

Optimization of Village Grouping Using Comparison of K-Means and K-Medoids Methods

Eza Rahmanita¹, Yeni Kustiyahningsih^{2*}, Putri Nihayatul Husna³, Adhelia Firdaus Al-Najib⁴

^{1,2,3,4}Information System Department, Faculty of Engineering, University of Trunojoyo Madura, East Java, Indonesia

E-mail: ¹eza.rahmanita@trunojoyo.ac.id, ^{2*}ykustiyahningsih@trunojoyo.ac.id, ³putrinihayatul7@gmail.com, ⁴adheliafirdaus11@gmail.com

(Received: 5 May 2025, revised: 27 May 2025, 4 Jun 2025, accepted: 5 Jun 2025)

Abstract

Sumenep Regency is the largest agricultural area in Madura and a major producer of food crops such as rice, corn, and vegetables. However, agricultural productivity in some areas has declined due to uneven fertilizer distribution and limited knowledge about plant diseases and their treatment. To address this, a village clustering system is needed to help the Agriculture Office identify areas with high and low productivity, enabling more targeted assistance and resource allocation. This study aims to classify villages based on agricultural productivity to support better decision-making in agricultural development. Two clustering methods, K-Means and K-Medoids, were applied and compared. K-Means determines cluster centers based on the average of data points, while K-Medoids selects the most representative data point within each cluster. The data used in this research is 690 datasets in 2023. Before clustering is carried out, the data is preprocessed first. Data Pre-processing steps include data transformation, label encoding, imputation, and Min-Max normalization. Cluster optimization was performed using the Sum of Squared Errors (SSE) method. The results show that K-Medoids offers more stable clustering, especially in the presence of outliers, while K-Means is more efficient in computation. The best clustering result was achieved using K-Means with five clusters ($K = 5$), producing the lowest SSE value of 29.8. These findings can assist local governments in prioritizing agricultural support based on village productivity profiles.

Keywords: Clustering, K-Means, K-Medoids, Village, Min-max, Sum of Squared Errors (SSE).

I. INTRODUCTION

The agricultural sector in Indonesia plays a major role in the Indonesian economy. In 2022, the agricultural sector (together with forestry and fisheries) contributed 12.4% of the national GDP. Madura consists of 4 districts, namely Bangkalan, Sampang, Pamekasan and Sumenep. Sumenep Regency is the largest agricultural area and the largest producer of agricultural commodities in Madura, especially food commodities. The agricultural sub-sector as a basic sub-sector for the economy in Sumenep Regency during 2015-2019 consists of the food crop sub-sector, plantation crop sub-sector, livestock and agricultural services and hunting. Village grouping is an important need in order to improve the management and development of the agricultural sector in Sumenep Regency [1]. Agriculture is the primary economic driver in Sumenep Regency, East Java, particularly as the region holds the largest agricultural land area and is the top producer of food commodities in Madura. Key crops such as rice, corn, and various vegetables form the backbone of both

subsistence and commercial farming in the area. Despite this potential, recent data show a decline in agricultural productivity in several villages. This issue is closely tied to uneven distribution of agricultural inputs, such as fertilizers and seeds, and a lack of timely information regarding plant diseases and appropriate interventions.

These problems suggest the need for a more structured and data-driven approach to monitoring and supporting village-level agricultural performance. One way to achieve this is by grouping villages based on their agricultural characteristics, allowing decision-makers to better identify target areas for intervention, resource distribution, and strategic development. Previous studies have applied clustering techniques in the agricultural sector to reveal patterns in crop production, land usage, or climate zones. However, most of these studies focus on national or provincial scales, leaving a gap in village-level analysis where local government interventions are most directly applied. Additionally, many rely on single clustering methods without comparing their robustness or suitability for different data conditions. This study seeks to fill that gap by

analyzing the agricultural productivity of villages in Sumenep Regency using multiple clustering algorithms. The research focuses on key agricultural indicators grouped into three thematic categories: Staple crop production, Horticultural yield, agricultural performance metrics. By applying both K-Means and K-Medoids clustering algorithms and comparing their performance, this study aims to produce a village classification model that is both analytically sound and practically useful for agricultural planning. The analysis also incorporates cluster interpretation to identify the strengths and challenges of each group, thereby informing policy design.

The main objective of this research is to develop a reliable village classification based on agricultural productivity that can assist the Agriculture Office in making targeted, data-driven decisions for resource allocation, program planning, and regional development strategies. In grouping villages based on agricultural productivity, an effective and efficient method is needed. One method that is often used in grouping is the k-means algorithm. The k-means method is very easy to implement and has efficient computing time, especially for large datasets. This method can handle large amounts of data with relatively stable performance, the cluster results from K-Means tend to form clear and separate clusters, facilitating further analysis, K-Means works quite well on data with dimensions that are not too high, suitable for use if the data cluster has a symmetrical shape or resembles a circle (round). The number of clusters (k) can be determined according to analysis needs [2]. While K-Medoids uses medoid (original data) as the cluster center, so it is not easily influenced by extreme data unlike K-Means which uses the average (mean). In this study, there are a number of indicators used, namely plant area, plant variety/type, plant stage/age, type of plant pest (OPT), subsidized organic and inorganic fertilizers used, area of plants attacked, intensity of attack, area of plants in alert condition, and the harvest obtained. K-Means is a non-hierarchical data clustering method that aims to divide data into one or more clusters, where data with similar characteristics are grouped into one cluster, while data with different characteristics are grouped into different groups [3]. The k-medoids method is one of the partition techniques in cluster analysis, which uses medoids objects as cluster centers. Medoids objects are selected as the midpoint of the data group in the cluster [4]. The data that has been obtained will be analyzed using the k-medoids clustering algorithm to group villages based on the needs required. This study compares the approaches of two methods, namely the k-means and k-medoids methods. Several previous studies on K-means and the k-medoids method for clustering tourist visits [5] and clustering drug data [6]. For tourist visits, the optimal cluster is not used in determining the best cluster, while for drug data, the silhouette coefficient is used to determine the optimal cluster. Furthermore, research using the k-means method with elbow cluster optimization to measure student readiness [7]. Both studies state that k-means has better performance than several other clustering methods. Based on the comparison of the two clustering methods, village grouping will be carried out based on agricultural productivity. By using the method to determine the optimal cluster, namely SSE, the smallest SSE

value will be determined. Sum of Squared Errors (SSE) is a statistical method used to measure the total difference between the actual value and the value obtained [8][9]. This method is useful in analyzing the accuracy or error of predictions in a model or algorithm. By squaring the difference between the actual value and the predicted value, SSE provides an overview of the extent to which the model or algorithm has succeeded in describing or predicting data accurately. The smaller the SSE value, the better the model or algorithm is at matching data with actual values [10][11]. By comparing the clustering methods between K-medoid and K Means and cluster optimization using SSE, one method of grouping villages based on agricultural productivity will be selected, with the smallest SSE value that will produce optimal clusters.

II. RESEARCH METHODOLOGY

A. Research Stages

The research stages carried out refer to the flow diagram as shown in Figure 1.

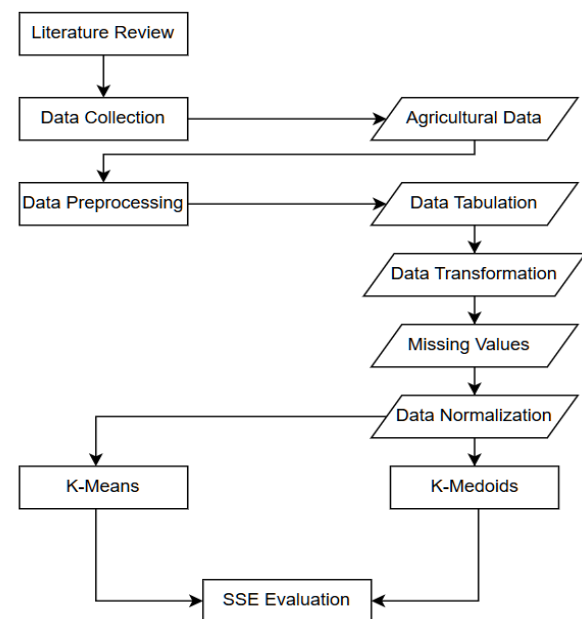


Figure 1. Research Stages

Figure 1 explains the stages carried out in the research process, starting from literature studies, then data were collected from the Sumenep Regency Agriculture Service, Madura. The agricultural dataset was pre-processed (i.e. Tabulation Process, data transformation, missing value, data normalization). The next stage is the clustering process, where the processed data will be clustered using a comparison of the K-Means and k-medoids methods. The last step is the calculation of the Sum of Squared Errors (SSE) to determine the optimal number of clusters.

B. Data Preprocessing

The data used in this study are data on agricultural productivity in Sumenep Regency. The data used are the

results of agricultural productivity in 2023. The data used has 9 indicators, namely plant area, variety, plant age, type of OPT, fertilizer, area attacked, intensity, alert area, and harvest results. Before analyzing the dataset, the data tabulation, data cleaning, data transformation, and data normalization processes were carried out first to ensure the quality of the data to be analyzed.

1. Data Tabulation

Data tabulation is a process where data is corrected or tidied up so that the data can be read and analyzed. This data tabulation aims to correct or tidy up the data so that the data can be analyzed optimally. From the agricultural productivity dataset in Sumenep Regency, it can be seen from the column that there is still a lot of empty data, this is the reason the data is not good enough to carry out the data mining process. Therefore, it requires data processing, namely Data Transformation and Missing Values [12].

2. Converting Data to Numeric (Label Encoding)[11]

Converting data to numeric is used for columns that have non-numeric values to numeric so that the analysis process runs well. In this study, the data transformed into numeric form is in the 'Variety or type' and 'Plant pests (OPT)' columns. Here are sample variables that are converted to numeric: Applied to any categorical variables, such as village types, using Python's sklearn.preprocessing.

Variety or type	Plant Pest Organisms (OPT)
Chrang+inpari = 1	Blas (Magnoportha) = 1
Ciherang = 2	Cercospora arachidicola = 2
Inpari 42 = 3	Kresak (X. Oryzae) = 3
Land/Local = 4	Rat (rattus argentiventer) = 4
	Stripe virus = 5

3. Missing Values

Missing Values are used to fill in unfilled columns with an average or many are called Mean with the formula Data Imputation = Number of Filled Data / Total Filled Data. so that the data can be analyzed optimally.

4. Data Normalization (Min Max Method)

Data normalization aims to make data more interpretable, comparable, or integrated with other data. MinMax normalization is used to change the data scale so that the data has a value that is similar to the Min-Max Normalization formula = $X - X_{min} / X_{max} - X_{min}$. The results of the min and max calculations are in Table 1. Min-Max normalization was used to scale the data between 0 and 1 to prevent features with large ranges from dominating the clustering results. Python programming language was used for all processing steps, along with the Pandas, NumPy, and scikit-learn libraries. All code was executed in a Jupyter Notebook environment.

Table 1. Sample Result of Min and Max Calculations

Variables	Min	Max
Plant Area (Ha)	11	1495
Variety/type	1	4

Variables	Min	Max
Plant Age (hst)	22	77
Types of OPT	1	5
Organic and Inorganic Fertilizers (Tons)	11	1495
Area Affected (Ha)	0.3	11
Intensity (%)	3.33	18.05
Alert Area (Ha)	0	38
Harvest Yield (Tons)	36	8970

Then after determining the minimum and maximum values, each data variable in the data column is calculated for normalization value by subtracting the initial data from the minimum value divided by the result of the maximum value minus the minimum value. So that the calculation is applied to all data variables.

C. K-Means Algorithm [5]

K-Means is one of the non-hierarchical partitioning methods for clustering data. This method divides data into two or more groups, where data with the same characteristics will be grouped together and data with different characteristics will be grouped into another group. K-Means algorithm research that will be carried out refer to the flow diagram as shown in Figure 2 below.

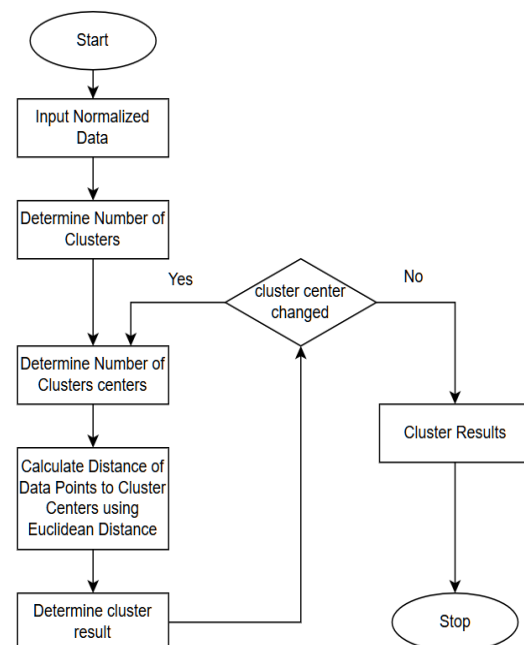


Figure 2. K-Means Algorithm Flowchart

K-Means Method Clustering generally has the following steps according in Figure 2 [13]:

1. Determine the desired number of clusters k.
2. Determine the initial center point (centroid) for each cluster. The initial center is usually determined randomly.
3. Calculate the distance of each data to the nearest cluster center using the following Formula 1:

$$D(i,j) = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + \dots + (x_{ki} - x_{kj})^2} \quad (1)$$

$D(i,j)$ = distance of data i to cluster center j

x_{ki} = i data in k data attribute

x_{kj} = the center point to j on the k th attribute

4. Each data is placed in the cluster with the closest center.
5. Recalculate new cluster centers based on the data that has been placed in the previous cluster.
6. If there is a change in the cluster center, then repeat steps 3-5 until there is no more change in the cluster center.
7. Clustering process is complete.

D. K-Medoids Algorithm [6]

K-Medoids algorithm research that will be carried out refers to the flow diagram as shown in Figure 3 below:

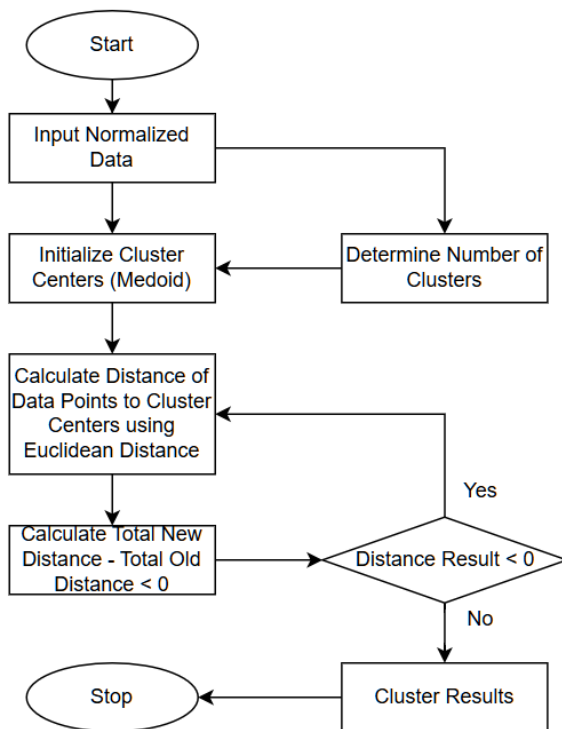


Figure 3. Flowchart of K-Medoids Algorithm

Below are some steps in calculating the K-Medoids Algorithm with Figure 3:

1. The initial step in the clustering algorithm is to initialize the cluster centers as many as k (number of clusters) as determined previously.
2. After that, for each data object (object) in the dataset, grouping is carried out into the closest cluster using the Euclidean Distance measure
3. The next step in the clustering algorithm is to randomly select objects in each cluster as new candidate medoids.
4. After that, the distance calculation is carried out between each object in each cluster with the new candidate medoids.
5. Then, the total deviation (S) is calculated by calculating the difference between the new total distance value and the

old total distance. If $S < 0$, then the object is exchanged with the cluster data to form a set of k new objects as medoids.

6. Steps 3 to 5 are repeated until there is no change in medoids and the clusters and their respective cluster members are obtained. K-Means: A partition-based clustering algorithm that determines cluster centers using the mean of all points in each cluster. It is computationally efficient but sensitive to outliers. K-Medoids: A more robust alternative that selects actual data points as cluster centers, making it less influenced by extreme values.

E. Sum of Squared Errors (SSE) [9]

The next step after the clustering process using the K-Means and K-Medoids methods is to calculate the Sum of Squared Errors (SSE) to determine the optimal method in analyzing agricultural productivity data. Determination of the optimal method is based on the resulting SSE value. The calculation of SSE depends on the results of the sum of the deviations. Deviation is a process in SSE by squaring the distance of each data point to its cluster center. More detailed information about these steps can be seen in the flowchart in Figure 4 below:

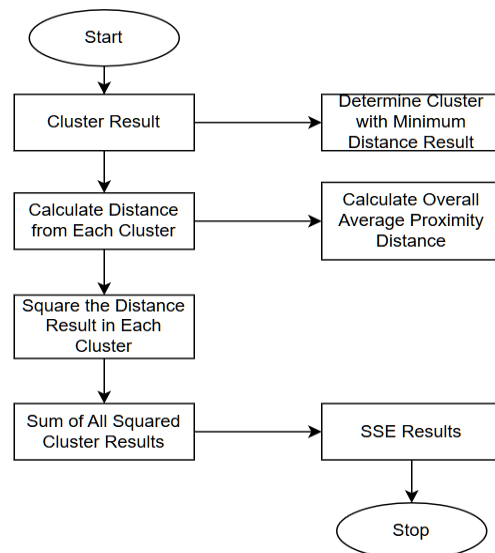


Figure 4. Flowchart of Sum of Squared Errors

III. RESULTS AND DISCUSSION

The results and discussion of this study are started from the data preprocessing process, which is an important stage in this study which aims to prepare raw data so that it is ready for further analysis. The preprocessing process applied is label encoding on categorical variables, data imputation using the mean method, data normalization using the min-max method. The dataset used is the productivity data of the Sumenep Regency harvest which has 9 indicators, namely plant area, variety, plant age, type of OPT, fertilizer, area attacked, intensity, alert area, and harvest yield. The dataset totaled 690 in 2023 as shown in Table 2.

Table 2. Data Preprocessing Results

No	Plant Area (Ha)	Variety/type	...	Harvest Yield (Tons)
1	0.011305	0.216216	...	0.007498
2	0.013312	0.216216	...	0.008836
3	0.011305	0.216216	...	0.007498
4	0.011305	0.216216	...	0.007498
5	1,000,000	0.162162	...	1,000,000
...	...	...	...	...
690	0.020001	0.351351	...	0.023330

The initial step of the preprocessing stage, namely Label Encoding, is to select the data column to be converted into numeric, namely the variety/type column and the OPT type column. Label encoding page is shown in Figure 5.

Data setelah Label Encoding:

	Luas Tanaman (Ha)	Varietas/jenis	Stadia/Umur Tanaman (hst)	Jenis OPT	Pupuk Bersubsidi Organik dan Anorganik (Ton)
0	17	8	35	42	17
1	20	8	32	42	20
2	17	8	35	42	17
3	17	8	34	5	17
4	1,495	6	27	42	1,495
5	198	6	25	42	198
6	199.5	6	22	42	199.5
7	None	8	23	42	None
8	None	8	35	42	None
9	None	17	45	42	None

Figure 5. Label Encoding Page

After completing the label encoding process, the next step in the preprocessing stage is data imputation or filling in empty values using the mean method as shown in Figure 6. Final step of the preprocessing process is the min-max method normalization. Min-Max Normalization is a technique used to change the range of values in a dataset into a certain range, usually between 0 and 1, as shown in Figure 7.

Data setelah Missing Values diisi dengan Mean:

	Luas Tanaman (Ha)	Varietas/jenis	Stadia/Umur Tanaman (hst)	Jenis OPT	Pupuk Bersubsidi Organik dan Anorganik (Ton)
0	17	8	35	42	17
1	20	8	32	42	20
2	17	8	35	42	17
3	17	8	34	5	17
4	1,495	6	27	42	1,495
5	198	6	25	42	198
6	199.5	6	22	42	199.5
7	63.6983	8	23	42	63.6983
8	63.6983	8	35	42	63.6983
9	63.6983	17	45	42	63.6983

Figure 6. Data Imputation Page (Mean)

Data setelah Min-Max Normalisasi:

	Luas Tanaman (Ha)	Varietas/jenis	Stadia/Umur Tanaman (hst)	Jenis OPT	Pupuk Bersubsidi Organik dan Anorganik (Ton)
0	0.0113	0.2162	0.16	0.8235	0.0113
1	0.0133	0.2162	0.1429	0.8235	0.0133
2	0.0113	0.2162	0.16	0.8235	0.0113
3	0.0113	0.2162	0.1543	0.098	0.0113
4	1	0.1622	0.1143	0.8235	1
5	0.1324	0.1622	0.1029	0.8235	0.1324
6	0.1334	0.1622	0.0857	0.8235	0.1334
7	0.0425	0.2162	0.0914	0.8235	0.0425
8	0.0425	0.2162	0.16	0.8235	0.0425
9	0.0425	0.4595	0.2171	0.8235	0.0425

Figure 7. Min-Max Normalization Page

The next step is clustering using the K-Means Algorithm and the K-Medoids Algorithm. Users are given the option to choose the number of clusters according to their preferences. However, in the context of the task of the Sumenep Agriculture Service, it has been determined that the clustering process must be carried out by dividing the data into three clusters. The results of this clustering process will be displayed in the form of a table, and in the table there will be an additional column called 'Cluster' as shown in Table 3. After that, under the clustering process there is a display of the SSE results (Table 4).

Table 3. K-Means Method Clustering Process Page

No	A	B	C	D	E	F
1	0.0113	0.0536	0.0061	0.093	0.0075	1
2	0.0113	0.0179	0.0061	0.110	0.0088	1
3	0.0113	0.0179	0.0122	0.093	0.0075	1
4	0.0113	0.0179	0.0061	0.093	0.0075	2
5	1	0.025	0.0208	0.027	1	1
6	0.1324	0.025	0.0302	0.093	0.1324	1
7	0.1334	0.025	0.0233	0.116	0.1334	1
8	0.0425	0.1429	0.024	0	0.0413	1
9	0.0425	0.1071	0.0079	0	0.0413	1
10	0.0425	0.1786	0.026	0	0.0413	1

A : Organic and Inorganic Fertilizers (Tons)
B : Area Affected (Ha)
C : Intensity (%)
D : Alert Area (Ha)
E : Harvest Yield (Tons)
F : Cluster

Table 4. K-Medoids Method Clustering Process Page

No	A	B	C	D	E	F
1	0.0113	0.0536	0.0061	0.093	0.007	2
2	0.0113	0.0179	0.0061	0.110	0.008	2
3	0.0113	0.0179	0.0122	0.093	0.007	2
4	0.0113	0.0179	0.0061	0.093	0.007	3
5	1	0.025	0.0208	0.027	1	2
6	0.1324	0.025	0.0302	0.093	0.132	2



No	A	B	C	D	E	F
7	0.1334	0.025	0.0233	0.116	0.133	2
8	0.0425	0.1429	0.024	0	0.041	2
9	0.0425	0.1071	0.0079	0	0.04	2
10	0.0425	0.1786	0.026	0	0.041	1

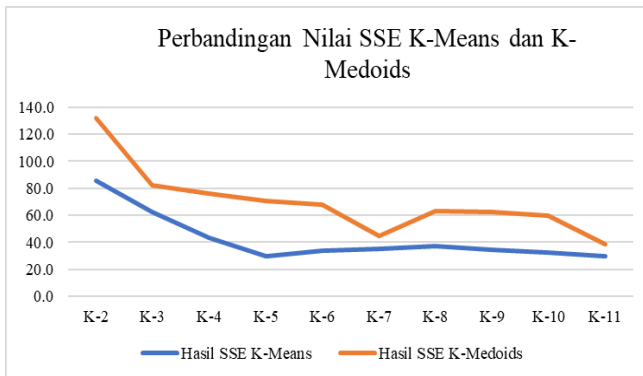


Figure 8. SSE Graph of K-Means and K-Medoids

The results of the analysis of the SSE K-Means graph show that the lowest SSE value is at K = 5, with an SSE value of 29.8 (Figure 8). The results of the analysis of the SSEK-Medoids graph show that the lowest SSE value is found at K = 7, with an SSE value of 44.80. Based on the trial of several clusters on the K-Means and K-Medoids algorithms, the smallest error value is K-Means, so it can be seen that the K-Means method is a more optimal choice for grouping Sumenep Regency harvest data in this study. Detailed result is shown in Table 5.

Table 5. Comparison of SSE K-Means and K-Medoids

Cluster	SSE K-Means Results	K-Medoids SSE Results
K-2	85.5	131.78
K-3	62.5	82.27
K-4	43.4	76.22
K-5	29.8	70.82
K-6	33.6	67.67
K-7	34.8	44.80
K-8	37.1	62.88
K-9	34.4	62.62
K-10	32.1	59.77
K-11	30.0	38.55

The results of village grouping using the K-Means method produced five distinct clusters based on harvest productivity indicators in Sumenep Regency. The distribution of villages in each cluster is as follows: Cluster 1 consists of 158 villages, Cluster 2 includes 108 villages, Cluster 3 has 190 villages, Cluster 4 includes 76 villages, and Cluster 5 consists of 158 villages. Each cluster exhibits unique characteristics based on the analyzed indicators: Cluster 1 demonstrates high average values in planting area, crop variety, and harvest yield. This suggests that these villages possess strong agricultural potential, with diverse crop varieties and expansive cultivation. These areas could benefit from infrastructure

development and access to markets to further increase productivity and economic returns. Cluster 2 is marked by high averages in plant age, type of pests and diseases (OPT), and the use of both organic and inorganic fertilizers. This profile indicates that these villages may focus on longer-cycle crops and actively manage pest threats through fertilizer usage. It highlights the importance of sustainable pest and nutrient management programs. Cluster 3 shows high values in the attacked area and damage intensity indicators. These villages are experiencing severe pest infestations and plant damage, implying the urgent need for targeted pest control measures, extension services, and agricultural resilience strategies. Cluster 4 records low average values across all indicators. These are likely underdeveloped agricultural zones with minimal input usage and poor output. Such areas may require basic agricultural assistance, investment in infrastructure, training, and access to farming technology to improve performance. Cluster 5 has high averages in alert area, harvest yield, and damage intensity. This mixed profile suggests that although these villages produce relatively well, they are at risk due to high alert conditions and pest intensity. Monitoring systems and early warning mechanisms would be essential here. These cluster profiles not only show statistical distinctions but also offer practical insights for designing location-specific agricultural policies. Understanding the dominant characteristics of each cluster helps policymakers and stakeholders prioritize resources, design appropriate interventions, and tailor support programs based on actual field conditions.

IV. CONCLUSION

In the clustering process using the K-Means and K-Medoids methods, evaluation was carried out using the Sum of Square Error (SSE) to assess the clustering quality. The K-Means method showed the lowest SSE value at K = 5, amounting to 29.8, while the K-Medoids method recorded its lowest SSE at K = 7, with a value of 44.80. Based on this comparison, the K-Means method is considered more optimal for clustering harvest data in Sumenep Regency. Beyond the numerical results, this study offers practical benefits, particularly for agricultural planning. Accurate clustering can support more targeted decision-making related to resource distribution, planting strategies, and agricultural development policies. However, the study has limitations, such as relying solely on SSE for evaluation and not accounting for external factors like climate conditions or agricultural infrastructure. Future research is recommended to incorporate additional evaluation metrics—such as the Silhouette Coefficient or Davies-Bouldin Index—and to expand the range of variables analyzed to produce more accurate and applicable clustering results.

ACKNOWLEDGMENT

We would like to express our sincere gratitude to LPPM UTM and the Department of Agriculture, Sumenep Regency,



for their valuable contribution and continuous support in the implementation of this research.

REFERENCES

- [1] E. Rahmanita, Y. D. P. Negara, Y. Kustiyahningsih, V. Sasmeka, and B. K. Khotimah, "Implementasi Metode Naïve Bayes dan Information Gain Untuk Klasifikasi Penyakit dan Hama Tanaman Jagung," *Teknika*, vol. 12, no. 3, pp. 198–204, 2023, doi: 10.34148/teknika.v12i3.684.
- [2] T. Terttiaavini, "A Hybrid Approach Using K-Means Clustering and the SAW Method for Evaluating and Determining the Priority of SMEs in Palembang City," *INSYST*, vol. 6, no. 1, pp. 46–53, Apr. 2024.
- [3] Y. D. Darmi and A. Setiawan, "Penerapan Metode Clustering K-Means Dalam Pengelompokan Penjualan Produk," *J. Media Infotama*, vol. 12, no. 2, pp. 148–157, 2017, doi: 10.37676/jmi.v12i2.418.
- [4] D. Marlina, N. Lina, A. Fernando, and A. Ramadhan, "Implementasi Algoritma K-Medoids dan K-Means untuk Pengelompokan Wilayah Sebaran Cacat pada Anak," *J. CoreIT J. Has. Penelit. Ilmu Komput. dan Teknol. Inf.*, vol. 4, no. 2, p. 64, 2018, doi: 10.24014/coreit.v4i2.4498.
- [5] Y. Kusnadi and M. S. Putri, "Clustering Menggunakan Metode K-Means Untuk Menentukan Prioritas Penerima Bantuan Bedah Rumah (Studi Kasus: Desa Ciomas Bogor)," *J. Teknol. Inform. dan Komput.*, vol. 7, no. 1, pp. 17–24, 2021, doi: 10.37012/jtik.v7i1.498.
- [6] R. A. Farissa, R. Mayasari, and Y. Umaidah, "Perbandingan Algoritma K-Means dan K-Medoids Untuk Pengelompokan Data Obat dengan Silhouette Coefficient di Puskesmas Karangsambung," *J. Appl. Informatics Comput.*, vol. 5, no. 2, pp. 109–116, 2021, doi: 10.30871/jaic.v5i1.3237.
- [7] N. T. Hartanti, "Metode Elbow dan K-Means Guna Mengukur Kesiapan Siswa SMK Dalam Ujian Nasional," *J. Nas. Teknol. dan Sist. Inf.*, vol. 6, no. 2, pp. 82–89, 2020, doi: 10.25077/teknosi.v6i2.2020.82-89.
- [8] A. Winarta and W. J. Kurniawan, "Optimasi cluster k-means menggunakan metode elbow pada data pengguna narkoba dengan pemrograman python," *J. Tek. Inform. Kaputama*, vol. 5, no. 1, pp. 113–119, 2021.
- [9] F. Marisa, S. S. S. Ahmad, Z. I. M. Yusof, Fachrudin, and T. M. A. Aziz, "Segmentation model of customer lifetime value in Small and Medium Enterprise (SMEs) using K-Means Clustering and LRFM model," *Int. J. Integr. Eng.*, vol. 11, no. 3, pp. 169–180, 2019, doi: 10.30880/ijie.2019.11.03.018.
- [10] A. Winarta and W. J. Kurniawan, "Optimasi cluster k-means menggunakan metode elbow pada data pengguna narkoba dengan pemrograman python," *J. Tek. Inform. Kaputama*, vol. 5, no. 1, pp. 113–119, 2021.
- [11] B. K. Khotimah, F. Agustina, O. R. Puspitarini, A. D. Cahyani, Y. Kustiyahningsih, and D. R. Anamisa, "Hyperparameters and Centroid Improvements in the K-Medoids Method for Grouping Processed Beef SMEs," *Commun. Math. Biol. Neurosci.*, vol. 2024, pp. 1–21, 2024, doi: 10.28919/cmbn/8369.
- [12] A. R. Ismail, N. Z. Abidin, and M. K. Mean, "Systematic Review on Missing Data Imputation Techniques with Machine Learning Algorithms for Healthcare," *Journal of Robotics and Control*, Vol. 3, No. 2, pp. 143–152, 2022, <https://doi.org/10.18196/jrc.v3i2.13133>.
- [13] R. Supardi and I. Kanedi, "Implementasi Metode Algoritma K-Means Clustering pada Toko Eidelweis," *J. Teknol. Inf.*, vol. 4, no. 2, pp. 270–277, 2020, doi: 10.36294/jurti.v4i2.1444.