SITEKIN: Jurnal Sains, Teknologi dan Industri, Vol. 20, No. Y, June 20xx, pp.549 - 558
ISSN 2407-0939 print/ISSN 2721-2041 online

Cluster Analysis of Indonesian Provinces Based On Harvest Area And Rice Productivity Using Single Linkage Method

Herlinda Nur'Afwa Sofhya[1]
[1] Department of Mathematics Education, Faculty of Education and Teachers Training, IAIN Syekh Nurjati Cirebon
Jl. Perjuangan bypass, Cirebon City, West Java, 45134
Email: herlindanurafwa@syekhnurjati.ac.id

In this article, a cluster analysis will be conducted for provinces in Indonesia based on the harvest area (ha) and rice productivity (ku/ha) of 34 provinces in Indonesia. Clustering is done using a hierarchical method, namely single linkage. The distance used as the basis for clustering is the euclidian distance. Based on the results of clustering using single linkage, 3 large clusters were obtained. In this article, a cluster analysis will be conducted for provinces in Indonesia based on the harvest area (ha) and rice productivity (ku/ha) of 34 provinces in Indonesia. clustering is done using a hierarchical method, namely single linkage. The distance used as the basis for clustering is the euclidian distance. Based on the results of clustering using single linkage, 3 large clusters were obtained. Cluster consists of 3 provinces, cluster 2 consists of 1 province and cluster 3 consists of 30 provinces. Cluster 1 is a province with high rice production with an average total rice production of 9,628,788 tons. Cluster 2 with an average total rice production of 5,341,021 tons. While cluster 3 with an average rice production of 863,995.34 tons. Furthermore, based on cluster validation using the anova test, the significance value is 0.00>0.05, which means that there is a significant difference between clusters. Thus it can be stated that the division of 34 Indonesian provinces in terms of land area and rice productivity into 3 large clusters using the single linkage method is valid.

Keywords: *harvested area, rice productivity, single linkage, euclidian distance, anova test*

Introduction

The amount of rice production is a very important thing to consider in Indonesia. Because rice is the staple food of the Indonesian people. Ensuring the availability of rice is one of the main tasks of the Indonesian government. One of the policies to fulfil the availability of rice is the government's rice import policy. Based on data from the Badan Pusat Statistik (BPS) Indoensia, rice imports in 2020 were 356.286,2 tonnes while in 2021 they were 407.741,4. This means that there is an increase in rice imports from 2020 to 2021 for 2022 from January to November the government has imported 326.450 tonnes of rice. Indonesia's rice imports are still quite high, even though the Indonesian government aspires to be self-sufficient in rice. Based on the 1999 FAO decree, a country is said to be self-sufficient if the amount of domestic production reaches 90% of national needs [1]. To reduce the level of rice imports, the amount of domestic rice production needs to be increased [2]. The government must determine the right policy to increase the amount of domestic rice production. The amount of rice production in each province in Indonesia varies. Based on BPS data, the amount of rice production is based on the harvest area and rice productivity in a region. This is in line with Nazzarudin's research which states that rice production is strongly influenced by harvest area [3]. This research will cluster provinces in Indonesia based on harvest area and rice productivity. This clustering is done to facilitate the government in seeing the potential of rice production in each province in Indonesia. Hopefully, the government can take the right policy for each region or province.

Cluster analysis is a multivariate analysis. Cluster analysis is carried out to group the objects of observation into several groups based on the observed variables. Grouping is done based on similarity so that variables that have similarities are in one group while variables between groups are not similar or different. There are many methods used in the clustering analysis process [4]–[8]. Hierarchical clustering methods are used to group observations in a structured manner based on similarity properties and the desired group number is not yet known while non-hierarchical methods are used to group data into k groups where the number of groups can be determined by yourself. In non-hierarchical methods, the number of clusters is divided at the beginning. [9]. Whereas if you only look at the initial data, it will be difficult for researchers to determine the number of clusters to be formed. Whereas in the hierarchical method the number of clusters is not determined at the beginning so that researchers can divide the data into several clusters seen from the comparison of distances between data. The distance between data states the similarity of the data [10]. Several cluster analysis studies using the hierarchical method have been conducted with different variables, some of which are cluster analysis based on air pollution levels [11], cluster analysis of the economic impact of covid 19 in Indonesia

SITEKIN: Journal of Science, Technology and Industry, Vol. X, No. Y, June 20xx, pp.1 - 4
ISSN 2407-0939 print/ISSN 2721-2041 online

[12], cluster analysis based on infectious diseases [13], cluster analysis based on calorie consumption of residents in maluku province [14] and many other studies.

In the hierarchical method, there are also several ways to cluster data based on the distance that has been obtained [15]–[19]. One of the most commonly used clustering methods is the single linkage method, where data is grouped based on the minimum distance between data. Previous research conducted by Rendy by comparing the k means and single linkage methods on document grouping concluded that the single linkage method has better performance than the K-means method seen from the silhouette coefficient and purity value [20]. Several cluster analysis studies using the single linkage method with different variables show excellent cluster results [21][22]. The cluster results obtained using the single linkage method show that there are very significant differences between clusters and homogeneous data are in one cluster. This is because the single linkage method classifies data with the closest distance into one group. Therefore, this study will conduct a cluster analysis using the single linkage method to classify 34 provinces in Indonesia based on the harvest area and rice productivity of each province.

Research Methods

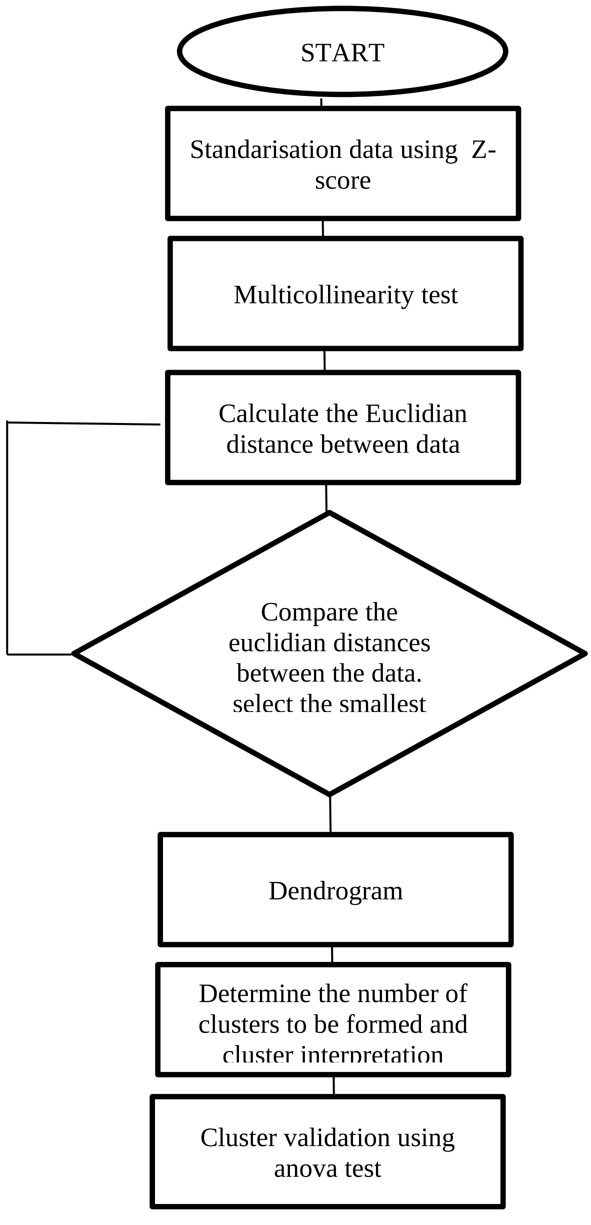

Figure 1. Cluster analysis steps using single linkage method

The cluster analysis method used in this study is the single linkage method. Several stages are passed in the first stage of clustering analysis, the most important thing in the cluster analysis problem is the selection of variables that will be used for clustering to include one or two variations of variables that are irrelevant to the clustering problem, the clustering results are likely to be very useful [23]–[26]. Basically, the variables to be selected must describe the similarity between objects, which is really relevant to the problem discussed. After selecting the data, the next step is data standardisation. Data standardisation usually needs to be done when using original data because usually the range of data is very large. Variables that have large values have a greater influence in making classification predictions than variables with small values[27]. To overcome this problem, a variable normalisation technique can be used so that all variables will differ in the same range. The way to determine the standardisation value is to calculate the mean and variance value of each variable and then look for the normalisation value. There are various ways to obtain the normalisation value but in this study using the Z-Score method [28]. The formula for obtaining data normality values using Z-score is

$$x_{norm} = \frac{x-\mu}{\sigma} \tag{1}$$

x_{norm} : the normalisation result value of x
x : original data
μ : data average
σ : standard deviation of data

Next is to detect multicollinearity. Correlation analysis needs to be done to determine whether there is a relationship between research variables. The correlation coefficient between independent variables must be uncorrelated or weakly correlated or below 0.8 [29]. If the correlation between research variables is strong or the correlation coefficient between variables is above 0.8, then there is a multicollinearity problem. In the previous stage, the research data has been standardised, which means that it is normally distributed so that for correlation analysis, we can use the commonly used classical correlation test, namely using the Pearson correlation test. The formula used to calculate the correlation coefficient using Pearson is as follows

$$r_{xy} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{\left(n\sum x_i^2 - (\sum x_i)^2\right)}\sqrt{\left(n\sum y_i^2 - (\sum y_i)^2\right)}} \tag{2}$$

r_{xy} : correlation coefficient between variables x and y
n : number of data
x_i : i-th x data value
y_i : i-th y data value

After the data is confirmed not to experience multicollinearity, the next step is to calculate the distance between data. In cluster analysis using hierarchical methods such as single likange, data classification is carried out based on the distance between data. Therefore, before clustering the data, it is necessary to calculate the distance between data first. The most commonly used distance is the Euclidean distance [30]. The Euclidian distance between two data expresses the similarity between the two data. The distance measure between the i-th data and the j-th data can be calculated through the calculation of the squared Euclidean distance as follows:

$$d_{ij} = \sqrt{\left(x_i - x_j\right)^2 + \left(y_i - y_j\right)^2} \tag{3}$$

d_{ij} : distance between i-th data and j-th data
x_i : i-th x data value
y_i : i-th y data value
x_j : j-th x data value
y_j : j-th y data value

After obtaining the distance between data, we then enter the clustering procedure using the single linkage method. In the single linkage method, the clustering process is carried out based on the minimum distance between data. In each iteration of the clustering process using the single linkage method is to find the smallest euclidian distance between two data, then the two data will form a cluster. So that the next iteration will compare the euclidian distance between a cluster with other data and so on until all data helps a cluster based on the closest euclidian distance [31]. The smallest euclidian distance states the similarity between data. So the smaller the euclidian distance between two data means the more similar the two data are. The results of classification using hierarchical methods such as single linkage are generally displayed in the form of a dendrogram. Through the dendrogram, researchers make decisions about how many clusters to choose. In this single linkage method, the class is not determined at the beginning but is determined after comparing the euclidian distance between data as will be presented in the dendrogram. Dendrogram is a visual representation of the cluster formation process based on the distance coefficient value at each step until the final cluster is formed[32]. After obtaining the dendrogram, the next step is to determine the number of clusters to be selected and interpret the cluster. Then the last stage that must be done in this cluster analysis is cluster validation. One of the Cluster validity tests that can be used is the Anova Test [33]. This Anova test is carried out to ascertain whether there is a significant difference between clusters or not. If there is a significant difference between clusters, it means that the clustering is valid. The decision-making criteria for the ANOVA test is if the significance value <0.05 means that there is a significant difference between groups[34].

Results and Discussion

The data used in this study is rice production data in terms of harvest area (ha) and rice productivity (ku/ha) from 34 provinces in Indonesia. Data on harvest area and rice productivity were obtained from the 2022 data published by BPS Indonesia.

Table 1. Harvest area and rice productivity by province in 2022

Province	Harvest Area (ha) 2022	Rice Productivity (ku/ha) 2022
Aceh	276622,14	55,03
North Sumatera	423522,28	52,00
West Sumatera	288510,67	48,36
Riau	54317,04	40,98
Jambi	63760,91	46,29
South Sumatera	516259,59	51,44
Bengkulu	58663,78	48,67
Lampung	516910,01	50,77
Kep, Bangka Belitung	15908,70	38,57
Kep, Riau	196,53	31,65
DKI Jakarta	535,63	58,03
West Java	1685295,13	56,81
Central Java	1699436,08	56,69
DI Yogyakarta	112148,00	51,77
East Java	1704759,48	56,02
Banten	338454,39	50,38
Bali	114790,87	58,83
West Nusa Tenggara	269827,26	51,39
East Nusa Tenggara	185737,54	41,85
West Kalimantan	272115,99	31,90
Central Kalimantan	109756,22	30,28
South Kalimantan	225483,04	39,97
East Kalimantan	64031,22	36,92
North Kalimantan	10550,13	33,74
North Sulawesi	59081,54	39,35
Central Sulawesi	173238,56	47,59
South Sulawesi	1042107,35	51,67
Southeast Sulawesi	119662,53	41,57
Gorontalo	48497,60	48,12
West Sulawesi	71470,11	52,05
Maluku	23991,26	41,24
North Maluku	6408,19	36,05
West Papua	5475,82	41,98
Papua	48987,63	44,05
Indonesia	10606513,22	52,26

Based on the data presented in Table.1, it can be seen that in the harvest area variable, the smallest data is 196,5 and the largest data is 1.704.759,48. The range between the largest data and the smallest data is too wide, which will affect the classification process. The classification process because variables that have large values have a greater influence on classification predictions than variables with small values. Therefore, it is necessary to standardise the data so that all variables will differ in the same range. By using formula (1), the data obtained after standardisation as presented in Table 2.

Table 2. Results of standardisation of harvest area and rice productivity data

Province	Harvest Area	Rice Productivity
Aceh	-0.07295	1.11190
North Sumatera	0.23033	0.74121
West Sumatera	-0.04840	0.29589
Riau	-0.53190	-0.60699
Jambi	-0.51241	0.04264
South Sumatera	0.42179	0.67270
Bengkulu	-0.52293	0.33381
Lampung	0.42313	0.59073
Kep, Bangka Belitung	-0.61120	-0.90184
Kep, Riau	-0.64364	-1.74844
DKI Jakarta	-0.64294	1.47893
West Java	2.83529	1.32967
Central Java	2.86449	1.31499
DI Yogyakarta	-0.41251	0.71307
East Java	2.87548	1.23302
Banten	0.05471	0.54302
Bali	-0.40705	1.57680
West Nusa Tenggara	-0.08698	0.66658
East Nusa Tenggara	-0.26058	-0.50056
West Kalimantan	-0.08225	-1.71786
Central Kalimantan	-0.41745	-1.91605
South Kalimantan	-0.17853	-0.73056
East Kalimantan	-0.51185	-1.10370
North Kalimantan	-0.62226	-1.49275
North Sulawesi	-0.52207	-0.80641
Central Sulawesi	-0.28639	0.20168
South Sulawesi	1.50741	0.70084
Southeast Sulawesi	-0.39700	-0.53481
Gorontalo	-0.54392	0.26652
West Sulawesi	-0.49649	0.74733
Maluku	-0.59451	-0.57519
North Maluku	-0.63081	-1.21014
West Papua	-0.63274	-0.48465
Papua	-0.54291	-0.23141

After the data has been standardised, the next step is the multicollinearity assumption test. This test is carried out to see whether there is a linear relationship between the research variables or not. Because the data has been standardised, a correlation test can be used using the Pearson correlation test. So that the correlation coefficient is obtained as presented in Table 3.

Table 3. Pearson correlation test results

Correlations

		Harvest (ha)	Rice Productivity (ku/ha)
Harvest (ha)	Pearson Correlation	1	.513**
	Sig. (2-tailed)		.002
	N	34	34
Rice Productivity (ku/ha)	Pearson Correlation	.513**	1

SITEKIN: Journal of Science, Technology and Industry, Vol. X, No. Y, June 20xx, pp.1 - 4
ISSN 2407-0939 print/ISSN 2721-2041 online

	Sig. (2-tailed)	.002	
	N	34	34

Table 2 shows that the correlation coefficient between harvested area and rice productivity is 0.513. The correlation coefficient is less than 0.8, meaning that there is no strong correlation between the variables of harvest area and productivity. Thus, there is no multicollinearity problem so that we can proceed to the data clustering process using the single linkage method.

In cluster analysis using the hierarchical method, the classifier process is based on the distance between data. The most commonly used distance between data is the Euclidian distance. Grouping dara using the single linkage method is to find data based on the closest Euclidian distance. Or in other words, in the single linkage method, data that has the closest euclidian distance will be in one group. Euclidian distance expresses the similarity between two data. So the closer the euclidian distance between two data means the more similar the two data are. Thus, to perform cluster analysis using single linkage, it is necessary to calculate the euclidian distance first. By using formula (3), the euclidian distance between can be obtained as presented in Table 4.

Table 4. Euclidian distance between data

No	1	2	3	4	…	34
1	0	0,199	0,271	0,275	…	0,347
2		0	0,072	0,076	…	0,148
3			0	0,071	…	0,143
4				0	…	0,072
⋮	⋮	⋮	⋮	⋮	0	⋮
34						0

The euclidian distance in Table 4 shows how similar the amount of production between provinces is in terms of land area and productivity. This euclidian distance is used as the basis in the classification process. In this research, the single likage method is used so that the provinces will be grouped with the closest euclidian distance. The clustering process using the Single linkage method is carried out by comparing the euclidian distance between provinces in terms of harvest area and productivity then a decision is made that the minimum euclidian distance will become one cluster. The clustering process stops until all data has entered the cluster. In the single linkage method, it is easier to see the data grouping process using a dendrogram. As described earlier that in hierarchical cluster analysis methods such as single linkage, the number of classes is not determined at the beginning but is determined after obtaining the dendrogram. Through the dendrogram we can see the data that has similarities will form a small cluster then a small cluster that has similarities will form a new larger cluster and so on until all data is formed in a large cluster. The results of clustering data of Indonesian provinces in terms of harvest area and productivity using the single linkage method are presented in the dendrogram in Figure 1.

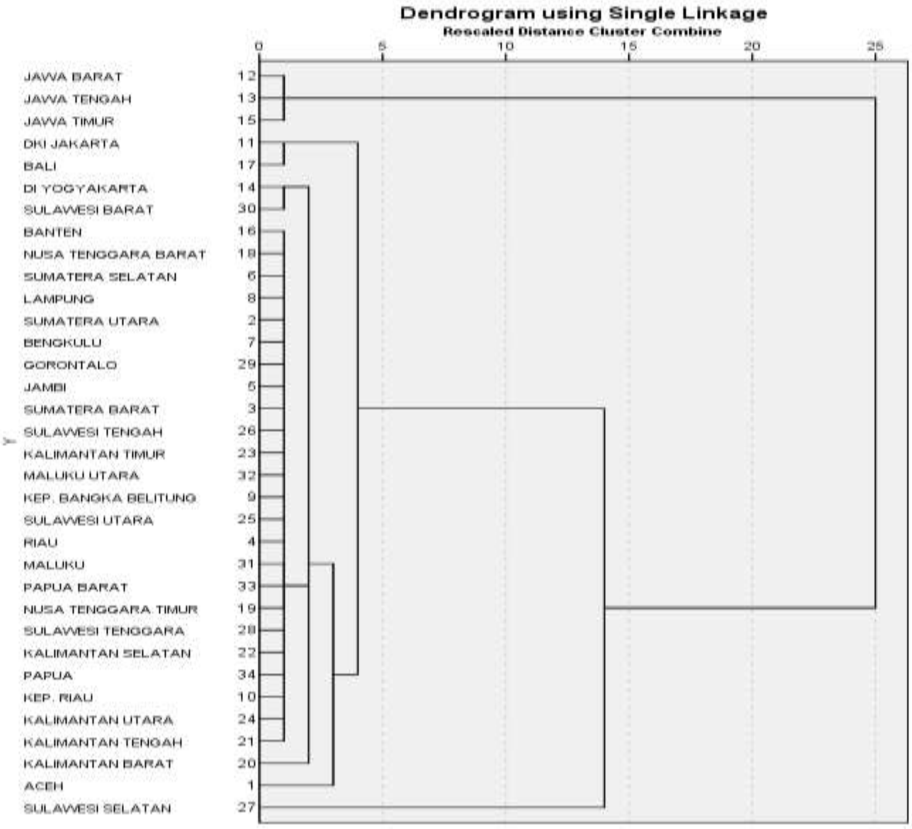

Figure 2. Dendrogram of clustering results of indonesian provinces data based on harvest area and rice productivity

To determine the number of clusters on the dendrogram, we can draw a vertical line. Based on the results on the dendrogram, we can divide the provinces in Indonesia into 3 large clusters. Cluster 1 consists of 3 provinces, cluster 2 consists of 1 province and cluster 30 provinces. The details of the 3 clusters are presented in Table 5.

Table 5. Clustering results of Indonesian provinces based on harvest area and rice productivity

Cluster 1	Cluster 2	Cluster 3
West Java	South Sulawesi	Aceh
Central Java		North Sumatera
East Java		West Sumatera
		Riau
		Jambi
		South Sumatera
		Bengkulu
		Lampung
		Kep, Bangka Belitung
		Kep, Riau
		DKI Jakarta
		DI Yogyakarta
		Banten
		Bali
		West Nusa Tenggara
		East Nusa Tenggara
		West Kalimantan
		Central Kalimantan
		South Kalimantan
		East Kalimantan
		North Kalimantan
		North Sulawesi
		Central Sulawesi
		Southeast Sulawesi
		Gorontalo

	West Sulawesi Maluku North Maluku West Papua Papua

Further interpretation of the clustering results. Total rice production can be calculated from the product of harvested area and productivity. The average total rice production of cluster 1 is 9.628.788 tonnes, cluster 2 has an average total rice production of 5.341.021 tonnes, while cluster 3 has an average rice production of 863.995,34 tonnes. There is a significant difference in the amount of rice production between clusters. The provinces in cluster 1 have very high rice production, reaching 9 million tonnes. While cluster two has a medium amount of rice production, cluster 3 shows provinces with a low amount of rice production with an average of not even reaching 1 million tonnes. Whenviewed in more detail in cluster 1 with high rice production, the average harvest area is 1.696.496,89 ha and the average rice productivity is 56.50667 (ku/ha). In cluster 2 with medium rice production, the average harvest area is 1.042.107,35 ha and the average rice productivity is 41,57 (ku/ha). In cluster 3 with low rice production, the average harvest area is 177.968,46 ha and the rice productivity is 47.72042 (ku/ha). Interestingly, the cluster with low rice production had a higher average rice productivity than the cluster with medium rice production. However, the average harvested area is much different. In the medium cluster the average harvested area reaches 1 million ha, but in the low cluster the average harvested area is only around 100 thousand. This shows that the main factor that differentiates the amount of rice production in each province is due to differences in harvest areas. Thus, to increase the amount of rice production in Indonesia, in addition to increasing productivity, the government must also be able to increase the harvest area in each province.

Furthermore, the final stage of this cluster analysis is cluster validation. Cluster validation will be carried out using the Anova Test. Through the Anova Test results we can see whether there are significant differences between clusters. A cluster analysis is said to be invalid if there is no significant difference between clusters. Because the purpose of the clustering process is to group similar data into one cluster. By using SPSS 20, the anova test results are presented in Table 6.

Table 6. Anova test results

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Harvest Area (ha)	Between Groups	7079020104767.	2	3539510052383.	165.409	.00
	Within Groups	663352644253.4	31	21398472395.27		
	Total	7742372749021.	33			
Rice Productivity (ku/ha)	Between Groups	414.372	2	207.186	3.587	.04
	Within Groups	1790.412	31	57.755		
	Total	2204.783	33			

Based on the Anova Test results, the significance value is 0.00 <0.05. So it can be concluded that there is a significant difference in the average data between clusters. Thus the results of clustering 34 Indonesian provinces into 3 large clusters based on harvest area and rice productivity using the single linkage method are valid. cluster 1 criteria are provinces with high rice production, cluster two with medium rice production and cluster 3 with low rice production.

Conclusion

Based on the results of clustering using the single linkage method, if the harvest area and rice productivity of 34 Indonesian provinces can be divided into 3 clusters. Cluster 1 consists of 3 provinces that are West Java, Central Java and East Java. Cluster 2 consists of 1 province, namely South Sulawesi. And Cluster 3 consists of 30 provinces namely Aceh, North Sumatra, West Sumatra, Riau, Jambi, South Sumatra, Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, DKI Jakarta, DI Yogyakarta, Banten, Bali, West Nusa Tenggara, East Nusa Tenggara, West Kalimantan, Central Kalimantan, South Kalimantan, East Kalimantan, North Kalimantan, North Sulawesi, Central Sulawesi, Southeast Sulawesi, Gorontalo, West Sulawesi, and Maluku.

Cluster 1 is a province with high rice production with the average harvest area is 1.696.496,89 ha and the average rice productivity is 56.50667 (ku/ha). In cluster 2 with medium rice production, the average harvest area is 1.042.107,35 ha and the average rice productivity is 41,57 (ku/ha). Cluster 3 with low rice production, the average harvest area is 177.968,46 ha and the rice productivity is 47.72042 (ku/ha). The results of cluster validation using the anova test show that there are significant differences between cluster groups. Thus it can be said that the clustering of Indonesian provinces based on harvest area and rice productivity in this research is valid.

References

[1] S. Konyep, "Upaya Pencapaian Swasembada Pangan Melalui Membumikan Padi Amfibi Balitbangtan di Provinsi Papua Barat," *Triton*, vol. 11, no. 2, pp. 32–41, 2020.

[2] S. Paipan and M. Abrar, "Determinan Ketergantungan Impor Beras Di Indonesia (Determinants of Rice Import Dependency in Indonesia)," *J. Ekon. dan Kebijak. Publik*, vol. 11, no. 01, pp. 53–64, 2020.

[3] Nazaruddin, "Luas Tanam Dan Luas Panen Padi Di Jawa Barat," *J. Trit.*, vol. 10, no. 1, pp. 59–68, 2019.

[4] E. P. Cynthia *et al.*, "Convolutional Neural Network and Deep Learning Approach for Image Detection and Identification," in *Journal of Physics: Conference Series*, 2022, vol. 2394, no. 1, p. 12019.

[5] M. L. Hamzah, M. Rizki, and M. I. H. Umam, "Integration of Fuzzy Logic Algorithms with Failure Mode and Effect Analysis for Decision Support Systems in Product Quality Improvement of Piano Cabinets," in *2022 International Conference on Electrical and Information Technology (IEIT)*, 2022, pp. 13–19.

[6] M. Yanti, F. S. Lubis, N. Nazaruddin, M. Rizki, S. Silvia, and S. Sarbaini, "Production Line Improvement Analysis With Lean Manufacturing Approach To Reduce Waste At CV. TMJ uses Value Stream Mapping (VSM) and Root Cause Analysis (RCA) methods," 2022.

[7] M. L. Hamzah, A. A. Purwati, S. Sutoyo, A. Marsal, S. Sarbani, and N. Nazaruddin, "Implementation of the internet of things on smart posters using near field communication technology in the tourism sector," *Comput. Sci. Inf. Technol.*, vol. 3, no. 3, pp. 194–202, 2022.

[8] S. Sarbaini, "Modeling of Traffic Flow Schemes at Road Intersections in Pekanbaru City Using Compatible Graphs," *Eduma Math. Educ. Learn. Teach.*, vol. 11, no. 2, pp. 213–222, 2022.

[9] R. H. B. Bangun, "Analisis Klaster Non-Hierarki Dalam Pengelompokan Kabupaten/Kota Di Sumatera Utara Berdasarkan Faktor Produksi Padi," *Agrica (Jurnal Agribisnis Sumatera Utara)*, vol. 4, no. 1, pp. 54–61, 2016.

[10] T. E. Nita and L. Zahrotun, "Penerapan Metode Single Linkage dengan Manhattan Distance Similarity dalam Mengelompokkan Trens Topik Kerja Praktik Application of the Single Linkage Method with Manhattan Distance Similarity in Grouping Trens of Practical Work Topics," *J. Ris. Sains dan Teknol.*, vol. 5, no. 1, pp. 39–44, 2021.

[11] D. Rachmatin, "Aplikasi metode-metode agglomerative dalam analisis klaster pada data tingkat polusi udara," *J. Ilm. Progr. Stud. Mat. STKIP Siliwangi Bandung*, vol. 3, no. 2, pp. 133–149, 2014.

[12] I. N. Hasanah and A. Sofro, "Analisis Cluster Berdasarkan Dampak Ekonomi Di Indonesia Akibat Pandemi Covid-19," *J. Ilm. Mat.*, vol. 10, no. 02, 2022.

[13] N. Ulinnuh and R. Veriani, "Analisis Cluster dalam Pengelompokan Provinsi di Indonesia Berdasarkan Variabel Penyakit Menular Menggunakan Metode Complete Linkage , Average Linkage dan Ward," *J. Nas. Inform. dan Teknol. Jar.*, vol. 5, no. 1, 2020.

[14] G. Haumahu and Y. W. Nanlohy, "Penerapan Analisis Klaster Hierarki Untuk Pengelompokkan Kabupaten / Kota Di Provinsi Maluku Berdasarkan Konsumsi Kalori (The Application of Hierarchical Cluster Analysis for Grouping District / City in Maluku Province Based on Population Calories Consum," *Var. J. Stat. its Appl.*, vol. 2, no. 2, pp. 75–79, 2020.

[15] V. Devani, M. I. H. Umam, Y. Aiza, and S. Sarbaini, "Optimization of Tire Production Planning Using The Goal Programming Method and Sensitivity Analysis," *Int. J. Comput. Sci. Appl. Math.*, vol. 8, no. 2, pp. 36–40, 2022.

[16] S. Sarbaini, Z. Zukrianto, and N. Nazaruddin, "Pengaruh Tingkat Kemiskinan Terhadap Pembangunan Rumah Layak Huni Di Provinsi Riau Menggunakan Metode Analisis Regresi Sederhana," *J. Teknol.*

dan Manaj. Ind. Terap., vol. 1, no. III, pp. 131–136, 2022.

[17] N. Nazaruddin and S. Sarbaini, "Evaluasi Perubahan Minat Pemilihan Mobil dan Market Share Konsumen di Showroom Pabrikan Honda," *J. Teknol. dan Manaj. Ind. Terap.*, vol. 1, no. II, pp. 97–103, 2022.

[18] S. Sarbaini, M. Imran, and A. Karma, "Metode Bertipe Steffensen dengan Orde Konvergensi Optimal untuk Menyelesaikan Persamaan Nonlinear." Riau University, 2014.

[19] S. Sarbaini, W. Saputri, and F. Muttakin, "Cluster Analysis Menggunakan Algoritma Fuzzy K-Means Untuk Tingkat Pengangguran Di Provinsi Riau," *J. Teknol. dan Manaj. Ind. Terap.*, vol. 1, no. II, pp. 78–84, 2022.

[20] R. Handoyo, "perbandingan metode clustering menggunakan metode single linkage dan k-means pada pengelompokan dokumen," *JSM SRMIK Mikroskil*, vol. 15, no. 2, pp. 73–82, 2014.

[21] Suyanto, Syarippudin, and Wasono, "Di Kabuapten Kutai KartannegaraTahun 2019 Single Linkage Cluster Analysis Based on Village Potential," *EKSPONENSIAL*, vol. 12, no. 1, pp. 59–64, 2021.

[22] N. Asiska, N. Satyahadewi, and H. Perdana, "Pencarian Cluster Optimum Pada Single Linkage , Complete Linkage Dan Average Linkage," *Bul. Ilm. Math, Stat, dan Ter.*, vol. 08, no. 3, pp. 393–398, 2019.

[23] S. Sarbaini, E. P. Cynthia, and M. I. Arifandy, "Pengelompokan Diabetic Macular Edema Berbasis Citra Retina Mata Menggunakan Fuzzy Learning Vector Quantization (FLVQ)," *SITEKIN J. Sains, Teknol. dan Ind.*, vol. 19, no. 1, pp. 75–80, 2021.

[24] M. I. Arifandy, E. P. Cynthia, and F. Muttakin, "Potensi Limbah Padat Kelapa Sawit Sebagai Sumber Energi Terbarukan Dalam Implementasi Indonesian Sustainability Palm Oil," *SITEKIN J. Sains, Teknol. dan Ind.*, vol. 19, no. 1, pp. 116–122, 2021.

[25] F. Muttakin, K. N. Fatwa, and S. Sarbaini, "Implementasi Additive Ratio Assessment Model untuk Rekomendasi Penerima Manfaat Program Keluarga Harapan," *SITEKIN J. Sains, Teknol. dan Ind.*, vol. 19, no. 1, pp. 40–48.

[26] S. Sarbaini and E. Safitri, "Penerapan Metode Single Exponential Smoothing dalam Memprediksi Jumlah Peserta Pelatihan Masyarakat," *Lattice J. J. Math. Educ. Appl.*, vol. 2, no. 2, pp. 103–117, 2022.

[27] R. D. Fitriani, H. Yasin, and Tarno, "PENANGANAN KLASIFIKASI KELAS DATA TIDAK SEIMBANG DENGAN RANDOM OVERSAMPLING PADA NAIVE BAYES (Studi Kasus: Status Peserta KB IUD di Kabupaten Kendal)," *J. GAUSSIAN*, vol. 10, no. 1, pp. 11–20, 2021.

[28] R. G. Whendasmoro and Joseph, "Analisis Penerapan Normalisasi Data Dengan Menggunakan Z-Score," *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 4, pp. 872–876, 2022, doi: 10.30865/jurikom.v9i4.4526.

[29] C. V. Bertan, A. K. T. Dundu, and R. J. M. Mandagi, "Pengaruh Pendayagunaan Sumber Daya Manusia (Tenaga Kerja) Terhadap Hasil Pekerjaan (Studi Kasus Perumahan Taman Mapanget Raya (TAMARA))," *J. Sipil Statik Vol.4*, vol. 4, no. 1, pp. 13–20, 2016.

[30] J. Supranto, *Analisis Multivariat: Arti dan interpretasi*. Jakarta.: PT. Rineka Cipta., 2004.

[31] M. Roux, "A Comparative Study of Divisive and Agglomerative Hierarchical Clustering Algorithms," *J. Classif.*, vol. 35, 2018, doi: 10.1007/s00357-018-9259-9.

[32] M. Paramadina, S. Sudarmin, and M. K. Aidid, "Perbandingan Analisis Cluster Metode Average Linkage dan Metode Ward (Kasus: IPM Provinsi Sulawesi Selatan)," *VARIANSI J. Stat. Its Appl. Teach. Res.*, vol. 1, no. 2, p. 22, 2019, doi: 10.35580/variansiunm9357.

[33] D. Noor, P. Sari, and Y. L. Sukestiyarno, "Analisis Cluster dengan Metode K-Means pada Persebaran Kasus COVID-19 Berdasarkan Provinsi di Indonesia," *PRISMA*, vol. 4, pp. 602–610, 2021.

[34] A. Septiadi and W. K. Ramdhani, "Penerapan Metode Anova untuk Analisis Rata-rata produksi donat, burger, dan Croissant pada Toko Roti Animo Bakery," *Bull. Appl. Ind. Eng. Theory*, vol. 1, no. 2, pp. 60–64, 2020.