The 2nd International Conference on Multidisciplinary Engineering (2nd ICOMDEN 2024)

The 2nd International Conference on Multidisciplinary Engineering: 00042 (2024) https://doi.org/10.29103/icomden.v2.xxxx
Volume 2, 2024
eISSN: xxxx-xxxx
Research Original Article

Poverty Level Clustering in Districts/Cities Using the K-Medoids Method Based on Population Data

Cut Syahira Salsabila✉[1] Eva Darnila[2] Cut Agusniar[3]
[1]Department of Informatics, Universitas Malikussaleh, Bukit Indah, Lhokseumawe, 24353, Indonesia, cut.200170166@mhs.unimal.ac.id
[2]Department of Informatics, Universitas Malikussaleh, Bukit Indah, Lhokseumawe, 24353, Indonesia, eva.darnila@unimal.ac.id
[3]Department of Informatics, Universitas Malikussaleh, Bukit Indah, Lhokseumawe, 24353, Indonesia, cutagusniar@unimal.ac.id
✉Corresponding Author: **cut.200170166@mhs.unimal.ac.id | Phone: +6282370178984**

Abstract
Poverty is a serious problem that hinders economic development, especially in developing countries like Indonesia. Aceh Province, especially Bireuen, Aceh Utara, and Lhokseumawe City have significant poverty rates due to high population and limited job opportunities. The K-Medoids algorithm used in this research works well in clustering the sub-districts in the region, with the aim of assisting the government in making more effective decisions. The implementation results show the clustering for the poverty rate in Bireuen in 2021 obtained C1 58.82%, C2 29.41%, C3 11.76%, in 2022 obtained C1 58.82%, C2 29.41%, C3 11.76%, in 2023 obtained C1 64.71%, C2 17.65%, C3 17.65%. In Aceh Utara District, C1 62.96% was obtained, C2 33.33%, C3 3.70%, in 2022 C1 62.96%, C2 33.33%, C3 3.70%, in 2023 C1 51%, C2 44.44%, C3 3.70%. In the city of Lhokseumawe City obtained C1 25%, C2 50%, C3 25%, in 2022 C1 25%, C2 50%, C3 25%, in 2023 C1 25%, C2 25%, C3 50%. The percentage of these results shows that the poverty rate in the three regions increases every year and this requires special attention from the government to minimize the level of poverty through increasing employment, controlling the birth rate, and cash and non-cash assistance programs for poor families.
Keywords: Clustering, K-Medoid, Poverty

Introduction
Poverty is a serious problem, especially in developing countries like Indonesia, as it can hinder a country's growth. Individuals or households are considered poor if they are unable to fulfill basic needs such as clothing, food, shelter, and social needs. Poverty is often associated with a high population which increases unemployment due to a lack of jobs. Countries with large populations tend to have higher poverty rates. The government also always moves to overcome this problem, such as opening up jobs, providing poor assistance, reducing the growth rate with family planning programs and others.

According to the Badan pusat Statistik (2023), Aceh is the poorest province in Sumatra with a poverty rate of 14.18%, far above the national average of 9.61%. Some areas with high poverty include Aceh Utara (16.64%), Bireuen (12.12%), and Lhokseumawe (10.73%), this is caused by population growth that hampers employment [1]. By using several indicators to measure the poverty rate, it must use a lot of data. With this large amount of data, it is certainly difficult to determine which areas in Bireuen Regency, Aceh Utara and Lhokseumawe City are included in the high poverty rate. So the use of data mining technology is needed to overcome these problems. Data mining technology is used because this technology is simple and fast to cluster which regions fall into the category of high or low poverty rates. An effort that can be made to assist the government in reducing the number of poverty rates is by clustering which areas have a high poverty rate so that the government can focus on the problems in the region,

Data mining is a tool that allows users to quickly access large amounts of data. In particular, data mining refers to tools and applications that use statistical analysis to process data. This process involves extracting or extracting information from large databases that was previously unknown, but can be understood and useful in making crucial business decisions [2]. Data mining includes various techniques aimed at finding undetected patterns in the data that has been collected. With data mining, users can identify knowledge in databases that was previously unseen [3].

The K-Medoids method has previously been used in various studies, including those conducted to group regions with increasing and decreasing amounts of panagan crop yields in North Sumatra. The K-Medoids method is classified

as a method that is quite efficient on small data and searches for points that are representative and can overcome outliers. The results showed that the K-Medoids algorithm produced a DBI (Davies Bouldin Index) value of 0.062 and a Silhouette Coefficient value of 0.8980. There are three clusters formed, where cluster 0 is dominated by an increase in corn and peanut production by 5% each, cluster 1 is dominated by a decrease in soybean production by 38%, and cluster 2 is dominated by a decrease in mung bean production by 33%."[4]

Research by Adinda Astasia to group all provinces in Indonesia based on poverty variables, life expectancy, average years of schooling, expected years of schooling, and per capita expenditure in 2020. there are indications of outliers in the data, so the K-Medoid method is suitable to be chosen. According to the results, three groups were formed. The first group has the highest poverty rate, as well as the lowest life expectancy, years of schooling, and per capita expenditure. The second group is characterized by a moderate average of the variables, and the third group has the highest poverty rate and the average of the other variables [5].

Based on the existing problems, the authors are interested in conducting research with the title "Clustering Poverty Levels in Districts / Cities Based on Population Data Using the K-Medoids Method (Case Study: Bireuen, Aceh Utara, Lhokseumawe)". This research was conducted in order to find out which district in the Regency / City is included in the poverty level of poor, vulnerable poor and Extremely poor poverty levels. It is hoped that the results of this study can provide information to the government in order to optimize poverty reduction in the region.

Literature Review

Clustering

Clustering is a data grouping method that involves the process of dividing a set of data objects into subsets called clusters. Observing and studying the attributes of each data object is part of the clustering process, which allows data objects to be grouped into different classes according to their similarities. that they have in common. To determine how close or similar objects are to each other in a clustering algorithm, one of the most common techniques for clustering is calculating the distance between geometric data [6]. Clustering is a trusted data mining method and is a valid tool for addressing complex problems in computer science and statistics. The clustering process involves grouping data points into two or more groups, where data points in one group tend to be more similar to each other than to other groups [7].

Method K-Medoids

K-Medoids is also known as the distribution method because it uses the most central object (medoid), which is the average of the objects in the cluster, as the cluster center to determine cluster values [8]. The K-Medoids method is a clustering technique that is closely related to the K-Medoids and Medoidshift methods. The K- Medoids algorithm, often referred to as PAM (Partition Around Medoids), uses medoids as representatives of each cluster. This algorithm aims to minimize the difference between data points in a cluster by selecting one data point as the center (medoid) of each cluster. In K-Medoids, the cluster center is between the data points. The distance between an object and the center is calculated using Euclidean distance, so the object most likely to be the center is randomly selected [9].

The K-Medoids algorithm does not use the average object in a cluster as a reference point, but chooses the medoid (center value). Thus, a fixed partitioning method can be performed with the principle of minimizing the difference between each object and the relevant medoid [10]. The main strategy of the K-Medoids algorithm is to identify k object clusters by randomly selecting the original object (medoid) as the representative of each cluster. Subsequently, every other object is clustered with the most similar medoid. The K-Medoids method focuses on the reference object, not on the average of the objects in each cluster. This algorithm accepts the number of clusters k as an input parameter, which is then distributed among n objects. [11]

Stages Of The K-Medoids Method

The completion stage in the K-Medoids algorithm involves several main steps, starting from initializing the cluster center to determining the final medoid in each cluster. Here are the complete steps, [12]:

1. Initialize a number of k points as clusters (number of clusters).
2. Distribute all data (objects) into clusters based on the closest distance using the Euclidean Distance formula:

   $$d(x, y) \sum_{i=0}^{n} (xi - yi)2 \quad (1)$$

   Description:
   d(x,y) = distance between the 1st data and the j data
   xi1 = value of the first attribute of the i data
   yi1 = value of the first attribute of the j data
   n = number of attributes used
3. Randomly select one object in each cluster as the new medoid candidate.
4. Calculate the distance of all objects in each cluster to the new candidate medoid.
5. Calculate the total deviation (D) by subtracting the new total distance from the old total distance. If D< 0, swap objects with the cluster data to form a new group of k objects as the medoid.
6. Repeat steps 3 to 5 until there is a change in the medoid, so that the clusters and members of each cluster are obtained.

K-Medoids Scheme

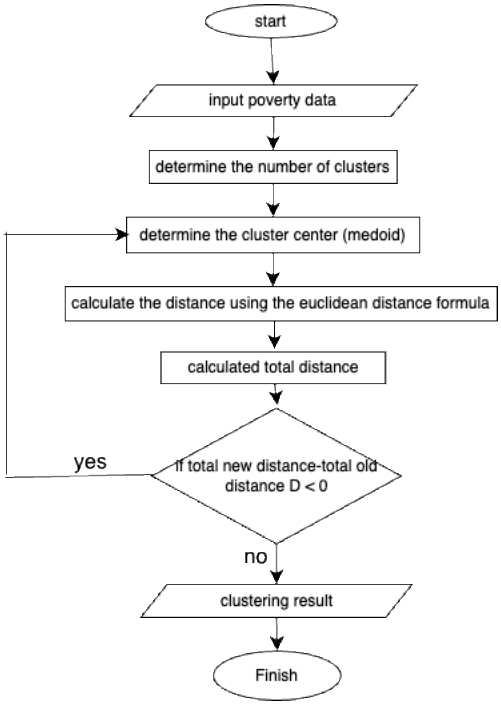

Figure 1. K-Medoids Scheme

Figure 1 is a process or system design scheme in this research, from start to finish:

1. The algorithm scheme process begins.
2. The process of inputting poverty criteria data, namely the Number of Poor People, Population Growth Rate, and Population Density.
3. Determining the number of clusters, in this study the cluster amounted to 3.
4. After determining the number of clusters, the euclidean distance will be calculated using equation (1).
5. The process of calculating the total deviation (D) is by calculating the value of the new total distance - the old total distance. If D<0 or repeat steps 3-5 until no more medoid changes occur.
6. When the calculation has been completed, the cluster results will come out which can determine whether the data is included in the vulnerable poor, poor and very poor clusters.
7. The calculation process is complete.

Calculation Analysis Of K-Medoids Method

The dataset that has been collected will be analyzed using the K-Medoids method. The analysis uses 3 criteria attributes, namely data on the number of poor families, population density and population growth rate. And the grouped areas are Bireuen Regency, North Aceh and Lhokseumawe City. Then in the calculation using the K-Medoids method, it is necessary to determine the number of clusters to be used, which can be seen in table 1 below.

Table 1. Claster Labels

Claster	Label
C1	Vulnerable Poor
C2	Poor
C3	Extremely Poor

Table 2. Variable initialization

Variable Name	Initialization
Number of Poor Families	X1
Population Density	X2
Population Growth Rate	X3

The following data for Bireuen Regency in 2021 is organized by sub-district to form a dataset to be grouped in table 3 below.

Table 3. Bireuen District Data In 2021

No	Kecamatan	X1	X2	X3

1	Samalanga	1008	199	0,64
2	Simpang Mamplam	1414	176	0,65
3	Pandrah	1665	78	0,95
4	Jeunib	1845	226	0,8
5	Peulimbang	1790	96	1,22
6	Peudada	994	90	1,05
7	Juli	1055	148	1,16
8	Jeumpa	568	342	0,94
9	Kota Juang	1980	2826	0,25
10	Kuala	1688	1106	1,04
11	Jangka	1496	768	0,67
12	Peusangan	772	898	0,62
13	Pesangan Selatan	1055	160	0,87
14	Pesangan Siblah Krueng	820	108	0,92
15	Makmur	876	220	0,77
16	Gandapura	945	510	0,85
17	Kuta Blang	1104	588	0,69

The steps taken in completing the calculation of the K-Meodids Method are as follows:

Determining the Number of Clusters; The number of clusters used in the poverty level grouping is 3 clusters. The clusters include Vulnerable Poor (C1), Poor (C2) and Extremely Poor (C3) based on poverty data with a total of 17 sub-districts in Bireuen Regency.

Normalization aims to equalize the value scale of each attribute in the data set, so that each variable has an equal weight and no attribute dominates the influence on the final result of the grouping. This process is carried out on poverty data including the Number of Poor Families, Population Density and Population Growth Rate, then processed using the min-max normalization method with a scale of 0 to 1 for the following results:

X1 = (1008 – 568)/(1980-568) = 0,31
X2 = (199-78)/(2826-78) = 0,04
X3 = (0,64-0,25)/(1,22-0,25) = 0,40

Continue until all rows and columns of data are last. So the final normalization result is as follows.

Table 4. Bireuen Normalization Data 2021

No	Kecamatan	X1	X2	X3
1	Samalanga	0,31	0,04	0,40
2	Simpang Mamplam	0,60	0,04	0,41
3	Pandrah	0,78	0,00	0,72
4	Jeunib	0,90	0,05	0,57
5	Peulimbang	0,87	0,01	1,00
6	Peudada	0,30	0,00	0,82
7	Juli	0,34	0,03	0,94
8	Jeumpa	0,00	0,10	0,71
9	Kota Juang	1,00	1,00	0,00
10	Kuala	0,79	0,37	0,81
11	Jangka	0,66	0,25	0,43
12	Peusangan	0,14	0,30	0,38
13	Pesangan Selatan	0,34	0,03	0,64
14	Peusangan Siblah Krueng	0,18	0,01	0,69
15	Makmur	0,22	0,05	0,54
16	Gandapura	0,27	0,16	0,62
17	Kuta Blang	0,38	0,19	0,45

1. Determining Initial Medoids; Determining the initial center of the cluster (Medoids) is done randomly or randomly. For the initial Medoids can be seen in table 5.

Table 5. Early Bireuen Medoids 2021

No	Kecamatan	X1	X2	X3
1	Samalanga	0,31	0,04	0,40
2	Kota Juang	1,00	1,00	0,00
3	Gandapura	0,27	0,16	0,62

2. Calculate the Euclidean Distance with the initial medoids that have been determined;
$d(x_1y_1) = \sqrt{(0{,}31 - 0{,}31)^2 + (0{,}04 - 0{,}04)^2 + (0{,}40 - 0{,}40)^2} = 0{,}00$

$d(x_2y_1) = \sqrt{(0{,}31 - 1{,}00)^2 + (0{,}04 - 1{,}00)^2 + (0{,}04 - 0{,}00)^2} = 1{,}24$

$d(x_3y_1) = \sqrt{(0{,}31 - 0{,}27)^2 + (0{,}04 - 0{,}16)^2 + (0{,}40 - 0{,}62)^2} = 0{,}25$

The process continues until the entire last row of data by calculating the Euclidean distance like the previous step. After all Euclidean distances are obtained, select the smallest distance and calculate the total distance of all data in the cluster. The results are presented in the form of table 6.

Table 6. Euclidean Distance Iteration-1 Bireuen 2021

No	C1	C2	C3	Min	Cluster
1	0,00	1,24	0,25	0,00	C1
2	0,29	1,12	0,41	0,29	C1
3	0,57	1,25	0,54	0,54	C3
4	0,62	1,11	0,65	0,62	C1
5	0,82	1,42	0,73	0,73	C3
6	0,42	1,47	0,26	0,26	C3
7	0,54	1,50	0,35	0,35	C3
8	0,44	1,52	0,29	0,29	C3
9	1,24	0,00	1,28	0,00	C2
10	0,71	1,05	0,60	0,60	C3
11	0,40	0,93	0,44	0,40	C1
12	0,31	1,17	0,30	0,30	C3
13	0,24	1,33	0,15	0,15	C3
14	0,32	1,46	0,19	0,19	C3
15	0,16	1,34	0,14	0,14	C3
16	0,25	1,28	0,00	0,00	C3
17	0,17	1,12	0,20	0,17	C1
	Total Proximity (Min 1+…+Min 17)			5,03	

3. Randomly Select a New Medoid; The new medoid is randomly selected under the condition that it cannot use the previously selected medoid. The new medoids can be seen in table 7.

Table 7. New Medoids Bireuen 2021

No	Kecamatan	X1	X2	X3
1	Simpang Mamplam	0,60	0,04	0,41
2	Pandrah	0,78	0,00	0,72
3	Jeunib	0,90	0,05	0,57

Recalculating Euclidean Distance; After obtaining the new medoid, the data is recalculated using the Euclidean method with the new center point (medoid). The calculation process is done in the same way as in the first iteration. The results of the second iteration calculation are shown in table 8.

Table 8. Euclidean Distance Iteration-2 Bireuen 2021

No	C1	C2	C3	Min	Cluster
1	0,29	0,57	0,62	0,29	C1
2	0,00	0,36	0,34	0,00	C1
3	0,36	0,00	0,21	0,00	C2
4	0,34	0,21	0,00	0,00	C3
5	0,65	0,29	0,44	0,29	C2
6	0,51	0,49	0,66	0,49	C2
7	0,58	0,48	0,67	0,48	C2
8	0,67	0,78	0,92	0,67	C1
9	1,12	1,25	1,11	1,11	C3
10	0,56	0,39	0,42	0,39	C2
11	0,22	0,40	0,34	0,22	C1
12	0,53	0,78	0,82	0,53	C1
13	0,34	0,44	0,56	0,34	C1

14	0,51	0,60	0,74	0,51	C1
15	0,40	0,59	0,69	0,40	C1
16	0,41	0,54	0,65	0,41	C1
17	0,27	0,51	0,55	0,27	C1
	Total Proximity (Min 1+…+Min 17)			6,39	

After getting the 2nd iteration distance, then calculate the total deviation. If the total deviation D>0 then the calculation will be stopped. If D>0 then it will be iterated again as before. The total deviation is asa follows:
Deviation (D) = total new distance – total old distance
D (Deviation) = 6.39 - 5.03 = 1.36
Since the total deviation result is more than 0, the object does not need to be replaced and this test stops at the 2nd iteration, the last iteration result will be the clustering parameter, so the final cluster result is as follows

Table 8. Final Result of K-Medoids Calculation For Bireuen 2021

No	Kecamatan	Cluster
1	Samalanga	Vulnerable Poor
2	Simpang Mamplam	Vulnerable Poor
3	Pandrah	Poor
4	Jeunib	Extremely Poor
5	Peulimbang	Poor
6	Peudada	Poor
7	Juli	Poor
8	Jeumpa	Vulnerable Poor
9	Kota Juang	Extremely Poor
10	Kuala	Poor
11	Jangka	Vulnerable Poor
12	Peusangan	Vulnerable Poor
13	Pesangan Selatan	Vulnerable Poor
14	Peusangan Siblah Krueng	Vulnerable Poor
15	Makmur	Vulnerable Poor
16	Gandapura	Vulnerable Poor
17	Kuta Blang	Vulnerable Poor

Final Results Of Poverty Level Clustering In Each District

Based on the K-Medoids calculation at the previous point, it is implemented on the entire dataset, namely 3 districts and cities from 2021 to 2023. The results of all clusters can be seen in the following table.

Table 9. Final Result Of Clustering The Poverty Rate Of Bireuen District

Year	Kecamatan		
	C1 (Vulnerable Poor)	C2 (Poor)	C3 (Extremely Poor)
Bireuen 2021	10	5	2
Bireuen 2022	10	5	2
Bireuen 2023	11	3	3

The table shows the percentage of poverty in Kabupaten Bireuen during 2021-2023, with three categories: Vulnerable to Poverty (C1), Poor (C2), and Extremely Poor (C3). In 2021 and 2022, the majority of kecamatan were classified as Vulnerable to Poverty (C1), but in 2023 there was a significant change: kecamatan in the C1 category increased to 64.71%, while kecamatan in the C2 category decreased to 17.65%, and the C3 category increased to 17.65%. Although most sub-districts are still classified as Vulnerable to Poor, the increase in sub-districts in the Very Poor category indicates serious economic challenges, requiring interventions in the form of job expansion, family planning programs, and assistance to poor families.

Table 10. Final Result Of Clustering The Poverty Rate Of Lhokseumawe City

Year	Kecamatan		
C1 (Vulnerable Poor)		C2 (Poor)	C3 (Extremely Poor)
Lhokseumawe 2021	1	2	1
Lhokseumawe 2022	1	2	1
Lhokseumawe 2023	1	1	2

Table 11. Final Result Of Clustering The Poverty Rate Of Aceh Utara District

Year	Kecamatan		
	C1 (Vulnerable Poor)	C2 (Poor)	C3 (Extremely Poor)
Aceh Utara 2021	17	19	**1**
Aceh Utara 2022	17	19	**1**
Aceh Utara 2023	14	12	**1**

Conclusions

This research successfully made calculations using the K-Medoids method to cluster poverty levels in Bireuen, Aceh Utara, and Lhokseumawe Districts based on population data. The implementation of the K-Medoids algorithm resulted in three clusters: vulnerable poor (C1), poor (C2), and extremely poor (C3). In Bireuen, the vulnerable poor cluster (C1) increased from 58.82% (10 kecamatan) in 2021-2022 to 64.71% (11 kecamatan) in 2023, while the poor cluster (C2) decreased. In Aceh Utara, the vulnerable poor cluster (C1) decreased from 62.96% (17 kecamatan) in 2021-2022 to 51% (14 kecamatan) in 2023, with an increase in the poor cluster (C2). In Lhokseumawe City, there was a significant increase in the extremely poor cluster (C3) from 25% in 2021-2022 to 50% in 2023.This conclusion shows that the K-Medoids method is effective in clustering poverty levels and helps provide relevant data for policy making in handling poverty in the area.

References

[1] Badan Pusat Statistik Provinsi Aceh, (2024) Provinsi Aceh Dalam Angka 2024. Aceh: BPS Provinsi Aceh.

[2] I. Hidayat, E. Darnila, And Y. Afrillia, (2023) "Clustering Zonasi Daerah Rawan Bencana Alam Di Kabupaten Mandailing Natal Menggunakan Algoritma K-Means," Vol. 7, No. 3, Pp. 1218–1226, Doi: 10.33379/Gtech.V7i3.2880.

[3] C. Zai, (2022) "Implementasi Data Mining Sebagai Pengolahan Data," Vol. 2, No. 3, Pp. 1–12.

[4] N. P. Dharshinni And C. Fandi, (2022)."Penerapan Metode K-Medoids Clustering Untuk Mengelompokkan Ketahanan Pangan," Vol. 6, Pp. 2301–2308, Doi: 10.30865/Mib.V6i4.4939.

[5] A. Astasia, (2020). "Analisis Cluster Kemiskinan Dan Indeks Pembangunan Manusia Di Indonesia Dengan K-Medoids," Vol. 4, Pp. 1–8.

[6] Lina Mardiana Harahap, W. Fuadi, L. Rosnita, E. Darnila, And R. Meiyanti, (2022). "Klastering Sayuran Unggulan Menggunakan Clustering Of Featured Vegetables Using The K-Means Algorithm," Vol. 8, Pp. 567–579.

[7] Gustientiedina, M. H. Adiya, And Y. Desnelita, (2019). "Penerapan Algoritma K-Means Untuk Clustering Data Obat- Obatan Pada Rsud Pekanbaru," Vol. 01, Pp. 17–24, Doi: 10.25077/ Teknosi.V5i1.2019.17-24.

[8] P. N. Safitri, R. Aristawidya, And S. B. Faradilla, (2021). "Klasterisasi Faktor-Faktor Kemiskinan Di Provinsi Jawa Barat Menggunakan K-Medoids Clustering," Vol. 4, No. 2, Pp. 75–80, Doi: //Doi.Org/10.32665/James.V4i2.242.

[9] T. Ramayanti, E. Haerani, And L. Oktavia, (2023). "Penerapan Algoritma K-Medoids Pada Clustering Penerima Bantuan Pangan Non Tunai (Bpnt)," Vol. 7, Pp. 1287–1296, Doi: 10.30865/Mib.V7i3.6475.

[10] T. N. P. Siti Nurlaela, Aji Primajaya, (2020). "Algoritma K-Medoids Untuk Clustering Penyakit Maag Di Kabupaten Karawang," Vol. 12, No. 2, Pp. 56–62, Https://Doi.Org/10.36723/Juri.V12i2.234.

[11] Y. Diana And F. Hadi, (2023). "Analisa Penjualan Menggunakan Algoritma K-Medoids Untuk Mengoptimalkan Penjualan Barang," Vol. 7, No. 1, Pp. 97–103, Https://Doi.Org/10.35145/Joisie.V7i1.2905.

[12] Fajriana, (2021). "Analisis Algoritma K-Medoids Pada Sistem Klasterisasi Produksi Perikanan Tangkap Kabupaten Aceh Utara," Vol. 7, No. 2, Pp. 263–269, Doi: Http://Dx.Doi.Org/10.26418/Jp.V7i2.47795.