20 Juni 2025].

[20] J. Sinuraya et al. (2024). *Analisis dan Perancangan Sistem*. Mega Press Nusantara. [On-line]. Available: [Diakses: 20 Juni 2025].

[21] S. Informasi dan P. Warga, "Penerapan Metode Prototype dalam Merancang Sistem Informasi," *Jurnal Sistem Informasi*, 2024.

[22] Y. Kurniawan. (2022). *Manajemen Sumber Daya Manusia*. Pascal Books. [On-line]. Available: https://books.google.co.id/books?id=t-18EAAAQBAJ [Diakses: 20 Juni 2025].

[23] Y. A. Prasetyo et al., "Pengembangan Web E-Commerce Bojana Sari Menggunakan Metode Prototype," *Jurnal Sistem Informasi*, vol. 2, no. 1, pp. 1042–1056, 2015.

[24] U. Pengembangan et al., "Pengembangan Sistem Informasi Berbasis Web," *Jurnal Teknologi Informasi*, vol. 5, no. 1, pp. 83–95, 2023.

[25] C. D. M. N. A. Azam et al. (2023). *Pengetahuan Dasar Pengembangan Sistem Keamanan Komunikasi Bergerak*. Narotama University Press. [On-line]. Available: https://books.google.co.id/books?id=tv2yEAAAQBAJ [Diakses: 20 Juni 2025].

[26] I. Umila, F. Ratnawati, dan E. Yumami, "Pengujian Kualitas Website Pemerintah Kabupaten Bengkalis Menggunakan Standar ISO 9126," *Jurnal Teknologi Informasi*, vol. 5, no. 1, 2025

[27] S. K. M. T. S. Pratasik. (2022). *Analisis Aplikasi PBB P-2 Integrated Tax System for Urban Need*. CV Azka Pustaka. [On-line]. Available: https://books.google.co.id/books?id=WW7KEAAAQBAJ [Diakses: 20 Juni 2025].

[28] M. Indriani et al., "Evaluasi Usability Aplikasi Streaming Film Menggunakan Metode Post Study System Usability Questionnaire (PSSUQ)," *Jurnal Algoritma*, pp. 221–231, 2024.