Implementation of the K-Medoids Clustering Method in Grouping Districts in Sleman Regency, Yogyakarta According to the Amount of Fruit Production in 2023

Puspa Chandra Wijayati[1]; Kariyam[2]
[1,2]Universitas Islam Indonesia
21611012@students.uii.ac.id [1], kariyam@uii.ac.id[2]

Abstract

Agriculture is one of the main factors in economic development in Indonesia, as Indonesia is an agrarian country. The government also expects agricultural production of food crops to increase every year. This study aims to determine the clustering results and characteristics of each sub-district in Sleman Regency based on the amount of fruit production. The method used in this research is K-Medoids Clustering. The K-Medoids method includes a clustering algorithm that is quite efficient in clustering small data and finding the most representative points and being able to overcome outliers. The results showed the application of the K-Medoids method resulted in 5 clusters where Cluster 1 has 4 sub-districts namely, Moyudan, Minggir, Seyegan, and Godean sub-districts, Cluster 2 has 5 sub-districts namely, Gamping, Mlati, Depok, Berbah, and Prambanan sub-districts, Cluster 3 has 4 sub-districts namely, Kalasan, Ngemplak, Ngaglik, and Sleman sub-districts, Cluster 4 has 2 sub-districts namely, Tempel and Turi sub-districts, and Cluster 5 has 2 sub-districts namely, Pakem and Cangkringan sub-districts.
Keywords: *Agriculture; Production; K-Medoids*

Abstrak

Pertanian adalah salah satu faktor utama dalam pembangunan ekonomi di Indonesia, karena Indonesia termasuk dalam daerah agraris. Pemerintah juga mengharapkan produksi pertanian tanaman pangan mengalami peningkatan tiap tahunnya. Pada penelitian ini bertujuan untuk mengetahui hasil pengelompokkan dan karakteristik dari masing-masing kecamatan di Kabupaten Sleman berdasarkan jumlah produksi buah-buahan. Metode yang digunakan dalam penelitian ini adalah *K-Medoids Clustering.* Metode *K-Medoids* termasuk algoritma pengelompokan yang cukup efesien dalam melakukan pengelompokan terhadap data yang kecil dan pencarian titik yang paling representatif dan mampu mengatasi *outlier*. Hasil penelitian menunjukkan penerapan metode *K-Medoids* menghasilkan 5 *cluster* dimana *Cluster* 1 memiliki 4 kecamatan yaitu, Kecamatan Moyudan, Minggir, Seyegan, dan Godean, *Cluster* 2 memiliki 5 kecamatan yaitu, Kecamatan Gamping, Mlati, Depok, Berbah, dan Prambanan, *Cluster* 3 memiliki 4 kecamatan yaitu, Kecamatan Kalasan, Ngemplak, Ngaglik, dan Sleman, *Cluster* 4 memiliki 2 kecamatan yaitu, Kecamatan Tempel dan Turi, dan *Cluster* 5 memiliki 2 kecamatan yaitu, Kecamatan Pakem dan Cangkringan.
Kata Kunci: Pertanian; Produksi; *K-Medoid*

INTRODUCTION

Agriculture is a key factor in economic development in Indonesia, as Indonesia is an agrarian region. The agricultural sector plays a significant role, including being the largest employer, producing food for the population, and determining price stability. The government also expects annual food crop production to increase. (Hadiyatullah, Arifianto, & Al Faruq, 2024). Regional economic growth is fundamentally influenced by the region's comparative advantages, regional specialization, and its economic potential. Therefore, utilizing and developing all economic potential is a top priority that must be explored and developed in implementing sustainable regional economic development. (Istiqamah & Novita, 2017).

Fruit is a food source that is rich in various vitamins, minerals and nutrients that are beneficial for the body. (Fadhila & Febriyanti, 2022). Fruit is also a food source of vitamins. Fruit is quickly damaged by mechanical, chemical and microbiological influences so it easily rots. Therefore, it is necessary to process fruit with the aim of extending its shelf life. Fruit processing is intended to transform fruit into products that are more durable and easy to consume. Apart from that, fruit processing is also aimed at increasing the variety of products so that people can taste fruit even if it is not in season (Yusmita & Wijayanti, 2018).

Production is an activity carried out to add value to an object or create an object so that it is more useful (Yogatama, 2024). According to (Sleman, 2024) Production is the yield of each vegetable, fruit, biopharmaceutical and ornamental plant taken based on the area harvested in the reporting month/quarter. Seasonal fruit plants are plants that are sources of vitamins, mineral salts and others that are consumed from the part of the plant that is in the form of fruit, are less than one year old, can be in the form of a tree/clump but creeping and have soft stems. Perennial fruit plants are plants that are sources of vitamins, mineral salts and others that are consumed from the part of the

plant that is in the form of fruit and are annual plants.

As a regency where the majority of its population relies on agriculture, Sleman Regency is committed to advancing the agricultural sector to improve the standard of living of its residents. Other supporting sectors, such as horticulture, also contribute to the development of fruit centers. According to (Rianto, 2023) The highest banana production is in Prambanan, followed by Ngemplak, Cangkringan, and Sleman. The cultivation system is mostly still spread out in residents' yards. Some locations that are cultivated in a stretch include PT. Pesona Republik Pisang, which has a 4-hectare area or a population of 8,000 Cavendish trees cultivated in Bedoyo, Wukirsari, and Cangringan. In Sleman Regency, per capita consumption of Ambon bananas is 0.078 kg per week and other types of bananas is 0.164 kg. Per capita consumption of this banana is second only to papaya at 0.171 kg and oranges at 0.124 kg.

Grouping of sub-districts in Jember Regency based on the production results of superior commodities has been carried out using the k-medoids partition method. (Alfian Hadiyatullah dkk, 2018). The optimum cluster results were obtained in the scenario of 2 clusters up to 10 clusters based on the smallest Davies Bouldin Index value of 0.087 in 2 clusters with 30 sub-districts in cluster 1 and 1 sub-district in cluster 2. However, in cluster 2, which has 1 sub-district, there were outliers indicated, so the further calculation process could not be carried out. Therefore, the calculation process was carried out on the second optimum cluster, namely on 4 clusters with a value of 0.612 (Hadiyatullah, Arifianto, & Al Faruq, 2024). Halimatusakdiah Pohan's research and friends The application of the k-medoids algorithm in clustering stunted toddlers in Indonesia aims to group provinces experiencing stunting with the highest and lowest clusters, which is useful as input for the government to quickly address stunting reduction in Indonesia. The application of the k-medoids partition method was also carried out by Athallah Widyatama Rafii. to group hotels in Jakarta based on the tiket.com website. Research comparing the k-means and k-medoids algorithms has also

been conducted in grouping districts and cities based on biopharmaceutical plants. (Anwar, Rinaldi, & Mulyawan, 2023).

Based on the above results, the researcher is interested in conducting research on grouping sub-districts in Sleman Regency, Yogyakarta, based on fruit production volume. The purpose of this research is to facilitate the Central Statistics Agency (BPS) of Sleman Regency in maintaining the highest quality of fruit production and prioritizing fruit production by sub-district in Sleman Regency.

RESEARCH METHODS

The data used in this study were obtained from the Fruit Production Amount by District in Sleman Regency. The four variables used refer to the names of fruits found in Sleman Regency. These variables are Durian, Siamese Orange, Banana, and Papaya. To determine the grouping of districts based on the amount of fruit production, K-Medoids Cluster analysis was used. The steps used to group the districts are as follows.

1. Input data on the amount of fruit production by sub-district in Sleman Regency in 2023.
2. Process the data on the amount of fruit production by sub-district in Sleman Regency in 2023 into ready-to-process data.
3. Conduct descriptive analysis.
4. Standardize the data based on fractional ranking.
5. Create sub-district groupings.

RESULTS AND DISCUSSION

Descriptive statistics is the process of classifying and describing information. It describes, displays, and summarizes the basic characteristics of a data set found in a particular study to make it easier to understand. (Dwiyanto, 2023). Descriptive statistics only relate to describing or providing information about data, conditions, or phenomena. In other words, descriptive statistics function to explain conditions, symptoms, or problems. (Nasution, 2017). The 2023 data on fruit production by sub-

district in Sleman Regency used were for durian, tangerine, banana, and papaya. Figure 1.1 below shows that the average production of durian is higher than that of other fruits. This indicates that the average value of durian production varies significantly across sub-districts in Sleman Regency.

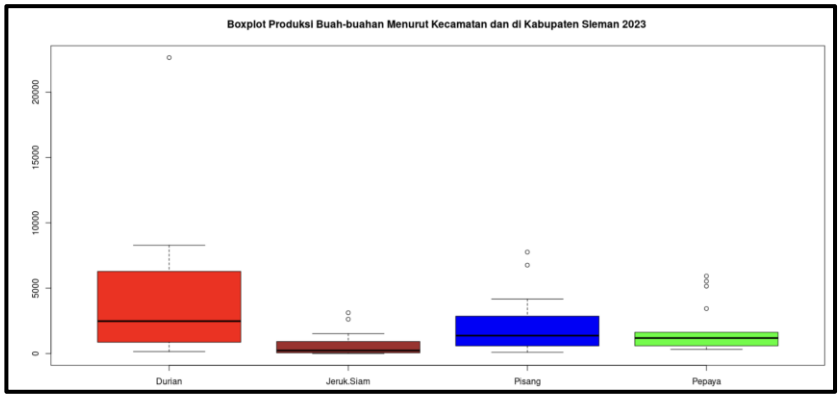

Figure 1. Boxplot of Data Before Standardization.

It is necessary to carry out a data standardization process by transforming the original data before further analysis if the variables being studied have large differences in unit sizes. (Alwi & Hasrul, 2018). Data standardization with fractional ranking is the process of converting the scale or range of values from data to a scale that makes it easier to analyze. Each value in the data is converted to a fractional rank relative to other values in the same data set. This process is done by assigning a rank to each data value, with the lowest value being given the lowest rank, and so on. If there are identical values, the average rank of the corresponding positions is assigned. Then, the original data values are replaced with their fractional ranks. The formula used is as follows: (Kariyam, Abdurakhman, Subanar, Utami, & Effendie, 2022).

$$z_{li} = f.\left(\frac{r_{li} - r_{l1}}{r_{lm} - r_{l1}}\right); i = 1, 2, \dots, n$$

Information:
$i = 1, 2, \dots, n$
$r_{li} = rank$ data ke $- i$ dalam variabel ke $- l$
$r_{l1} = rank$ data terendah dalam variabel ke $- l$

$r_{lm} = rank$ data tertinggi dalam variabel ke $- l$
$Z_{li} =$ nilai Z ke $- l$ dalam variabel ke $- l$

Data standardization results in data with previously different means and variances becoming equal. This can be seen in Figure 2.

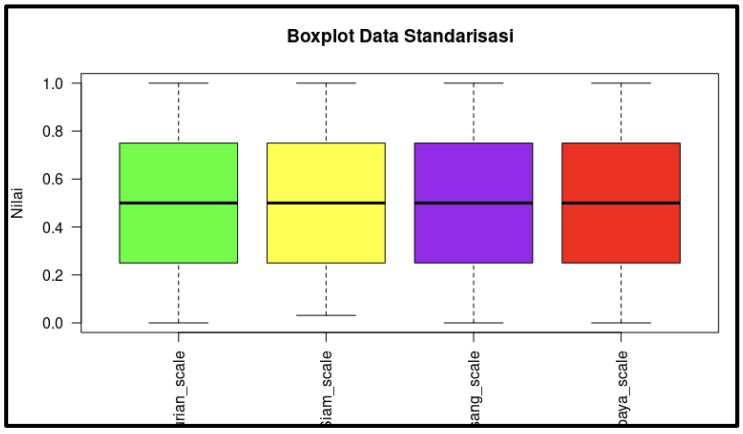

Figure 2. Boxplot of Data After Standardization

The data resulting from standardization has the same distribution as the original data. Standardization using fractional ranking focuses only on the relative order of the values in the data. Therefore, data standardization using fractional ranking is considered superior because it can handle outliers. *K-Medoids* is a non-hierarchical clustering algorithm. K-medoids is a non-hierarchical clustering algorithm derived from the k-means algorithm. Through partition clustering, the algorithm groups x objects into k clusters. At the center of each cluster, objects that are robust to outliers are called medoids. Clusters are formed by calculating the distance between medoid and non-medoid objects. (Dinata, Retno, & Hasdyna, 2021).

K-Medoids or Partitioning Around Medoids (PAM) is a clustering partitioning method for grouping a set of (n) objects into a number of (k) clusters. (Budiaji & Leisch, 2019). In addition to PAM, the k-medoids (KM) and simple and fast k-medoids (SFKM) algorithms have been developed. Both algorithms have limited medoid

exchange steps. These algorithms work within clusters only, i.e., similar to centroid updating in k-means. Meanwhile, SFKM selects initial medoids using a set of sorted vectors vj. The value of vj is defined as the standardized number of rows or standardized number of columns of the Euclidean distance matrix.

Euclidean distance is a calculation to measure the distance between two points in Euclidean space that relates angles and distances. In mathematics, Euclidean distance is used to measure two points in one dimension, giving results similar to the Pythagorean theorem. (Miftahuddin et al., 2020). Euclidean distance has several advantages, including having more optimal results compared to other calculations, and the concept of Euclidean distance is relatively easy to understand and easy to interpret. (Pribadi et al., 2022).

From the clustering using the k-medoids method, we can obtain a grouping of sub-districts in Sleman Regency according to fruit production volume, as presented in Table 1. The first column in Table 1 represents the sub-districts within Sleman Regency, while column 2 represents the resulting clusters.

The number of clusters is determined by the number k obtained, namely k = 5. This cluster formation is also suggested by the results of the Silhouette Method, where the dotted line refers to the number 5, indicating the optimum cluster value. The following graph shows the results of cluster formation using the Silhouette Method.

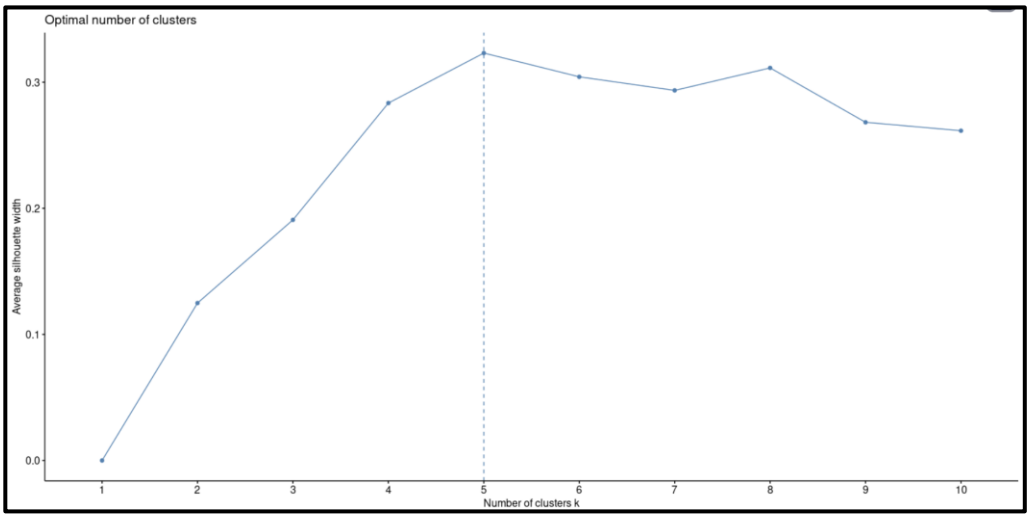

Figure 3. Silhouette Graphics

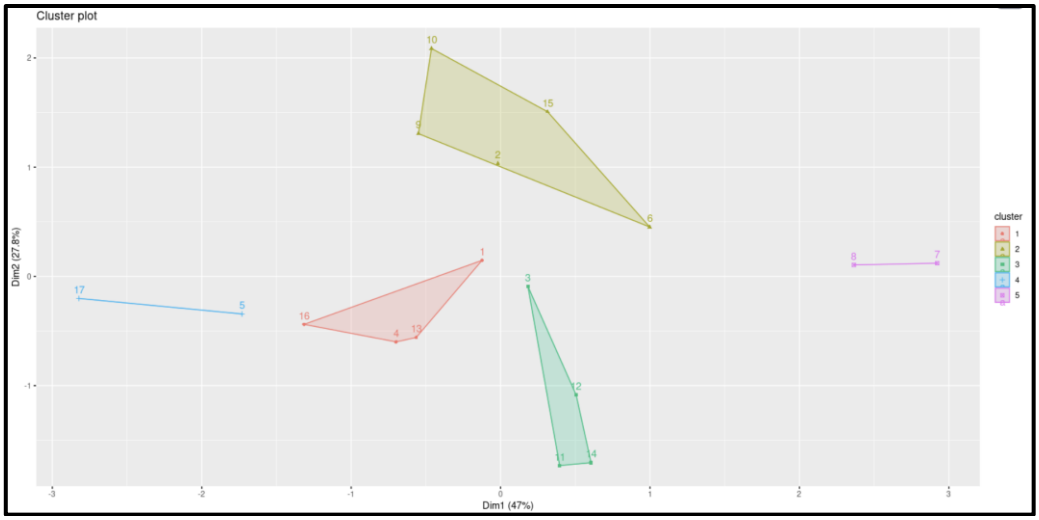

Figure 4. District Grouping

Table 1. Clustering Results

Kecamatan	Klaster
Moyudan	1
Minggir	1
Seyegan	1
Godean	1
Gamping	2
Mlati	2
Depok	2
Berbah	2
Prambanan	2
Kalasan	3
Ngemplak	3
Ngaglik	3
Sleman	3
Tempel	4
Turi	4
Pakem	5
Cangkringan	5

Table 2. Variable Grouping

Cluster	Durian	Siamese Oranges	Banana	Papaya
1	0.422	0.859	0.594	0.500
2	0.725	0.338	0.588	0.213
3	0.328	0.422	0.313	0.828
4	0.875	0.813	0.906	0.875
5	0.063	0.031	0.063	0.188

In cluster 1, it has a larger average variable than clusters 2, 3, 4, and 5. Therefore, it can be concluded that cluster 1 has a very high amount of fruit production. For cluster 2, it has a larger average variable than clusters 3, 4, and 5. Therefore, it can be concluded that cluster 2 has a high amount of fruit production. Meanwhile, for

cluster 3, it has a larger average variable than clusters 4 and 5. Therefore, it can be concluded that cluster 3 has a moderate amount of fruit production. Furthermore, cluster 4 has a larger average variable than cluster 5. Therefore, it can be concluded that cluster 4 has a low amount of fruit production. Also, in cluster 5, it has a smaller average variable than clusters 1, 2, 3, and 4. Therefore, it can be concluded that cluster 5 has a very low amount of fruit production.

CONCLUSION

The characteristic results obtained were that in cluster 1, the Siamese orange had a higher average value. In cluster 2, the durian had a higher average value. Meanwhile, in cluster 3, the papaya had a higher average value. Furthermore, in cluster 4, the banana had a higher average value. Finally, in cluster 5, the papaya had a higher average value. It can be concluded that the higher average value of the variable is what produces more fruit in each sub-district within the cluster group.

Cluster 1 had a higher average value of the variable than clusters 2, 3, 4, and 5. Therefore, it can be concluded that cluster 1 has a very high fruit production volume. In cluster 2, the average value of the variable is higher than clusters 3, 4, and 5. Therefore, it can be concluded that cluster 2 has a high fruit production volume. Meanwhile, cluster 3 has a larger average variable than clusters 4 and 5. Therefore, it can be concluded that cluster 3 has a moderate amount of fruit production. Furthermore, cluster 4 has a larger average variable than cluster 5. Therefore, it can be concluded that cluster 4 has a low amount of fruit production. Also, cluster 5 has a smaller average variable than clusters 1, 2, 3, and 4. Therefore, it can be concluded that cluster 5 has a very low amount of fruit production.

REFERENCE

Alwi, W., & Hasrul, M. (2018). *ANALISIS KLASTER UNTUK PENGELOMPOKAN KABUPATEN/KOTA DI PROPINSI SULAWESI SELATAN BERDASARKAN INDIKATOR KESEJAHTERAAN RAKYAT*. Retrieved from file:///C:/Users/Asus/Downloads/4782-Article%20Text-12413-1-10-20180703.pdf

Anwar, A. M., Rinaldi, A. R., & Mulyawan. (2023, Desember). *JATI*. Retrieved from C:/Users/Asus/Downloads/8188-Article%20Text-32589-1-10-20240203.pdf

Budiaji, W., & Leisch, F. (2019). *Simple K-Medoids Partitioning Algorithm for Mixed Variable Data*. Retrieved from file:///C:/Users/Asus/Downloads/algorithms-12-00177.pdf

Dinata, R. R., Retno, S., & Hasdyna, N. (2021). *Minimization of the Number of Iterations in K-Medoids Clustering with Purity Algorithm*. Retrieved from file:///C:/Users/Asus/Downloads/35.03_02.pdf

Dwiyanto, A. (2023, Januari). Retrieved from https://www.researchgate.net/publication/367487749_Statistika_Deskriptif_Pengertian_Fungsi_dan_Jenisnya

Fadhila, A., & Febriyanti, E. (2022). *https://jurnal.umsu.ac.id/index.php/JIH/article/view/11801/pdf*. Retrieved from https://jurnal.umsu.ac.id/index.php/JIH/article/view/11801/pdf

Hadiyatullah, A., Arifianto, D., & Al Faruq, H. A. (2024). *PENGELOMPOKAN KECAMATAN BERDASARKAN HASIL PRODUKSI KOMODITAS UNGGULAN DI KABUPATEN JEMBER MENGGUNAKAN ALGORITMA PARTITIONING AROUND MEDOIDS*. Retrieved from http://repository.unmuhjember.ac.id/5445/10/j.%20JURNAL.pdf

Istiqamah, N., & Novita, U. D. (2017). KAJIAN PENGEMBANGAN KOMODITAS UNGGULAN BUAH-BUAHAN DI KABUPATEN SAMBAS. Retrieved from file:///C:/Users/Asus/Downloads/757-2500-1-PB.pdf

Kariyam, Abdurakhman, Subanar, Utami, H., & Effendie, A. R. (2022). *Block-Based K-Medoids Partitioning Method with Standardized Data to Improve Clustering Accuracy*. Retrieved from file:///C:/Users/Asus/Downloads/09.06_22.pdf

Nasution, L. M. (2017). *STATISTIK DESKRIPTIF*. Retrieved from file:///C:/Users/Asus/Downloads/16-Article%20Text-62-1-10-20200814.pdf

Pohan, H., Zarlis, M., Irawan, E., Okprana, H., & Purba, Y. P. (2021, November). Retrieved from https://www.ioinformatic.org/index.php/JUKI/article/view/69/83

Rafii, A. W., & Hakim, R. F. (2023). Retrieved from https://journal.uii.ac.id/esds/article/view/28159/15347

Rianto, F. (2023, Januari Jumat). *Popularitas Tinggi, Kabupaten Sleman Kembangkan Potensi Buah Pisang*. Retrieved from https://jogja.times.co.id/news/ekonomi/h9x4un514m/Popularitas-Tinggi-Kabupaten-Sleman-Kembangkan-Potensi-Buah-Pisang

Sleman, B. K. (2024). KABUPATEN SLEMAN DALAM ANGKA 2024. In B. K. Sleman. Sleman: BPS Kabupaten Sleman.

Yogatama, I. (2024). *TEORI PRODUKSI*. Retrieved from http://eprints.umsida.ac.id/7043/1/191020700121_Teori%20Produksi.pdf

Yusmita, L., & Wijayanti, R. (2018). *Jurnal Teknologi dan Industri Pertanian Indonesia*. Retrieved from file:///C:/Users/Asus/Downloads/10152-26807-1-PB.pdf