JOURNAL LA MULTIAPP

VOL. 06, ISSUE 05 (1250-1265), 2025
DOI: 10.37899/journallamultiapp.v6i5.2411

Livestock Population Map Based on Provinces in Indonesia Using the K-Medoids Method

Riri Qorib Nurhakim[1], Wina Witanti[1], Agus Komarudin[1]

[1]*Faculty of Science and Informatics, Informatics Engineering Study Program, Jenderal Achmad Yani University, Cimahi, Indonesia*

**Corresponding Author: Riri Qorib Nurhakim*
Email: riri.qorib@student.unjani.ac.id

Article Info

Article history:
Received 5 July 2025
Received in revised form 8 August 2025
Accepted 29 September 2025

Keywords:
Clustering
Davies Bouldin Index
K-Medoids
Livestock Population Map
Livestock

Abstract

Indonesia is one of the countries with a large livestock population. A healthy and stable livestock population can affect the production and availability of livestock products, such as meat, milk, eggs, and skin. FAO's Domestic Animal Diversity - Information System (DAD-IS) data (2020) recorded around 206 large farms, small farms, poultry and pigs. Clustering is a technique for grouping data without unknown class labels. Clustering is used to find data that has similarities. The clustering technique is to determine the initial cluster center. This study is intended to determine the best cluster value using the selected method. The purpose of this study is to create a system that can process and group data. With data obtained from the central statistics agency. This study uses the topic of Livestock Population Map in Indonesia using K-Medoids. The algorithm used in this study is K-Medoids. The K-Medoids method is a variation of the K-Means method to retrieve k data, the number of clusters in a data set with n objects. There are several processes carried out in this study including collecting data, then entering the preprocessing stage, grouping data that has similarities between data. After clustering using K-Medoids, it was found that Cluster 0 had 3 provinces with the highest average population with types of livestock such as Dairy cattle, Beef cattle, Sheep and Goats, Cluster 1 had 29 provinces with the lowest average population, Cluster 2 had 2 provinces with the highest average number for types of livestock such as Buffalo, Horse and Pig.

Introduction

Indonesia is one of the countries with a large livestock population. Livestock plays an important role in agriculture, animal husbandry, economy, and public health (Budiman et al., 2024; Lubis et al., 2024; Kamel et al., 2025). Livestock makes a significant contribution to the Indonesian economy, as it is a source of income for many livestock breeders and farmers. Livestock is also a source of essential protein for the Indonesian people. A healthy and stable livestock population can affect the production and availability of livestock products, such as meat, milk, eggs, and skin. FAO's Domestic Animal Diversity - Information System (DAD-IS) data (2020) recorded around 206 large farms, small farms, poultry and pigs (Suharno, 2020). The animals raised vary such as dairy cattle, beef cattle, buffalo, horses, goats, sheep and pigs. Most of the animals raised have adapted well to their rearing environment (BPS Indonesia, 2022). Clustering is a part of unsupervised learning that is used as an effective data mining. Clustering methods can reveal previously unclear relationships and structures from a data set. Because clustering is unlabeled, all attributes are considered the same. Clustering aims to group data that have the same characteristics into the same group (Mustofa & Iman

ISSN: 2716-3865 (Print), 2721-1290 (Online)

Saufik Atmosfer, 2020; Zhang & Peng, 2024; Chen et al., 2024). Data visualization is a system that actively restructures, sketches, authors and personalizes visuals (Dimara & Perin, 2020; Hallmen et al., 2025). Clustering tries to collect similar data into only one group and that data should not appear in other groups. Data grouping is a difficult job depending on the form of the data. Therefore, in the data grouping that will be carried out, the selection of the clustering method must be right. One technique that can be done in data grouping is the K-Medoids method. The K-Medoids method aims to minimize the shortcomings of outliers in the data set that often occur in the K-Means method. The K-Medoids method is also called a partition method that combines n objects into k clusters (Abbas et al., 2020; Khotimah et al., 2024; Nurdin et al., 2025). K-Medoids is a variation of the K-Means method, where data points are selected as medoids instead of selecting the mean as the centroid in the K-Means algorithm. Medoids can be thought of as objects in a group that have a minimum average difference with other objects from other clusters (Alfiah et al., 2021; Kariyam et al., 2024; Alfitra et al., 2025).

Several previous studies have proven that the K-medoids method can be used to determine clusters for each district and city in West Java based on the characteristics of the livestock population (Khoirunisa et al., 2023; Aqilah et al., 2025; Azzahro et al., 2024). In addition, another study, namely the second study, found that the K-medoids method obtained an optimal cluster of 2 clusters. There are 30 districts/cities in cluster 1 and 8 districts/cities in cluster 2. Cluster 1 has the characteristics of a high Percentage of Households with Proper Sanitation, AHH, and Percentage of Literacy Rates Aged 15-55 Years. While cluster 2 has the characteristics of a high Percentage of Poor Households Receiving Raskin, Percentage of Poor Population, Percentage of Per Capita Expenditure on Food with Poor Status (Alfiah et al., 2021). In the third study, the average number of small and micro businesses in 2017-2019 was the attribute used. The number of clusters is determined by the Davies Bouldin Index (DBI) method, where k = 2 is the best value (0.111). According to the cluster results (k = 2), the labels are divided into two (high cluster (clt1) and low group) (clt2). The results of the k-medoids calculation show that 90% of Indonesia's area is in the low class, with the case of "Application of the clustering algorithm to the small and micro industrial companies for mapping regions with k-medoids" (Jatnika et al., 2021). For the fourth study by conducting cluster mapping with a comparison of the number of students and teachers in Indonesia by region, especially at the elementary school level, using the Davies Bouldin Index (DBI) and Clustering parameters. By using three cluster labels, namely high cluster (K1), normal cluster (K2) and poor cluster (K3), 3 provinces were found to be in the high cluster, 9 provinces were in the normal cluster and 22 provinces had fewer clusters. By testing the cluster results (k = 3) through the DBI parameter, the value = 0.587 was obtained, with the case of "Unsupervised Data Mining with K-Medoids Method in Mapping Areas of Student and Teacher Ratio in Indonesia" (Hermawati et al., 2020; Putri, 2025; Hidayat et al., 2025). And the fifth study applied the K-medoids Clustering method to group provinces based on the characteristics of their broiler chicken population. The results of the analysis showed that there were 2 clusters, namely cluster 0 containing 31 provinces while cluster 1 contained 3 provinces, with a Davies Bouldin value of 0.011 (Ainisa et al., 2024; Tajbar et al., 2024; Saleh et al., 2025). From previous research that has been done, the update in this study lies in how to convey information about data distribution patterns using the K-medoids method which is carried out by visualizing specifically for the distribution of livestock population distribution patterns in each province in Indonesia using the elbow method in determining the optimal number of clusters. The purpose of this study is to provide an overview of distribution patterns in supporting optimization of progress, especially in the livestock sector.

ISSN: 2716-3865 (Print), 2721-1290 (Online)

Methods

Research Stages

This research went through several stages such as data collection, then entering the preprocessing stage, then the clustering stage and finally visualization of the clustering results.

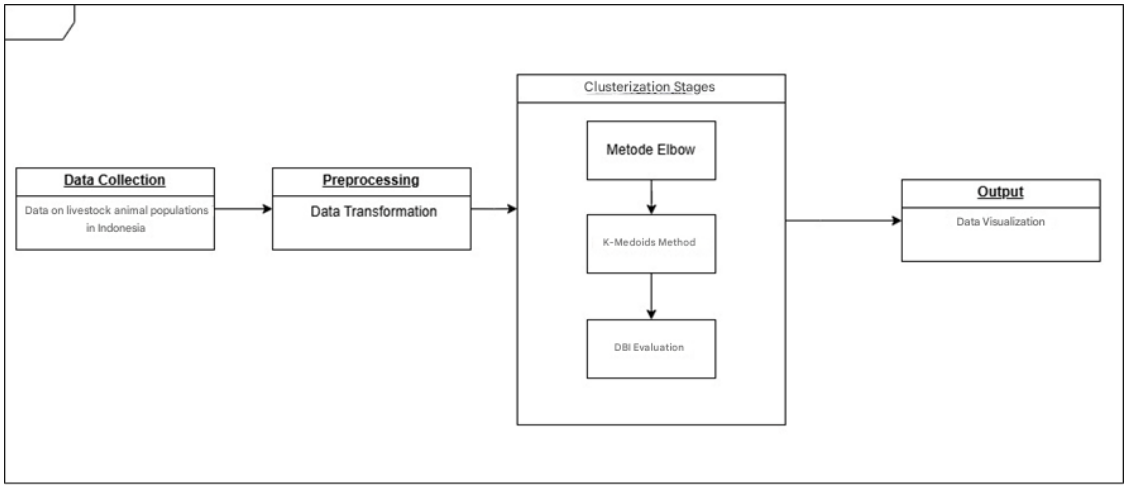

Figure 1. Research Stages

Data collection

In this study, the dataset used is the number of livestock population in each province in Indonesia, data taken from the Central Statistics Agency in a span of 5 years starting from 2018 to 2022 with a total of 34 provinces. The data obtained are in the form of Province Name, Type of Livestock, Year and Population.

Pre-Processing

The pre-processing stage is carried out on research data to ensure the quality of the data used. This process aims to handle incomplete data, eliminate noise, and correct inconsistent data (Kenger et al., 2023).

Data Transformation

To process the data to be analyzed, data transformation is needed. The goal is to change the selected data into data that is more suitable for carrying out the data mining process. One technique that is often used in data transformation is Z-Score Normalization (Mathematics & Generation, 2023). This technique changes data into a standard form by calculating the average value and standard deviation of the data. Z-Score normalization can be calculated using the available formula (Kenger et al., 2023), as follows:

$$Z = \frac{x - \mu}{\sigma}$$

Information:

x = observed value (raw score)

μ = population average

σ = population standard deviation

ISSN: 2716-3865 (Print), 2721-1290 (Online)

Z = Z-score (standard value)

Elbow Method

The elbow method is a technique for determining the number of clusters by observing the point on the graph where there is a sharp decline, forming a clear curve. This point indicates the optimal k value or number of clusters. To find the k value, it can be done by comparing the Sum of Squared Errors (SSE) values displayed in the form of a graph (Syahfitri et al., 2023). The purpose of the elbow method is to determine the optimal k value by considering the within-cluster sum of squares (WSS) value which is low but still effective. The selection of the right number of clusters is done by analyzing the comparison between the SSE values in each number of clusters, where there will be a point where the SSE decline begins to slow down, forming an elbow pattern. The greater the number of clusters k, the smaller the resulting SSE value (FIRDAUS, 2022).

$$SSE = \sum_{k-1}^{k} \sum_{x1 \in k} \|xi - ck\|_2^2$$

Information :

SSE : Sum Of Square Error

Xi : Attribute value of data 1

Ck : attribute value of Cluster center point 1

K-Medoids Method Clustering

The clustering process in this study will use the K-medoids method. The K-Medoids algorithm is able to overcome the weaknesses of the K-Means algorithm, especially because K-Means tends to be sensitive to outliers and relies on the squared error function. One of the main reasons for developing a K-Medoids-based algorithm is its approach that uses real objects as a representation of each cluster. Each cluster is represented by one object that functions as a center, and other objects are grouped based on their similarity to the center. Unlike K-Means which uses cluster center points, K-Medoids is more resistant to the influence of outliers and extreme values (Barik & Centeno, 2020). One of the well-known K-Medoids-based algorithms is PAM (Partition Around Medoids). PAM works by accepting two input parameters, namely the number of clusters (k) and a data set consisting of n objects. This algorithm then forms k clusters based on the principle of minimizing the total difference between each object and its reference point. In the process, the algorithm randomly selects k objects from the dataset as initial medoids to be used as cluster representations (Barik & Centeno, 2020).

a. Input the data to be clustered.

b. Determine the number of clusters

c. Initialize the cluster center as much as k (number of clusters).

d. Allocate data (objects) to the nearest cluster using the euclidian distance measurement equation with the equation.

ISSN: 2716-3865 (Print), 2721-1290 (Online)

$$d(x,y) = \sqrt{\sum_{i-1}^{N} (x_i - y_i)^2}$$

e. Randomly select objects in each cluster to become new medoid candidates.

f. Calculating the range between objects contained in each cluster.

g. Calculate the deviation (S), namely the reduction of the total new distance and the total old distance. If S < 0, then swap objects using cluster data to form a collection of k objects as a new medoid.

Davies Bouldin Index

Davies Bouldin Index is a cluster validation method developed by D.L. Davies and Donald W. Bouldin, so the name of this method is taken from their names (Septhya et al., 2023). DBI is used to evaluate the quality of a cluster in the data grouping process. This method works by calculating the ratio between the level of data spread in a cluster (cohesion) and the level of separation between clusters. DBI measures cluster quality based on two main factors: cohesion and separation. Cohesion refers to the level of similarity of data in a cluster to the cluster center, where a smaller cohesion value indicates that the data in the cluster is more centralized. Meanwhile, separation measures the distance between different cluster centers, with a larger separation value indicating that the clusters are more separated from each other. A cluster is considered optimal if it has a low cohesion value and a high separation (K- et al., 2020).

Visualization

To build visualizations, remove distracting information to enrich visualizations with external knowledge, smoothly control both data presentation and pre-processing statistical function data. Data that has been processed makes it easier for users to visualize in real time. Data Visualization provides accessible data to understand a pattern in data. Data visualization uses processed data rather than raw data, so that someone can easily understand and know what it means (Peringkat et al., 2022). Shapefile is a simple format that does not use topology to store geometric location and attribute information from geographic features. Geographic features in shapefiles can be points, lines, or polygons (areas), and can be equipped with additional attributes associated with the feature (Baru & Di, 2019). Meanwhile, this shapefile data was obtained from the Indonesia Geospatial website, which includes the latest village and administrative boundaries for 2022/2023, including the expansion of 38 provinces with a scale of 10K from BIG (Lewis et al., 2023).

Results and Discussion

In accordance with the research method used, this section will explain in detail the stages that have been carried out in the research. This research is focused on the evaluation stage, where at that stage the final results of the research are obtained.

Data Attributes

Attribute description is an explanation of each attribute that will be used from the Livestock Population data in Indonesia. The attribute description is in table 1.

Table 1. Data Attributes

Attribute	Description
Province	Name of the province
Dairy Cattle	Livestock found on the farm
Beef Cattle	Livestock found on the farm

ISSN: 2716-3865 (Print), 2721-1290 (Online)

Sheep	Livestock found on the farm
Goat	Livestock found on the farm
Buffalo	Livestock found on the farm
Horse	Livestock found on the farm
Pig	Livestock found on the farm

Livestock Population Data

Table 2. Livestock Population Data

Province	Dairy Cattle	Beef Cattle	Sheep	Goat	Buffalo	Horse	Pig
Aceh	120	2,181,918	3,173,776	510,621	512,577	13,237	9,999
North Sumatra	22,816	4,639,538	3,776,673	3,694,642	472,326	5,818	2,865,177
West Sumatra	3,638	2,050,063	1,241,589	25,007	410,613	7,031	163,753
Riau	389	986,161	1,171,041	138,923	139,179	455	238,071
Jambi	104	799,889	2,111,058	356,797	210,902	794	11,994
South Sumatra	260	1,504,111	2,039,299	177,352	136,332	488	86,831
Bengkulu	1,255	773,626	1,036,956	46,845	126,016	228	12,512
…	…	…	…	…	…	…	…
Papua	43	348,140	345,407	616	3,882	13,322	4,537,710

The data was taken from the Central Statistics Agency of Indonesia with a span of 5 years starting from 2018 to 2022 in 34 provinces, with four-legged livestock, there are 7 types of animals, namely dairy cows, beef cattle, buffalo, horses, goats, sheep and pigs.

Z-score Transformation Data

The stages here process the original data into numeric data according to needs. The attributes used are province, dairy cattle, beef cattle, sheep, goats, buffalo, horses and pigs. The results can be seen in table 2 below. From Table 2 it can be seen that the numeric values are quite large, so data normalization must be carried out to facilitate the data mining process. Z-score normalization is one of the data normalization methods that is carried out using the average value and standard deviation of the data. This method has the advantage of being quite stable against outliers and is able to handle values that exceed the maximum limit (maxA) or are smaller than the minimum limit (minA). The formula for normalizing with Z-score is as follows:

$$Z = \frac{x - \mu}{\sigma}$$

$$Z = \frac{120 - 82471{,}9}{287354{,}9}$$

Z = -0,286

Normalization of Dairy Cow Attribute Data

Description:

x = observed value (raw score)

ISSN: 2716-3865 (Print), 2721-1290 (Online)

μ = population mean

σ = population standard deviation

Z = Z-score (raw score).

The value of x is the value of the number of livestock populations in the data to be calculated. For example, the number of dairy cattle used in the table is 120 and μ is the average value that has been calculated in the Dairy Cattle livestock column, which is 82471.9, the value of σ is the standard deviation that has been calculated, which is 287354.9. So that the value of the number of dairy cattle in one province with a total of 120 livestock produces a z-score value of -0.286. The following is done on each other attribute to get the results that can be seen in table 3.

Table 3. Z-score Transformation Data

Province	Dairy Cattle	Beef Cattle	Sheep	Goat	Buffalo	Horse	Pig
Aceh	-0.286	-0.080	0.099	-0.201	1.704	-0.243	-0.531
North Sumatra	-0.207	0.478	0.234	0.132	1.510	-0.285	0.826
West Sumatra	-0.274	-0.110	-0.335	-0.252	1.213	-0.278	-0.458
Riau	-0.285	-0.352	-0.350	-0.240	-0.092	-0.316	-0.423
Jambi	-0.286	-0.394	-0.139	-0.218	0.252	-0.314	-0.530
South Sumatra	-0.286	-0.234	-0.155	-0.236	-0.106	-0.316	-0.495
Bengkulu	-0.282	-0.400	-0.381	-0.250	-0.155	-0.317	-0.530
…	…	…	…	…	…	…	…
Papua	-0.286	-0.497	-0.536	-0.255	-0.743	-0.242	1.622

Outlier Checking

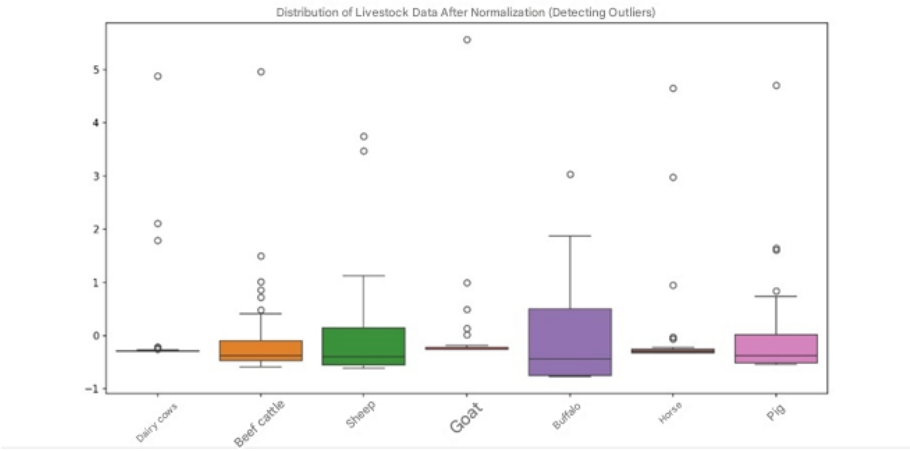

Figure 2. Boxplot Diagram

Based on Figure 2, there is a point that comes out of the boxplot which indicates that the data used contains outliers. So the cluster method that is suitable for use in this analysis is the K-Medoids Clustering method. In the K-Medoids method, the center of the cluster is the representative object (medoids) which then places each object against the closest cluster to the medoids, so it is suitable for data containing outliers.

ISSN: 2716-3865 (Print), 2721-1290 (Online)

Elbow Method

The Elbow method is a technique used to determine the optimal number of clusters from available data, such as livestock population data. This process is done by analyzing the differences between clusters. If there are clusters with significant differences, the number of clusters can be considered as the optimal number of clusters.

Determine the number of Clusters as many as n, the calculation starts from K=2. Determine the Cluster center randomly.

Table 4. Livestock Population Data

No.	Province	Dairy Cattle	Beef Cattle	Sheep	Goat	Buffalo	Horse	Pig
C1	Aceh	-0.286	-0.080	0.099	-0.201	1.704	-0.243	-0.531
C2	North Sumatra	-0.207	0.478	0.234	0.132	1.510	-0.285	0.826

Calculate SSE using the equation formula.

Table 5. SSE Equations

Formula	SSE Calculation	Results
$SSE = \sum_{k-1}^{k} \sum_{x1 \in k} \|xi - ck\|_2^2$	(-0,286--0,286)^2+(-0,080--0,080)^2+(0,099-0,099)^2+(-0,201--0,201)^2+(-1,704-1,704)^2+(-0,243--0,243)^2+(-0,531--0,531)^2+(-0,28665--0,286)^2	0,000
$SSE = \sum_{k-1}^{k} \sum_{x1 \in k} \|xi - ck\|_2^2$	(-0,286--0,207)^2+(-0,080-0,478)^2+(0,099-0,234)^2+(-0,201-0,1323)^2+(-1,704-1,510)^2+(-0,243--0,285)^2+(-0,531-0,826)^2+(-0,286--0,207)^2	3,55485

Add the number of Clusters, calculate SSE by adding 1 Cluster to the number of Clusters that have been determined and determine new medoids. Calculate SSE again, with 3 medoids. Calculate until the calculation with the number of Clusters that have been determined, until getting the SSE results for each Cluster. So that it forms an elbow as in Figure 2, the elbow graph shows that the optimal number of k is 3.

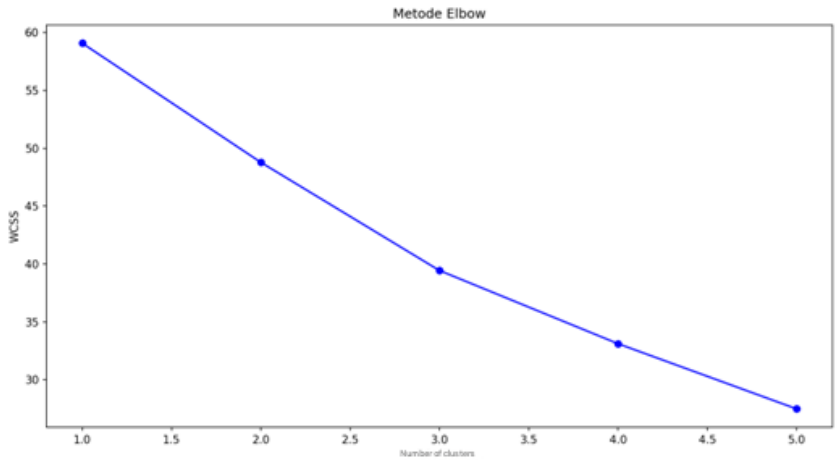

Figure 3. Elbow Graph

ISSN: 2716-3865 (Print), 2721-1290 (Online)

K-Meoids Clustering

After determining the number of k, the next step is to perform the clustering technique using the K-Medoids algorithm. After knowing using the elbow method the number of k is 3, then the K-Medoids calculation is tried using k = 3. The stages of the K-Medoids calculation can be seen below:

Randomly select centroids on the data according to the number of clusters, the centroids used are from C0 of Central Java province, C1 of Maluku province and C2 of NTT province.

Table 6. Initial Centroids

Centroid	Dairy Cattle	Beef Cattle	Sheep	Goat	Buffalo	Horse	Pig
C1	2.079	1.478	3.693	0.979	0.572	-0.066	-0.284
C2	-0.287	-0.454	-0.488	-0.249	-0.375	-0.279	-0.326
C3	-0.286	0.7065	0.381	-0.218	2.989	2.934	4.640

After determining the centroid, then perform the Euclidean distance calculation to determine the proximity of the distance to the existing centroid.

Data 1 to C1

$$b(x_1, y_1) = \sqrt{\begin{matrix}(-0{,}286 - 2{,}079)^2 + (-0{,}080 - 1{,}478)^2 + \\ (0{,}099 - 3{,}693)^2 + (-0{,}201 - 0{,}979)^2 + \\ (-1{,}704 - 0{,}572)^2 + (-0{,}243 - -0{,}066)^2 + \\ (-0{,}531 - -0{,}066)^2\end{matrix}}$$

$b(x_1, y_1) = 4{,}869$

Data 1 to C2

$$b(x_1, y_2) = \sqrt{\begin{matrix}(-0{,}286 - -0{,}287)^2 + (-0{,}080 - 0{,}454)^2 + \\ (0{,}099 - -0{,}488)^2 + (-0{,}201 - 0{,}249)^2 + \\ (-1{,}704 - 0{,}375)^2 + (-0{,}243 - 0{,}279)^2 + \\ (-0{,}531 - 0{,}326)^2\end{matrix}}$$

$b(x_1, y_2) = 2{,}203$

Data 1 to C3

$$b(x_1, y_3) = \sqrt{\begin{matrix}(-0{,}286 - -0{,}286)^2 + (-0{,}080 - 0{,}706)^2 + \\ (0{,}099 - 0{,}381)^2 + (-0{,}201 - 0{,}218)^2 + \\ (-1{,}704 - 2{,}989)^2 + (-0{,}243 - 2{,}934)^2 + \\ (-0{,}531 - 4{,}640)^2\end{matrix}}$$

$b(x_1, y_3) = 6{,}261$

Determine the smallest distance of each data to the initial medoid, and calculate the resulting proximity of the initial medoid.

Table 7. Proximity Calculation Results on Medoid 1

Data	C1	C2	C3	Proximity	Cluster

ISSN: 2716-3865 (Print), 2721-1290 (Online)

1	4.869	2.203	6.261	2.203	2
2	4.590	2.536	5.225	2.536	2
3	5.128	1.638	6.376	1.638	2
…	…	…	…	…	…
34	5.858	1.985	5.952	1.985	2
Total initial medoid proximity distance = 36.651					

Calculation of the 2nd iteration with new centorids randomly selected from the C0 data of Central Java Province, C2 data of Maluku Province and C3 data of South Sumatra Province.

Centroid	Dairy Cattle	Beef Cattle	Sheep	Goat	Buffalo	Horse	Pig
C1	1.765	-0.122	0.923	5.481	0.922	-0.027	-0.520
C2	-0.287	-0.508	-0.594	-0.255	-0.762	-0.318	-0.343
C3	-0.265	1.002	0.288	-0.255	1.848	4.583	1.582

Data 1 to C1

$$b(x_1, y_1) = \sqrt{\begin{matrix}(-0{,}286 - 1{,}765)^2 + (-0{,}080 - -0{,}122)^2 + \\ (0{,}099 - 0{,}923)^2 + (-0{,}201 - 5{,}481)^2 + \\ (-1{,}704 - 0{,}922)^2 + (-0{,}243 - -0{,}027)^2 + \\ (-0{,}531 - -0{,}520)^2\end{matrix}}$$

$b(x_1, y_1) = 6{,}151$

Data 1 to C2

$$b(x_1, y_2) = \sqrt{\begin{matrix}(-0{,}286 - -0{,}287)^2 + (-0{,}080 - 0{,}454)^2 + \\ (0{,}099 - -0{,}488)^2 + (-0{,}201 - 0{,}249)^2 + \\ (-1{,}704 - 0{,}375)^2 + (-0{,}243 - 0{,}279)^2 + \\ (-0{,}531 - 0{,}326)^2\end{matrix}}$$

$b(x_1, y_2) = 2{,}203$

Data 1 to C3

$$b(x_1, y_3) = \sqrt{\begin{matrix}(-0{,}286 - -0{,}265)^2 + (-0{,}080 - 1{,}002)^2 + \\ (0{,}099 - 0{,}288)^2 + (-0{,}201 - -0{,}255)^2 + \\ (-1{,}704 - 1{,}848)^2 + (-0{,}243 - 4{,}583)^2 + \\ (-0{,}531 - 1{,}582)^2\end{matrix}}$$

$b(x_1, y_3) = 5{,}385$

After getting the new centroid, continue with the approximation calculation to get the latest proximity, namely:

Table 8. Proximity Calculation Results on the 2nd Medoids

Data	C1	C2	C3	Proximity	Cluster
1	4,869	2,203	6,261	2,203	2
2	4,590	2,536	5,225	2,536	2
3	5,128	1,638	6,376	1,638	2

ISSN: 2716-3865 (Print), 2721-1290 (Online)

…..	…..	…..	…..	……	…..
34	5,858	1,985	5,952	1,985	2
Total proximity distance of initial medoids = 38,855					

Calculating the total deviation

$S = Total\ Jarak\ Kedekatan\ Baru - Total\ Jarak\ Kedekatan\ Awal$

S = 38.855-36.651 = 2.204

If the total deviation >0, then the loop is complete. So that a cluster is obtained for each province. As in table 9

Table 9. Clusters in Each Province

C1	1	Jawa Barat	3	Jawa Tengah
	2	Jawa Timur		
C2	1	Aceh	16	Kalimantan Barat
	2	Sumatera Utara	17	Kalimantan Tengah
	3	Sumatera Barat	18	Kalimantan Selatan
	4	Riau	19	Kalimantan Timur
	5	Jambi	20	Kalimantan Utara
	6	Sumatera Selatan	21	Sulawesi Utara
	7	Bengkulu	22	Sulawesi Tengah
	8	Lampung	23	Sulawesi Tenggara
	9	Kepulauan Bangka Belitung	24	Gorontalo
	10	Kepulauan Riau	25	Sulawesi Barat
	11	DKI Jakarta	26	Maluku
	12	DI Yogyakarta	27	Maluku Utara
	13	Jawa Timur	28	Papua Barat
	14	Banten	29	Papua
	15	Bali		
C3	1	Nusa Tenggara Timur	2	Sulawesi Selatan

In the table, you can see the determination of clusters for each province, where in C0 there are 3 provinces, C1 there are 29 provinces, and C2 there are 2 provinces.

Table 10. New Medoid Data

Types of Animals	C1	C2	C3
Dairy Cattle	679918	0	6185
Beef Cattle	9034650	538284	6942694
Sheep	19169996	558807	4014868
Goats	11755412	57658	4072
Buffalo	277399	80368	542479
Horses	4391	6982	852075
Pigs	529405	440850	4453538

The table is livestock population data taken from the new medoid where C0 is livestock population data for Central Java province, C1 livestock population data for Maluku province and C2 livestock data for South Sulawesi province. In C0 of the three livestock data in each province, livestock with the type of Dairy Cows as many as 679,918, Beef Cattle as many as

ISSN: 2716-3865 (Print), 2721-1290 (Online)

9,034,650, Sheep as many as 17,950,945 and Goats as many as 11,755,412 have the highest middle value characteristics compared to other clusters, while in C1 with a total of 29 data for each province, has the lowest middle value characteristics for each type of livestock compared to other clusters and in C2 with a total of 2 data for each province, has the highest middle value characteristics for the type of livestock Buffalo as many as 661,140, Horses as many as 708,776 and Pigs as many as 7,666,498.

Table 11. Indicators

Cluster	High Indicators	Low Indicators
Cluster 0	Dairy Cattle, Beef Cattle, Sheep, Goat	None
Cluster 1	None	Dairy Cattle, Beef Cattle, Sheep, Goat, Buffalo, Horse, Pig
Cluster 2	Buffalo, Horse, Pig	None

Davies Bouldin-index evaluation

In this study, to obtain the best provincial cluster based on livestock population, an experiment was conducted to find the smallest DBI value. The results of the experiment can be seen in the table below.

Table 12. Davies Bouldin-index evaluation

Experiment	Number of Clusters	DBI Value
1	2	0,871
2	3	0,759

The results of the Province grouping process using the K-Medoids algorithm by determining the DBI value of each cluster, and based on the results of the Davies Bouldin Index (DBI) recapitulation on the livestock population dataset shown by the results of the value of k = 2, namely 0.871, k = 3, namely 0.759, then the k value that will be selected and become the number of clusters is in the second experiment with a value of k = 3 because it has the smallest DBI value, namely 0.759, because the smaller the DBI value, the more optimal the cluster produced, so that in the livestock population dataset, the clusters used are 3 clusters.

Visualization of Calculation Results

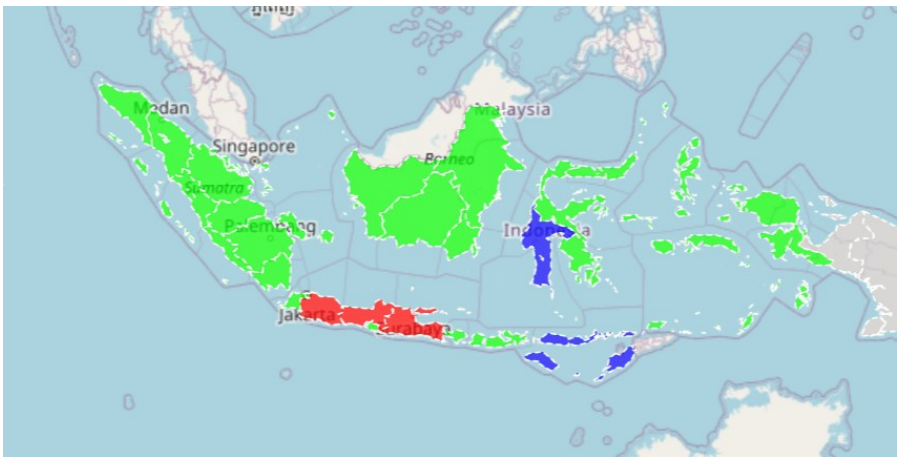

Figure 4. Map of Distribution of Livestock Population

In Figure 4, the livestock population on the island of Java includes the provinces of West Java, Central Java and East Java marked in red with the highest number of livestock for types of animals such as Dairy cattle, Beef cattle, Sheep and Goats, West Java is the province with the largest number of livestock in a period of 5 years, the highest type of livestock in West Java

ISSN: 2716-3865 (Print), 2721-1290 (Online)

is Sheep with a total of 54,593,092 heads, and the province marked in blue is a province that includes the provinces of South Sulawesi and East Nusa Tenggara has the highest number of livestock with the highest population for the types of Buffalo, Horse and Pig. While those marked in green are provinces that have the lowest number of animals.

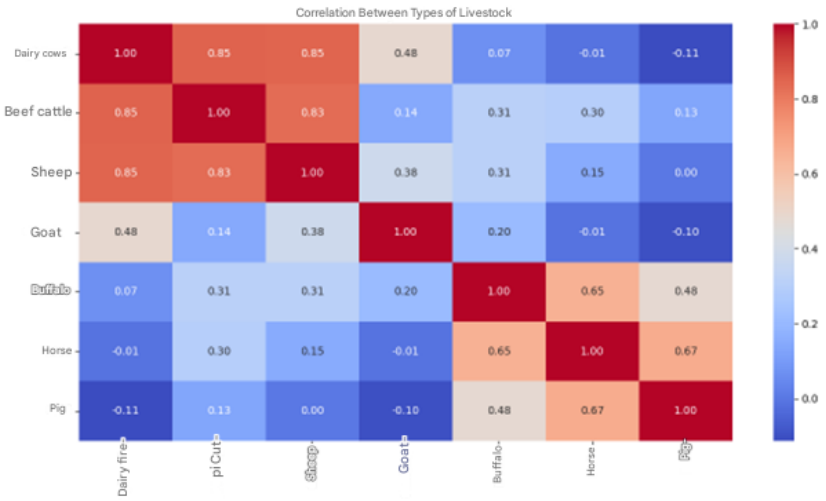

Figure 5. Correlation Between Types of Livestock

Provinces that have a lot of dairy cattle tend to also have a lot of beef cattle and sheep. This indicates that these provinces have a complete livestock industry, both for milk, beef, and sheep. While provinces that have a lot of buffalo tend to also have a lot of horses and pigs. This could reflect that the region tends to be traditional or based on certain customs. And of course its distribution is not evenly distributed in all provinces due to cultural and religious factors. While goats, this type of animal tends to stand alone because the characteristics of goat farming are different, such as narrow land or rural areas.

Conclusion

In the study of the Livestock Population Map in Indonesia using the K-Medoids method which has been carried out based on SSE using the elbow method, it is known that the ideal K value is formed K = 3, therefore the grouping is carried out with a K number of 3. After clustering using K-Medoids, it was found that Cluster 0 had 3 provinces with the highest average population with types of livestock Dairy cattle, Beef cattle, Sheep and Goats, Cluster 1 had 29 provinces with the lowest population compared to other clusters, Cluster 2 had 2 provinces with the highest average number for types of animals Buffalo, Horse and Pig. Based on the Davies Bouldin Index evaluation for 3 Clusters, namely 0.75903755. This shows that the Cluster that was formed has not proven the best cohesion value which is influenced by the dataset, the value of the attribute has a small diversity value.

Confession

I would like to express my gratitude to all parties who supported me in developing this research, especially to my family and of course to my two supervisors, Mrs. Wina Witanti and Mr. Agus Komarudin, who always supported and encouraged me to complete this research.

References

Abbas, S. A., Aslam, A., Rehman, A. U., Abbasi, W. A., Arif, S., & Kazmi, S. Z. H. (2020). K-Means and K-Medoids: Cluster Analysis on Birth Data Collected in City Muzaffarabad, Kashmir. *IEEE Access*, *8*, 151847–151855. https://doi.org/10.1109/ACCESS.2020.3014021

ISSN: 2716-3865 (Print), 2721-1290 (Online)

Ainisa, N., Suarna, N., & Prihartono, W. (2024). Analisis Populasi Ayam Ras Pedaging Di Indonesia Dengan Penerapan Data Mining K-Medoids Clustering. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *7*(6), 3679–3687. https://doi.org/10.36040/jati.v7i6.8251

Alfiah, F., Almadayani, A., Al Farizi, D., & Widodo, E. (2021). Analisis Clustering K-Medoids Berdasarkan Indikator Kemiskinan di Jawa Timur Tahun 2020. *Jurnal Ilmiah Sains*, *22*(1), 1. https://doi.org/10.35799/jis.v22i1.35911

Alfitra, A., Nurdin, N., & Meiyanti, R. (2025). Comparison of K-Means and K-Medoids Methods in Clustering High Population Density Areas in Bireuen Regency. *JOURNAL OF INFORMATICS AND TELECOMMUNICATION ENGINEERING*, *9*(1), 292-302. https://doi.org/10.31289/jite.v9i1.15602

Aqilah, N. Y., & Fauzan, A. (2025, January). K-medoids and agglomerative clustering approach with outlier handling in livestock production clusterization (A case study in Brebes district in 2022). In *AIP Conference Proceedings* (Vol. 3248, No. 1, p. 040011). AIP Publishing LLC. https://doi.org/10.1063/5.0236675

Azzahro, Z., Chyntia, L. F. A. R., Nasar, M., & Widianto, W. (2024, March). Implementation of the k-means clustering method for grouping poverty data: Case studies in Indonesia. In *AIP Conference Proceedings* (Vol. 2927, No. 1, p. 040026). AIP Publishing LLC. https://doi.org/10.1063/5.0193516

Barik, T. K., & Centeno, V. A. (2020). K-Medoids Clustering of Setting Groups in Directional Overcurrent Relays for Distribution System Protection. *2020 IEEE Kansas Power and Energy Conference, KPEC 2020*. https://doi.org/10.1109/KPEC47870.2020.9167531

Baru, M., & Di, T. (2019). *Implementasi Metode K-Medoids Clustering Untuk Mengetahui Pola Pemilihan Program Studi*. *1*(3), 54–69.

BPS Indonesia. (2022). Catalog: 1101001. *Statistik Indonesia 2022*, *1101001*, 790.

Budiman, H., Yuhandra, E., & Faisal, K. A. (2024). Preventive Measure and Action towards Livestock disease: Law on Animal Husbandry and Animal Health. *Unifikasi: Jurnal Ilmu Hukum*, *11*(01), 1-10. https://doi.org/10.25134/unifikasi.v11i01.758

Chen, Z. S., Vaitheeshwari, R., Wu, E. H. K., Lin, Y. D., Hwang, R. H., Lin, P. C., ... & Ali, A. (2024). Clustering apt groups through cyber threat intelligence by weighted similarity measurement. *IEEE Access*. https://doi.org/10.1109/ACCESS.2024.3469552

Dimara, E., & Perin, C. (2020). What is Interaction for Data Visualization? *IEEE Transactions on Visualization and Computer Graphics*, *26*(1), 119–129. https://doi.org/10.1109/TVCG.2019.2934283

FIRDAUS, M. S. (2022). Segmentasi Mutu Jalan Nasional Menggunakan Metode Clustering K-Medoids di Kabupaten Bandung Barat dan Kota Bandung. *Prosiding Sains Nasional Dan Teknologi*, *12*(1), 431. https://doi.org/10.36499/psnst.v12i1.7173

Hallmen, T., Schiller, D., Vehlen, A., Eberhardt, S., Baur, T., Withanage, D., ... & André, E. (2025). DISCOVER: A Data-driven Interactive System for Comprehensive Observation, Visualization, and ExploRation of Human Behavior. *Frontiers in Digital Health*, *7*, 1638539. https://doi.org/10.3389/fdgth.2025.1638539

Hermawati, A., Jumini, S., Astuti, M., Ismail, F., & Rahim, R. (2020). Unsupervised Data Mining with K-Medoids Method in Mapping Areas of Student and Teacher Ratio in Indonesia. *TEM Journal*, *9*(4), 1614–1618. https://doi.org/10.18421/TEM94-37

ISSN: 2716-3865 (Print), 2721-1290 (Online)

Hidayat, M. L., Abdurahman, S. G., Astuti, D. S., Prabawati, R., Anif, S., Hariyatmi, H., & Zannah, F. (2025). Pilot study of digital competency mapping of Indonesian preservice teachers: Rasch model analysis. *Indonesian Journal on Learning and Advanced Education (IJOLAE)*, 100-116.

Jatnika, H., Jamaludin, H., Rahman, A., Ibrahim, I., Sianipar, J. H., Waluyo, A., Maulana Siregar, A. R., & Qomario, Q. (2021). Application of the clustering algorithm to the small and micro industrial companies for mapping regions with k-medoids. *Journal of Physics: Conference Series*, *1933*(1). https://doi.org/10.1088/1742-6596/1933/1/012033

K-, P. M. A., Untuk, M., & Performa, E. (2020). *Pelita teknologi*. *15*(1), 49–55.

Kamel, M., Aleya, S., Almagharbeh, W. T., Aleya, L., & Abdel-Daim, M. M. (2025). The emergence of highly pathogenic avian influenza H5N1 in dairy cattle: implications for public health, animal health, and pandemic preparedness. *European Journal of Clinical Microbiology & Infectious Diseases*, 1-17. https://doi.org/10.1007/s10096-025-05147-z

Kariyam, Abdurakhman, & Effendie, A. R. (2024, November). Modified fast k-medoids algorithm to guarantee no empty clusters. In *AIP Conference Proceedings* (Vol. 3201, No. 1, p. 060009). AIP Publishing LLC.

Kenger, O. N., Kenger, Z. D., Ozceylan, E., & Mrugalska, B. (2023). Clustering of Cities Based on Their Smart Performances: A Comparative Approach of Fuzzy C-Means, K-Means, and K-Medoids. *IEEE Access*, *11*(October), 134446–134459. https://doi.org/10.1109/ACCESS.2023.3333753

Khoirunisa, I., Astuti, R., & Suprapti, T. (2023). *PENERAPAN ALGORITMA K-MEDOIDS DALAM MENENTUKAN CLUSTER KABUPATEN DAN KOTA BERDASARKAN POPULASI PETERNAKAN DI PROVINSI JAWA BARAT*. *7*(6).

Khotimah, B. K., Agustina, F., Puspitarini, O. R., Cahyani, A. D., Kustiyahningsih, Y., & Anamisa, D. R. (2024). Hyperparameters and centroid improvements in the K-medoids method for grouping processed beef SMEs. *Commun. Math. Biol. Neurosci.*, *2024*, Article-ID.

Lewis, G., Snider, B., Vamvakeridou-Lyroudia, L., Chen, A. S., Djordjević, S., & Savić, D. A. (2023). Web-based GIS for simulation, and visualisation: EPANET and data-rich shapefiles. *IOP Conference Series: Earth and Environmental Science*, *1136*(1). https://doi.org/10.1088/1755-1315/1136/1/012007

Lubis, S. N., Menglikulov, B., Shichiyakh, R., Farrux, Q., Khakimboy Ugli, B. O., Karimbaevna, T. M., ... & Jasur, S. (2024). Temporal and spatial dynamics of bovine spongiform encephalopathy prevalence in Akmola Province, Kazakhstan: Implications for disease management and control. *Caspian Journal of Environmental Sciences*, *22*(2), 431-442. https://doi.org/10.22124/cjes.2024.7734

Matematika, T., & Generasi, M. (2023). *SEMINAR NASIONAL JURUSAN MATEMATIKA 2023* (Issue November).

Mustofa, Z. & Iman Saufik Suasana. (2020). Algoritma Clustering K-Medoids Pada E-Government Bidang Information and Communication Technology Dalam Penentuan Status Edgi. *Jurnal Teknologi Informasi Dan Komunikasi*, *9*(1), 1–10. https://doi.org/10.51903/jtikp.v9i1.162

Nurdin, N., Meiyanti, R., & Maulita, M. (2025). Clustering and Mapping of Agricultural Production Based on Geographic Information System Using K-Medoids

ISSN: 2716-3865 (Print), 2721-1290 (Online)

Algorithm. *Journal of Artificial Intelligence and Technology*, *5*, 116-124. https://doi.org/10.37965/jait.2025.633

Peringkat, A., Sinta, (, Berdasarkan,), Dirjen, K., Tinggi, P., Kemendikbud, D. T., Didik Madyatmadja, E., Ridho, M. N., Pratama, A. R., Fajri, M., Novianto, L., & Informasi, S. (2022). *Infotech: Journal of Technology Information Penerapan Visualisasi Data Terhadap Klasifikasi Tindak Kriminal Di Indonesia*. *8*(1), 61–68.

Putri, D. (2025). The Impact of Social Inequality on Educational Quality in Indonesia: Challenges and Policy Recommendations. *Solo Universal Journal of Islamic Education and Multiculturalism*, *3*(01), 43-56. https://doi.org/10.61455/sujiem.v3i01.248

Saleh, R. D., Rustiadi, E., Fauzi, A., & Rahma, H. (2025). Mapping Island Village Typology Based on Spatial and Socio-Economic Characteristics in Maluku Province, Indonesia. *International Journal of Sustainable Development & Planning*, *20*(6). https://doi.org/10.18280/ijsdp.200609

Septhya, D., Rahayu, K., Rabbani, S., Fitria, V., Rahmaddeni, R., Irawan, Y., & Hayami, R. (2023). Implementasi Algoritma Decision Tree dan Support Vector Machine untuk Klasifikasi Penyakit Kanker Paru: Implementation of Decision Tree Algorithm and Support Vector Machine for Lung Cancer Classification. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, *3*(1), 15–19. https://doi.org/10.57152/malcom.v3i1.591

Suharno, B. (2020). Prospek Peternakan Di Era Normal Baru Pasca Pandemi Covid-19 (Sudut Pandang Media). In *Prosiding Seminar Teknologi dan Agribisnis Peternakan*. http://jnp.fapet.unsoed.ac.id/index.php/psv/article/view/467

Syahfitri, N., Budianita, E., Nazir, A., & Afrianty, I. (2023). Pengelompokan Produk Berdasarkan Data Persediaan Barang Menggunakan Metode Elbow dan K-Medoid. *KLIK: Kajian Ilmiah Informatika Dan Komputer*, *4*(3), 1668–1675. https://doi.org/10.30865/klik.v4i3.1525

Tajbar, S., Tajbar, A., Pashaie, Z., Khorshiddoust, A. M., & Rafiq, L. (2024). Application of hierarchical cluster analysis to spatiotemporal variability of monthly precipitation over Khyber Pakhtunkhwa, Pakistan. *Acta Geophysica*, *72*(2), 1159-1174. https://doi.org/10.1007/s11600-023-01161-x

Zhang, H., & Peng, Y. (2024). Image clustering: An unsupervised approach to categorize visual data in social science research. *Sociological Methods & Research*, *53*(3), 1534-1587. https://doi.org/10.1177/00491241221082603

ISSN: 2716-3865 (Print), 2721-1290 (Online)