JTAM (Jurnal Teori dan Aplikasi Matematika)
http://journal.ummat.ac.id/index.php/jtam
p-ISSN 2597-7512 | e-ISSN 2614-1175
Vol. 8, No. 3, July 2024, pp. 800-810

Implementation of Data Mining and Spatial Mapping in Determining National Food Security Clusterization

Sifriyani[1*], I Nyoman Budiantara[2], M. Fariz Fadillah Mardianto[3], Eka Riche Febriyani[4], Nurul Rizky Chairunnisa[4], Asyifa Charmadya Putri[4]
[1]Department of Mathematics, Mulawarman University, Samarinda, Indonesia
[2]Department of Statistics, Sepuluh Nopember Institute of Technology, Indonesia
[3]College of Arts and Sciences, University of Pennsylvania, Pennsylvania, United States
[4]Laboratory of Applied Statistics, Mulawarman University, Indonesia
sifriyani@fmipa.unmul.ac.id

Article History:
Received : 30-10-2023
Revised : 22-04-2024
Accepted : 23-04-2024
Online : 17-07-2024

Keywords:
K Medoids Algorithm;
Data Mining;
Food Security Data;
Silhouette Coefficient.

ABSTRACT

This study proposes a cluster analysis of provinces based on national food security data. The research objective is to determine provincial clusters based on food indicators which include rice harvest area, distribution of rice stocks, percentage of trade margin and transportation of rice distribution, percentage of average per capita expenditure, and total per capita consumption of rice. The source of observation data for the Rice Harvested Area by Province variable is the Ministry of Agriculture, Central Bureau of Statistics and Agriculture Services throughout Indonesia. This study uses data mining techniques in data processing with the K-Medoids algorithm. The K-Medoids method is a clustering method that functions to break down data sets into several groups. The advantage of this method is that it can overcome the weakness of the K-Means method which is sensitive to outliers. Another advantage of this algorithm is that the results of the clustering process do not depend on the order in which the dataset is entered. The k-medoids clustering method can be applied to food security data by province. From grouping the data obtained three clusters, with silhouette coefficient values for cluster 1, cluster 2, and cluster 3 respectively 0.33; 0.32; and 0.44. With the largest silhouette coefficient value obtained in cluster 3 and the cluster has entered into a strong cluster structure. The research results can provide information to the government about food security grouping data in Indonesia which has an impact on the distribution and availability of food in Indonesia.

Crossref
https://doi.org/10.31764/jtam.v8i3.19912

——————————— ◆ ———————————

A. INTRODUCTION

Food security in Indonesia is an issue that is often discussed, because of the large population in Indonesia and the country's rapid economic growth. This study discusses food security in Indonesia to find out how the current index of food security in Indonesia is described by looking at the grouping in each province that has the same characteristics. Data mining is the process of automatically finding useful information in large data repositories (Tan et al., 2019), (Arora & Varshney, 2016). Clustering is known as a segmentation technique which helps in dividing large data sets. This technique groups past data. Clustering is also part of artificial intelligence techniques (Maheshwari, 2015), (Chen et al., 2023). The purpose of clustering is to determine intrinsic grouping in a data set that is not labeled depending on some measure of similarity, for example Euclidean distance (Shukur & Alrashid, 2014), (Chen et al., 2022). The clustering method used in this research is K-Medoids. K-Medoids is a clustering algorithm that aims to minimize the absolute error criterion of Sum of Squared Errors (SSE). Similar to the K-means clustering algorithm, the K-Medoids algorithm runs iteratively until each representative object is actually a medoid of the cluster (Aggarwal & Reddyck, 2014), (Harikumar & Pv, 2015).

Several previous studies used cluster algorithms, such as the use of a clustering system to facilitate the process of identifying halal food menu ingredients (Sucipto et al., 2021); research processing COVID-19 pandemic data through data mining techniques using K-Means clustering with the help of three software, namely KNME, Weka, and Microsoft Excel to classify areas based on the number of infected and dead (Indraputra & Fitriana, 2020). Grouping data on areas infected with COVID-19 with the best grouping was carried out using the K-Medoids algorithm as many as 3 clusters using the Rapid Miner software (Sindi et al., 2020); comparative analysis of K-Means with K-Medoids of the two algorithms in different data sets to explain the strengths and weaknesses of both (Arbin et al., 2015); evaluation of the performance of the basic K-Means algorithm was carried out using various distance metrics (Thakare & Bagal, 2015); increase in clustering results based on the Davies Bouldin Index in determining the initial centroid in the K-Means algorithm (Sitompul et al., 2019).

In this study, the K-Medoids method was used to classify each province and differentiate it from other studies. The research results are distributed in spatial mapping to determine the national food security clusters. The variables used are paddy harvested area, distribution of rice stocks per regional office in Bulog, percentage of trade and transport margins as well as main rice distribution chains, percentage of average per capita expenditure per month for food in rural and urban areas according to total rice consumption per capita per year. The importance of achieving sustainable food security and ensuring that all regions have adequate access to food in sufficient quantities, nutritious, of quality in a sustainable manner, and easy to reach.

B. METHODS

1. Data Mining

Data mining is a method for finding certain patterns from large data sets. Data mining can be interpreted as knowledge mining from large data (Chikohora, 2014). Data mining is an amalgamation of several scientific disciplines such as computers which are defined as the process of discovering new patterns from very large data sets, including cutting methods from artificial intelligence, machine learning, statistics, and database systems (Khatami et al., 2017). In general, the uses of data mining can be divided into two: descriptive and predictive. Descriptive means data mining is used to look for patterns that can be understood by humans that explain the characteristics of the data. While predictive means data mining is used to form a knowledge model to make predictions. Based on their functionality, data mining tasks can be grouped into the following six groups (Shen et al., 2023): (a) Classification is to generalize known structures to be applied to new data; (b) Clustering is grouping data, whose class label is unknown, into a certain number of groups according to the size of their similarity; (c) Regression is finding a function that models data with the minimum error (prediction error); (d) Anomaly detection is identifying data that is not common, can be in the form of outliers, changes or deviations that may be very important and need further investigation; (e)

Association rule learning is dependency modeling: looking for relations I or between variables; and (f) Summarization is to provide a simpler data representation, including visualization and report generation. The analysis used in this research is clustering.

2. Clustering and Distance Measures

Clustering is the process of grouping objects based on information obtained from data that describes the relationship between objects with the principle of maximizing the similarities between members of one class and minimizing the similarities between clusters (Pinheiro et al., 2020). Euclidean distance calculates the distance of two points by knowing the value of each attribute at these two points. Euclidean distance is calculated using equation (1).

$$Distance(p,q) = \left(\sum\nolimits_{k}^{n} \mu k|Pk - qr|r\right)\frac{1}{r} \tag{1}$$

Distance is a commonly used approach to determine the similarity or dissimilarity of two feature vectors which is expressed by ranking. If the resulting ranking value is smaller, then the closer the similarity between the two vectors. Distance measurement technique with the Euclidean method is one of the most used methods. Distance measurement using the Euclidean method is written in equation (2).

$$j(v_1, v_2) = \sqrt{\sum_{k=1}^{N} \left(v_1(k) - v_2(k)\right)^2} \tag{2}$$

where v_1 and v_2 are two vectors whose distance will be calculated and N denotes the length of the vector (Irwansyah, 2015).

3. K-Medoids Clustering

K-Medoids is a group of partitional clustering methods that minimizes the distance between labeled points in the cluster and the cluster center point (Sun et al., 2018). The K-Medoids algorithm is also known as the Partitioning Around Mendoids (PAM) algorithm. The K-Medoids algorithm uses a representative object (medoid) as the cluster center for each cluster (Kaur, 2014). K-Medoid is a classic clustering partitioning technique that groups a dataset of ni objects into k groups known by Apriori. The K-Medoids algorithm has the advantage of overcoming the weaknesses of the K-Means algorithm which are sensitive to noise and outliers, where objects with large values may deviate from the data distribution. Another advantage is that the results of the clustering process do not depend on the order in which the dataset is entered (Pramesti, 2017). The K-Medoids algorithm (Bhat, 2014), (Sureja et al., 2022) is carried out in the following steps: (a) Initialize k cluster centers (number of clusters); (b) Calculate each object to the closest cluster using the Euclidian Distance measure equation; (c) After calculating the Euclidian Distance, initialize a new cluster center randomly on each object as a non medoid candidate; (d) Calculate the distance of each object in each cluster with non-medoids candidates; (e) Calculate the total deviation (S) by calculating the new total distance –

the old total distance. If S < 0 then swap objects with non-medoids data clusters to form a new set of k objects as medoids; and (f) Repeat steps c to e until there is no change in the medoid, so that the clusters and their respective cluster members are obtained.

4. Cluster Goodness Measure with Silhoutte Coefficient Method

Silhouette Coefficient (SC) is a method used to see the quality and strength of clusters. The Silhouette Coefficient method is a combined method of the cohesion method and the separation method. The cohesion method is a measure of how close the relationships are between objects in a cluster, while the separation method is a measure of how far or apart a cluster is from other clusters (Rendy, 2014). The steps for calculating the Silhouette Coefficient include:

a. Calculation of the average distance from an object, for example, the i-th object with all other objects in one cluster. Calculation of the average distance according to Equation (3).

$$a(i) = \frac{1}{[A]} \sum j \in_{A,j \neq i} d(i,j) \tag{3}$$

b. The calculation of Equation (4) is the average distance from the i-th object to all objects in the other clusters, then taking the smallest value.

$$d(i,C) = \frac{1}{[A]} \sum j \in C \; d(i,j) \tag{4}$$

c. The Silhouette Coefficient value calculation uses Equation (5). The sum of *s(i)* is obtained by combining *a(i)* and *b(i).*

$$SC = s \begin{cases} 1 - \dfrac{a(i)}{b(i)} & if \; a(i) < b(i) \\ 0 & if \; a(i) = b(i) \\ \dfrac{a(i)}{b(i)} - 1 & if \; a(i) > b(i) \end{cases} \tag{5}$$

$$SC(i) = \frac{b(i) - a(i)}{\max\big(a(i), b(i)\big)}$$

The calculated value using the Silhouette Coefficient method lies in the range between -1 to 1. The average value of the Silhouette Coefficient of each object in a cluster is a measure that shows how closely the data are grouped in one cluster. The closer the average value of the Silhouette Coefficient is to 1, the better the grouping of data in one cluster. Conversely, if the average value of the Silhouette Coefficient is close to -1, the worse the grouping of data in one cluster. The criteria for measuring the value of the Silhouette Coefficient (Wira et al., 2019), (Yu et al., 2018), (Zhang & Shang, 2022) are shown in Table 1 below.

Table 1. Silhouette Coefficient Value Size

Silhouette Coefficient	Proposed interpretation
0.7 < SC ≤ 1.0	Stronge Structure
0.5 < SC ≤ 0.7	Medium Structure
0.25 < SC ≤ 0.5	Weak Structure
SC ≤ 0.25	No Structure

5. Research Methods and Data Analysis Stages

This study uses Food Security data with observation units of 34 provinces from 2020 to 2022. The initial stage of determining the dataset, then pre-processing the data. The next stage is the data cleaning process to remove data that is less relevant or less valid so that it becomes relevant data. Then proceed with data transformation before applying the K-Medoid algorithm. Data analysis using R-Studio software. The stages of data analysis for the K-Medoids cluster method are described as follows: (a) Transforming data for the K-Medoids calculation process; (b) Determining the cluster center then calculate the distance of the data (object) to the cluster center using Euclidean Distance; (c) Calculating the total distance of all data in the cluster; (d) Determining the new cluster center randomly and then calculate the data distance from the cluster center using Euclidean Distance; (e) Determining the difference between the total distance by subtracting the new total distance from the old total distance; (f) Obtaining the final cluster results; and (g) Measuring the goodness of clusters with the silhouette coefficient method.

6. Variables and Research Data Sources

The source of observation data for the Rice Harvested Area by Province variable is the Ministry of Agriculture, Central Bureau of Statistics and Agriculture Services throughout Indonesia (Pusat Data Informasi Pertanian Sekretariat Jenderal, 2022), (Pusat Data dan Sistem Informasi Pertanian Sekretariat Jenderal Kementerian Pertanian, 2022). Variables and research data sources are shown in Table 2.

Table 2. Variables and Data Sources

Variable	Source of observation data	Unit	Scale
Rice Harvested Area (x_1)	Ministry of Agriculture, Central Bureau of Statistics and Agriculture Services throughout Indonesia (Pusat Data Informasi Pertanian Sekretariat Jenderal, 2022), (Pusat Data dan Sistem Informasi Pertanian Sekretariat Jenderal Kementerian Pertanian, 2022)	Ha (Hectare)	34 Provinces in Indonesia
Distribution of End of Month Rice Stocks per Regional Office in Bulog (x_2)	Perum Bulog is processed by Ministry of Data and Information Center (Badan Pusat Statistik, 2022)	Ton	34 Provinces in Indonesia
Percentage of Trade and Transportation Margins and Main Chains of Rice Distribution (x_3)	Central Bureau of Statistics (Badan Pusat Statistik, 2022)	% (Percent)	34 Provinces in Indonesia

Variable	Source of observation data	Unit	Scale
Percentage of Average Per Capita Monthly Expenditures for Food in Rural and Urban Areas by Province in 2021 (x_4)	National Socioeconomic Survey, Central Bureau of Statistics (Badan Pusat Statistik, 2022)	% (Percent)	34 Provinces in Indonesia
Total Consumption of Rice per Capita per Year (x_5)	National Socioeconomic Survey, Central Bureau of Statistics processed by the Ministry of Data and Information Center (Badan Pusat Statistik, 2022)	Kg (Kilogram)	34 Provinces in Indonesia

C. RESULT AND DISCUSSION

1. Characteristics of Research Data and Descriptive Statistics

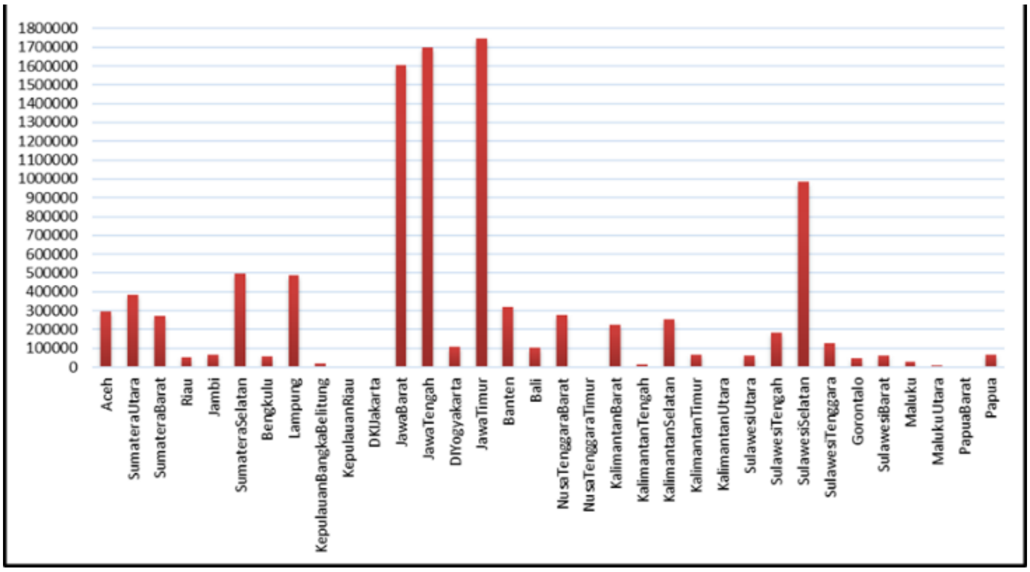

Figure 1. Rice Harvested Area by Province in 2021

2. Descriptive Statistics

Table 3. Descriptive Statistics

No.	Variables	Average	Min	Max
1	Rice Harvested Area by Province Year (x_1)	297804.9	270	1747481
2	Distribution of End of Month Rice Stocks per Regional Office in Bulog (x_2)	34512.11	1	223834.7
3	Percentage of Trade and Transportation Margins and Main Chains of Rice Distribution (x_3)	17.04059	6.09	27.12
4	Percentage of Average Per Capita Monthly Expenditures for Food in Rural and Urban Areas by Province in 2021 (x_4)	49.63382	39.54	57.94
5	Total Consumption of Rice per Capita per Year (x_5)	94.63265	59.96	118.85

Based on Table 3, the average data value of rice harvest area in Indonesia is 297804.9 Ha, the lowest rice harvest area is 270 Ha located in the Riau Islands province, and the highest rice harvest area is 1,747,481 Ha located in East Java Province.

3. K-Medoids Clustering

The results of data mining implementation research using the K-Medoids Cluster method are discussed in this chapter. The initial stage in the cluster process is determining the cluster center for each cluster. The cluster center is determined to be in the 7th, 32nd, and 12th data, namely the provinces of Bengkulu, West Papua, and West Java. The cluster center results are given in Table 4.

Table 4. K-Medoids Cluster Center for 34 Provinces in Indonesia

Medoids	ID	x_1	x_2	x_3	x_4	x_5
Bengkulu	7	-0.5027813	-0.5884503	-0.6821450	0.33123649	0.4425978
West Papua	32	-0.6051444	-0.3748402	1.4811451	-0.34813256	-0.9415500
West Java	12	2.7128691	2.9594480	0.1574268	-0.07089004	0.1667961

Table 5. Numerical Information per Cluster

Cluster	Size	Maximal Distance	Average Distance
1	20	2.572374	1.383486
2	10	3.162857	1.298141
3	4	2.183877	1.328783

The results of calculating the maximum distance and average distance are given in Table 5, obtained for the first cluster there are 20 data, for the second cluster there are 10 data, and for the third cluster there are 4 data. The maximum distance for each cluster is 2.572374; 3.162857; and 2.1838877.

4. Cluster Goodness Measure with Silhoutte Coefficient

The evaluation method used to see the quality and strength of clusters is the Silhouette Coefficient method. The results of the average value of the silhouette method of each cluster are given in Table 6.

Table 6. Cluster Goodness Measure

Cluster	Average
1	0.33
2	0.32
3	0.44

5. Cluster Results of 34 Provinces in Indonesia

K-Medoids cluster results from 34 provinces in Indonesia are given in Table 7.

Table 7. Cluster Results of 34 Provinces

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

The plot of the cluster results is shown in Figure 3. The red color indicates the province in cluster 1, the green color indicates the province in cluster 2, and the blue color indicates the province in cluster 3.

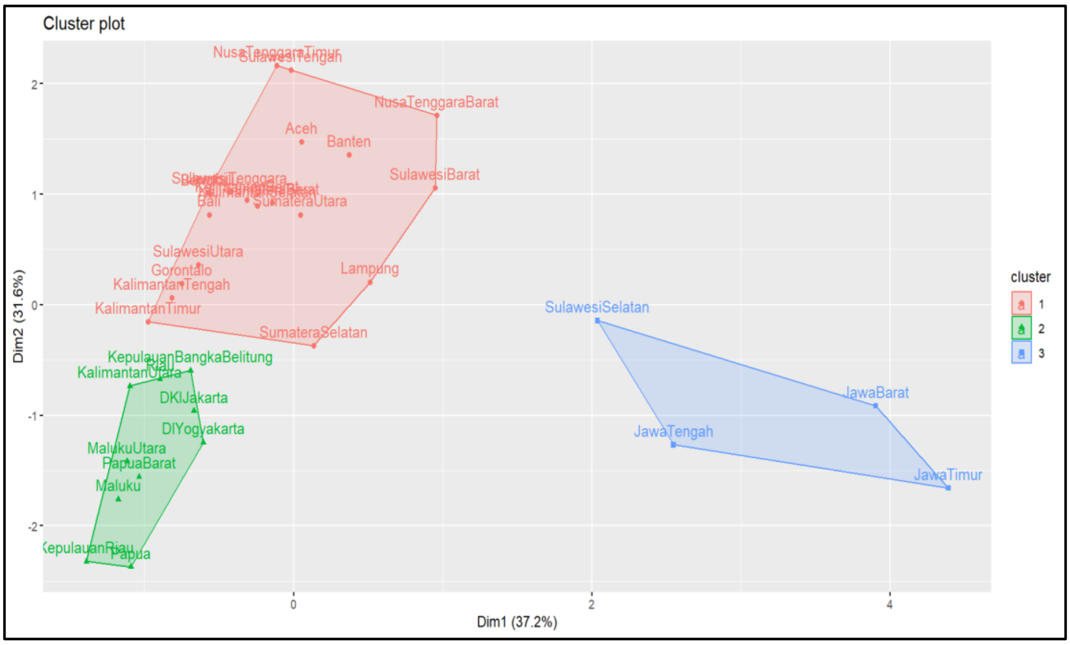

Figure 2. K-Medoids Plot of Cluster Results from 34 Provinces in Indonesia

6. Spatial Mapping Distribution of 34 Provinces Based on Cluster Results

The distribution map of the cluster results that have been obtained for 34 provinces in Indonesia is shown in Figure 3, where the red color indicates cluster 1, the green color indicates cluster 2, and the blue color indicates cluster 3.

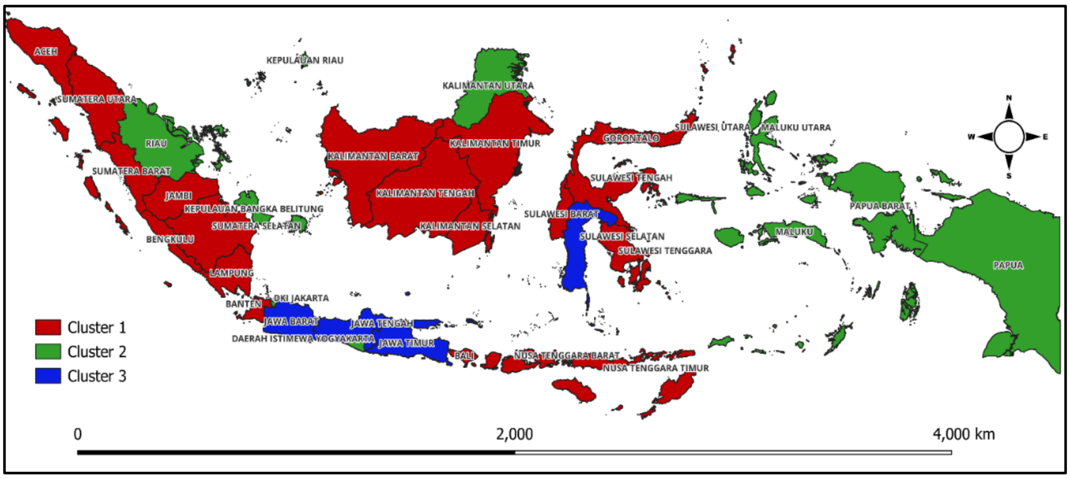

Figure 3. Food Spatial Mapping of 34 Provinces Based on Cluster Results

7. Interpretation of Cluster Results

The interpretation of the K-Medoids cluster results is to look at the characteristics of each cluster. The average variable value of each cluster shows this characteristic. The results of calculating cluster characteristics are shown in Table 8.

Table 8. Characteristics of Cluster Results

Cluster	x_1	x_2	x_3	x_4	x_5
1	189792	20371	13.9	50.6	101
2	29606	15480	22.7	47.9	82.9
3	1508365	152798	18.7	49.3	94.6

Table 8 shows that the provinces included in cluster 1 are provinces with higher food security compared to cluster 2. Where cluster 2 has the lowest food security between clusters 1 and 3. Meanwhile, cluster 3 has the highest food security among clusters 1 and cluster 2.

D. CONCLUSION AND SUGGESTIONS

The grouping using the K medoids method for the characteristics of food security in Indonesia was carried out with 3 clusters. Of the 34 units of observation, 20 data were found in cluster 1, 10 data were in cluster 2, and 4 data were in cluster 3. The silhouette coefficient values for cluster 1, cluster 2, and cluster 3 were 0.3331810; 0.3293861; and 0.4493752. The greatest silhouette coefficient value is obtained in cluster 3 and the cluster is already included in a strong cluster structure. Future research can be conducted using different variables such as the population of livestock commodities, agricultural commodity production, food consumption participation of the Indonesian population, the number of poor people and poverty lines and the development of inflation and food commodity prices using different methods such as K-Means.

ACKNOWLEDGEMENT

The title for the thank you to the institution or the person who has contributed during the research and references is not numbered.

REFERENCES

Aggarwal, C.C. & Reddyck. (2014). *Data Clustering Algoritms and Applications*. CRC Press. https://doi.org/10.1201/9781315373515

Arbin, N., N.S. Suhaimi, N.Z. Mokhtar & Z. Othman. (2015). Comparative Analysis between K-Means and K-Medoids for Statistical Clustering. *2015 Third International Conference on Artificial Intelligence, Modelling and Simulation*, 116–121. doi: https://doi.org/10.1109/AIMS.2015.82

Arora, P. & Varshney, S. (2016). Analysis of K-Means and K-Medoids Algorithm For Big Data. *Procedia - Procedia Comput. Sci.* 78:507–512. https://doi:10.1016/j.procs.2016.02.095

Bhat, A. (2014). K-Medoids Clustering Using Partitioning Around Medoids for Performing Face Recognition. *International Journal of Soft Computing, Mathematics and Control (IJSCMC)*, 3–4. https://doi.org/10.14810/IJSCMC.2014.3301

Badan Pusat Statistik, Statistik Indonesia - Statistical Yearbook of Indonesia 2022, Badan Pusat Statistik BPS-Statistics Indonesia, 2022 https://www.bps.go.id/id/publication/2022/02/25/0a2afea4fab72a5d052cb315/statistik-indonesia-2022.html

Chen, Y., H. Li, H. Lin, Y. Wang & Y. Zhao. (2023). Computers and Geotechnics Critical Slip Line Recognition and Extraction Method of Slope Based on Modified K-Medoid Clustering Algorithm. *Comput. Geotech* 15 (105125). doi:https://doi:10.1016/j.compgeo.2022.105125.

Chen, Z, H. Li, H. Lin, Y. Wang, and Y. Zhao. (2022). A New Parallel Adaptive Structural Reliability Analysis Method Based on Importance Sampling and K-Medoids Clustering. *Reliab. Eng. Syst. Saf.* 218 (108124). doi: https://doi:10.1016/j.ress.2021.108124.

Chikohora, T.T. (2014). A Study of the Factors Considired When Choosing an Appropriate Data Mining Algorithm. *International Journal of Soft Computing and Engineering* 4 (3):42–45. https://www.ijsce.org/wp-content/uploads/papers/v4i3/C2270074314.pdf

E. Irwansyah, M. F. (2015). *Advanced Clustering Teori Dan Aplikasi*. Yogyakarta: Deepublish. https://balaiyanpus.jogjaprov.go.id/opac/detail-opac?id=322626

Harikumar, S. & Pv, S. (2015). K-Medoid Clustering for Heterogeneous DataSets, 70:226–237. doi:https://doi:10.1016/j.procs.2015.10.077.

Indraputra, R.A. & Fitriana, R. (2020). K-Means Clustering Data COVID-19. *Jurnal Teknik Industri* 10, (3):275–282. doi: https://doi.org/10.25105/jti.v10i3.8428.

Kaur, K.N. (2014). K-Medoid Clustering Algorithm- A Review. *International Journal of Computer Application and Technology* 14. https://www.academia.edu/8446443/K_Medoid_Clustering_Algorithm_A_Review

Khatami, A., S. Mirghasemi, A. Khosravi, C. Peng & S. Nahavandi. (2017). A New PSO-Based Approach to Fire Flame Detection Using K-Medoids Clustering. *Expert Syst. Appl,* 68:69–80. doi:https://doi:10.1016/j.eswa.2016.09.021.

Maheshwari, A. (2015). *Data Analytics Mode Accessible*. https://dokumen.pub/data-analytics-made-accessible-t-8692526.html

Pinheiro, D. N., D. Aloise & S. J. Blanchard. (2020). Convex Fuzzy k -Medoids Clustering. *Fuzzy Sets Syst,* 389:66–92. doi:https://doi:10.1016/j.fss.2020.01.001.

Pramesti, F.D. (2017). Implementasi Metode K-Medoids Clustering Untuk Pengelompokkan Data Potensi Kebakaran Hutan/Lahan Berdasarkan Persebaran Titik Panas (Hotspot). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 723–732. https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/204

Pusat Data Informasi Pertanian Sekretariat Jenderal, Statistik Ketahanan PanganTahun 2022, (2022) 138–138. https://satudata.pertanian.go.id/assets/docs/publikasi/Statistik_Ketahanan_Pangan_2022.pdf

Pusat Data dan Sistem Informasi Pertanian Sekretariat Jenderal Kementerian Pertanian, Statistik Ketahanan Pangan Tahun 2022 Volume 2 Nomor 1, Pusat Data dan Sistem Informasi Pertanian Sekretariat Jenderal Kementerian Pertanian, 2022. https://satudata.pertanian.go.id/assets/docs/publikasi/Statistik_Ketahanan_Pangan_2022.pdf

Rendy, H.D. (2014). Perbandingan Metode Clustering Menggunakan Metode Single Linkage Dan K-Means Pada Pengelompokkan Dokumen. *JSM STMIK Mikroskil*, 76–77. https://media.neliti.com/media/publications/280965-perbandingan-metode-clustering-menggunak-bf005dbb.pdf

Shen, Y., D. Zhang, R. Wang, J. Li & Z. Huang. (2023). Transportation Geotechnics SBD-K-Medoids-Based Long-Term Settlement Analysis of Shield Tunnel. *Transp. Geotech,* 42(101053). doi:https://doi:10.1016/j.trgeo.2023.101053.

Shukur, B.K. & S.Z.N. Alrashid. (2014). Evaluation of Clustering Image Using Steady State Genetic and Hybrid K-Harmonic Clustering Algorithms. *IJCCCE,* 14 (1):166–173. https://www.iasj.net/iasj/download/4f003c95c33607ee

Sindi, S., W.R.O. Ningse, I.A. Sihombing, F.I. RHzer & D. Hartama. (2020). Analisis Algoritma K-Medoids Clustering Dalam Pengelompokan Penyebaran COVID-19 Di Indonesia. *Jurnal Teknologi Informasi,* 4 (1):166–173. doi:https://doi.org/10.36294/jurti.v4i1.1296.

Sitompul, B.J.D., O.S. Sitompul & P. Sihombing. (2019). Enhancement Clustering Evaluation Result of Davies-Bouldin Index with Determining Initial Centroid of K-Means Algorithm. *IOP Conf. Series: Journal of Physics,* 1235(1):6–12. doi:https://doi.org/10.1088/1742-6596/1235/1/012015.

Sucipto, S., M.S. Al-Mubarok & D.T. Setiyawan. (2021). Hierarchical Clustering Bahan Menu Di Kantin Universitas Untuk Menunjang Implementasi Sistem Jaminan Halal. *Jurnal Teknologi Industri Pertanian,* 31(1):103–124. doi:https://doi.org/10.1201/b19706.

Sun, D., H. Fei & Q. Li. (2018). A Bisecting K-Medoids Clustering Algorithm Based on Cloud Model, 308–315. doi:https://doi:10.1016/j.ifacol.2018.08.301.

Sureja, N., B. Chawda & A. Vasant. (2022). An Improved K -Medoids Clustering Approach Based on the Crow Search Algorithm. *J. Comput. Math. Data Sci.,* 3(100034). doi:https://doi:10.1016/j.jcmds.2022.100034.

Tan, P.N., Steinbach, M. & Karpatne, A. (2019). *Introduction to Data Mining Second Edition*. Pearson Education Limited. https://link.springer.com/chapter/10.1007/978-1-4302-3325-1_14

Thakare, Y.S. & S.B. Bagal. (2015). Performance Evaluation of K-Means Clustering Algorithm with Various Distance Metrics. *International Journal of Computer Applications,* 110(11):12–16. doi:https://doi.org/10.5120/19360-0929.

Wira, B., Budianto, E. Alexius, Wiguna & S. Anggri. (2019). Implementasi Metode K-Medoids Clustering Untuk Mengetahui Pola Pemilihan Program Stusi Mahasiswa Baru Tahun 2018 Di Universitas Kanjuruhan Malang. *Jurnal Terapan Sains Dan Teknologi*, 55–57. https://doi.org/10.21067/jtst.v1i3.3046

Yu, D., G. Liu, M. Guo & X. Liu. (2018). An Improve d K-Medoids Algorithm Based on Step Increasing and Optimizing Medoids, 92:464–473. doi:https://doi:10.1016/j.eswa.2017.09.052.

Zhang, Y. & P. Shang. (2022). Communications in Nonlinear Science and Numerical Simulation KM-MIC : An Improved Maximum Information Coefficient Based on K-Medoids Clustering. *Commun. Nonlinear Sci. Nu Mer. Simul*, 106418:111. doi:https://doi:10.1016/j.cnsns.2022.106418.