ENHANCED CLUSTERING USING PSO-KMEDOIDS FOR GOVERNMENT AID DISTRIBUTION

Aulia Nur Fitriani[1], Kartika Maulida Hindrayani[2], Trimono Trimono[3]
Data Science, UPN "Veteran" Jawa Timur
Jl. Rungkut Madya, Gn. Anyar, Kec. Gn. Anyar, Surabaya, Jawa Timur, Indonesia
e-mail: *[1]**21083010051@student.upnjatim.ac.id**, [2]kartika.maulida.ds@upnjatim.ac.id, [3]trimono.stat@upnjatim.ac.id

Abstract
The distribution of social assistance in Indonesia often experiences problems due to inaccuracies in recipient data between those recorded in government systems and field conditions. In Kalipuro Village, Mojokerto District, data mismatches caused difficulties in screening assistance, requiring village officials to manually re-filter the data. This triggered protests from citizens who should have received assistance but did not get their rights. To overcome this problem, this research proposes the use of the K-Medoids algorithm which is able to overcome sensitivity to outliers. This algorithm is used to cluster data based on criteria such as occupation, number of assets, number of dependents, and income. In addition, this research incorporates the Particle Swarm Optimization (PSO) technique to optimise the clustering process, which is expected to improve accuracy and efficiency in social assistance distribution. The results of clustering analysis using the K-Medoids algorithm show that the best cluster is obtained at the number of clusters K=5, with the distribution of cluster 0 (179 households), cluster 1(89 households), cluster 2 (296 households), cluster 3 (354 households), and cluster 4 (94 households). The Silhouette Score value of 0.6531 indicates good cohesion and separation between clusters. Based on the analysis, cluster 1 is the top priority group of aid recipients, followed by clusters 4, 2, 3, and 0. The K-Medoids algorithm effectively identifies the most needy community groups, supporting targeted and efficient decisions in aid distribution.

Keyword: *Aid Priorities, K-Medoids, Particle Swarm Optimization, Poverty*

INTRODUCTION

Poverty is a crucial issue faced by many countries, including Indonesia, due to the lack of access to education, employment, and basic needs (Novianti et al., 2021). Inequality in income distribution worsens the situation, although various policies have been implemented to address it (Trimono et al., 2023). Data from the Central Bureau of Statistics shows a decline in the national poverty rate from 28.17 million people in March 2023 to 25.22 million people in March 2024, while East Java recorded a poverty rate of 9.79% in 2024 (BPS, 2024). The Integrated Social Welfare Data (DTKS) and the Targeting for the Acceleration of the Elimination of Extreme Poverty (P3KE) are government programmes aimed at improving the accuracy of beneficiary data collection for social assistance. Despite the implementation of these programmes, data inaccuracies remain prevalent (Kemenko Perekonomian, 2018). Iskan Qolba Lubis, a member of Commission VIII of the House of Representatives, emphasised the urgency of updating DTKS to address this persistent issue (EMedia RI, 2024). Moreover, research has shown that inaccurate data collection can prevent the Family Hope Programme (PKH) from reaching the intended beneficiaries, particularly the poor, and may lead to social jealousy within communities (Musaddad & Kriswibowo, 2021). Kalipuro Village in Mojokerto is a clear example, where government data and field conditions do not match, causing assistance to be received by unauthorised parties, such as government employees, triggering protests from residents.

The K-Medoids Clustering approach is proven to provide an optimal solution in the process of grouping recipients of poor community assistance, as applied in the Southeast Aceh Social Service Office (Rispandi, 2023). Other research shows that the K-Medoids algorithm can be used to group prospective Bidikmisi scholarship recipients at Budi Darma University, making the selection process more efficient and structured (Buulolo et al., 2020). In addition, recent research has shown that optimising the initial centroid selection using Particle Swarm Optimization (PSO) can improve the quality of clustering results in K-Medoids, reduce dependence on outliers, and produce more uniform clusters (Wijaya et al., 2024).

K-Medoids Clustering algorithm proved effective in overcoming the weakness of K-Means which is vulnerable to outliers (Muhima et al., 2018). In addition, this algorithm is superior in handling large-scale data (Hu, 2024). By using PSO for initial centroid selection and clustering quality evaluation through Silhouette Score index, K-Medoids can be optimised to provide more accurate and efficient results. This research aims to develop a K-Medoids Clustering-based approach optimised with PSO to improve the accuracy and efficiency of social assistance distribution in Kalipuro Village, Mojokerto Regency. It is hoped that this technology can reduce errors in beneficiary data collection, improve the screening process, and support sustainable poverty alleviation efforts.

RESEARCH METHOD

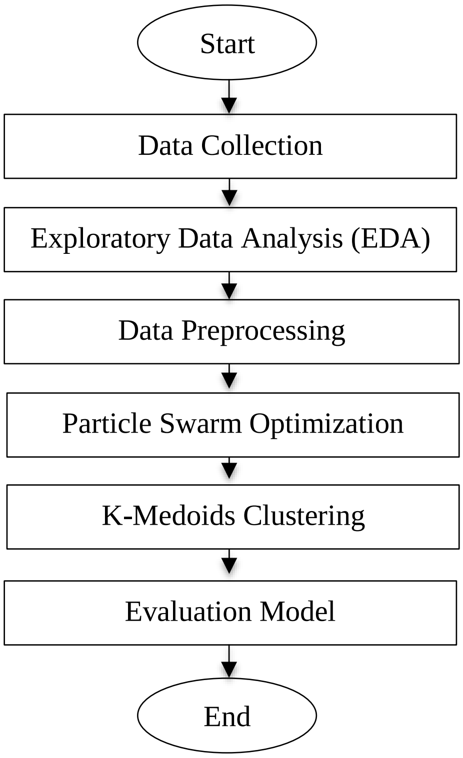

Figure 1. Flowchart Research Method

1. **Data Collection**
   Research data were obtained from five hamlets in Kalipuro Village, namely Madyopuro, Kaliurip, Lamongan, Sidomukti, and Wonoayu, with a total of 1,012 household heads. Factors analysed included occupation, monthly income, and assets (rice fields, land, houses, and vehicles). Data sources came from village websites (https://kalipuro.desa.id), which provide basic information and village archives containing more detailed data on assets and income. To ensure the accuracy of the data, interviews were conducted with relevant parties, including village officials who manage the funds and aid distribution.

2. **Exploratory Data Analysis (EDA)**
   Exploratory Data Analysis (EDA) is a method used to analyse and visualise data with the aim of gaining a better understanding of the information contained in the data (Radhi et

P-ISSN: 2622-6901
E-ISSN: 2622-6375

al., 2022). Through Exploratory Data Analysis (EDA), data can be presented visually, making the process of understanding and interpretation easier. In this research, the Exploratory Data Analysis (EDA) stage includes:

- Variable Distribution Analysis
  The first stage is variable distribution analysis, which aims to explore the distribution and relationships between variables in the dataset. The use of pair plots helps visualise the distribution of each variable through histograms or density plots, as well as the interrelationship patterns between variables through scatter plots. In addition, pair plots also enable the detection of outliers that may affect further analyses.
- Correlation Analysis Between Variables
  The next step is Inter-Variable Correlation Analysis examines the relationship between dataset variables through correlation coefficients that range from 0 to +1 in this case. The correlation matrix displays this coefficient, where values closer to +1 indicate a stronger positive correlation. Seaborn's heatmap visualises this matrix using colours - dark red for strong positive correlations (highest range), medium red for moderately strong correlations (middle range), pink for medium correlations, and lighter/white colours for weaker relationships (lowest range). This visualisation helps analysts to quickly identify patterns and variable relationships that may not be obvious in the raw data, aiding in feature selection and understanding the structure of the dataset.
- Descriptive Statistical Analysis
  The next step is Descriptive Statistical Analysis to see a summary of the data, such as the number of samples (count), mean, standard deviation (std), as well as minimum, maximum and quartile values. This analysis helps to understand the distribution and scale of the data. It is also useful for evaluating the variation between variables.

3. Data Preprocessing

Data preprocessing is the initial stage in data processing that aims to ensure data quality remains optimal. This process involves cleaning the data to remove noise to make the data more reliable and ready for further analysis. Systematic steps are taken so that the data used is in the best condition possible (Alghifari & Juardi, 2021). Following are the steps in data preprocessing:

- Missing Value Detection
  After conducting the Exploratory Data Analysis stage, the next step is to check for missing values. This method helps avoid missing important information and ensures that the dataset can still be used for further analysis.
- Outlier Identification
  At this stage, outliers are identified using the Interquartile Range (IQR) method without changing the data. Outliers are calculated based on the lower and upper bounds of the IQR for each numeric column in the dataset. The purpose of this check is to understand the distribution of the data and ensure that the presence of outliers does not affect the analysis results, especially since the K-Medoids method used is robust to outliers.
- Data Normalization
  At this stage, normalization was performed to scale the features using MinMaxScaler with a scale of 0 to 1. In addition, feature weighting was applied to determine the level of influence of each variable in the analysis. The highest weight is given to income, followed by the number of house/land/rice assets, then the number of car assets, and the lowest is the number of motorbike assets. This is done to adjust the level of importance of each variable in the analysis.

P-ISSN: 2622-6901
E-ISSN: 2622-6375

4. Particle Swarm Optimization

Particle Swarm Optimization (PSO) is a search method that uses collaboration between individuals in a group (‘swarm’) to obtain the most optimal solution. It is based on the collective intelligence and behaviour of insect swarms. In PSO, potential solutions are represented by particles moving in the search space. Each particle has an updated position and velocity based on its own personal best (P_{best}) and the best experiences from across the herd (global best atau G_{best}) (Pratiwi et al., 2024).

PSO is performed to determine the initial cluster centre point before performing k-medoids, so that the clustering results become more optimal. The PSO process starts with initialising the position and speed of the particles randomly. The particle position is calculated using the formula:

$$S(i) = 1 + decimal(s) \times \frac{a-1}{2^n-1} \tag{1}$$

$S(i)$ is the position of the i particle, decimal (s) is a random value, a is the upper bound of the search space, and (n) is the precision of the binary number representation.

Next, the fitness value of each particle is calculated to measure the quality of the resulting solution. Based on the fitness value, the value of P_{best} , which is the best position ever reached by a particle dan G_{best} , which is the best position ever reached by all particles, is updated By:

$$c_2 r_2 \left[P_{best} - x_j(i-1)\right], j = 1,2,\ldots,N \tag{2}$$

$$G_{best}(c) = Pbest(i) \tag{3}$$

The velocity of each particle is updated using the formula:

$$V_1(C_1) = \omega V_0 + c_1 r_1 (p_{best} - C_1) + c_2 r_2 (g_{best} - C_1) \tag{4}$$

$V_1(C_1)$ is the new velocity, V_0 is the new velocity, ω is the inertia factor, c_1 dan c_2 are learning coefficients, r_1 and r_2 are random numbers in the interval $[0,1]$ $Pbest$ is the individual best position G_{best} is the global best position, and C_1 is the current position. Iteration process is performed until convergent, and the final result is the optimal solution that can be used to determine the cluster centre.

5. K-Medoids Clustering

K-Medoids is a clustering method that serves as a more robust variant of K-Means. It addresses one of the key limitations of K-Means, namely its sensitivity to outliers (Muhima et al., 2018). Due to its effectiveness in handling outliers and non-numerical data, the K-Medoids algorithm has been widely applied in fields such as bioinformatics, image analysis, and recommendation systems (Hu, 2024).

The process starts with determining the number of clusters (k) and randomly selecting initial medoids from the dataset, which serve as the reference points for grouping. Next, the distance between each data point and the selected medoids is calculated using the Euclidean distance formula, allowing the algorithm to assess similarity and assign each point to the nearest medoid.

$$d_{(x,y)} = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{5}$$

After that, the closest distance of each object to the medoid is calculated, and the closest object is assigned to the medoid cluster. The total deviation (S) is calculated to evaluate how good the cluster division is, using the formula:

$$S = \sum_{i=1}^{k} \sum xj \in c_i d_{ij} \quad (6)$$

C_i is cluster i, d_{ij} is the distance between object j and medoid i. Iterations are performed by replacing medoids to new objects in the cluster to try to reduce the total deviation (S). The change in deviation is calculated as:

$$\Delta S = S_0 - S_1 \quad (7)$$

S_0 s the total deviation before the medoid change, and S_1 is the total deviation after the medoid change. If $\Delta S > 0$, the new medoid is accepted because it produces better clustering.

This process continues until the total deviation does not change significantly or reaches the maximum iteration. The end result is a cluster with members determined by proximity to the selected medoid.

6. Evaluation Model with Silhouette Score

Silhouette Score is an evaluation metric commonly used in cluster analysis. The silhouette coefficient is calculated by considering the average intra-cluster distance (a) and the average distance to the nearest cluster (b) for each data point (Shahapure & Nicholas, 2020). Mathematically, the silhouette coefficient for a sample is expressed by the following formula:

$$S = \frac{(b-a)}{max(a,b)} \quad (8)$$

A Silhouette Score close to +1 indicates that a data point is in the right cluster with high conformity. Conversely, a score close to 0 indicates that the data point is in an area of poor fit with its cluster. Meanwhile, a score close to -1 indicates that the cluster result is not appropriate (Shahapure & Nicholas, 2020).

RESULT AND DISCUSSION

1. Data Collection

Table 1. Village Population Data

ID	Occupation	Total Car Assets	Total Motorcycle Assets	Total Assets House / Land/ Rice Field	Revenue Category
K001	Self-employed	0	2	2	6
K002	Private Employee	0	1	1	4
…	…	…	…	…	…
…	…	…	…	…	…
K008	Private Employee	0	2	1	6

P-ISSN: 2622-6901
E-ISSN: 2622-6375

ID	Occupation	Total Car Assets	Total Motorcycle Assets	Total Assets House / Land/ Rice Field	Revenue Category
K009	Private Employee	0	2	1	5
K010	Self-employed	0	1	1	4
K011	Self-employed	0	2	1	4

Table 1 presents data on the residents of Kalipuro Village, consisting of 1,012 rows and 6 columns, where each row represents a household. The dataset includes key attributes such as resident identification numbers (IDs), types of occupations, and details on asset ownership, specifically the number of cars, motorcycles, and properties like houses, land, or rice fields. These variables offer valuable insights into the socioeconomic conditions of the village, revealing differences in asset accumulation and income potential among residents. By analyzing this dataset, researchers can better understand the overall financial landscape of Kalipuro Village, identify patterns of wealth distribution, and support more accurate clustering for aid prioritization.

Table 2. Revenue Range

Revenue Category	Nominal Revenue
1	≤ Rp. 500.000
2	Rp. 500.000 – Rp. 1.000.000
3	Rp. 1.000.000 – Rp. 2.500.000
4	Rp. 2.500.000 – Rp. 3.500.000
5	Rp. 3.500.000 – Rp. 4.500.000
6	Rp. 4.500.000 – Rp. 6.000.000
7	> Rp 6.000.000

Table 2 shows residents' income grouped into seven categories, providing a clearer picture of economic disparities within Kalipuro Village. The lowest income group (≤ Rp. 500,000) likely represents individuals engaged in informal, unstable, or seasonal jobs, with limited financial security. Meanwhile, a significant portion of residents fall within the Rp. 1,000,000 to Rp. 3,500,000 range, which may indicate a concentration of middle-income earners, such as farm laborers, small traders, or lower-level employees. These groups may have more consistent, but still modest, earnings. On the other end of the spectrum, those earning above Rp. 6,000,000 are relatively few and are likely employed in formal sectors, run their own businesses, or hold higher-paying professional roles. This classification helps identify the income gap among residents and supports the targeting of social aid to those most in need.

P-ISSN: 2622-6901
E-ISSN: 2622-6375

2. Exploratory Data Analysis (EDA)

- **Variable Distribution Analysis**

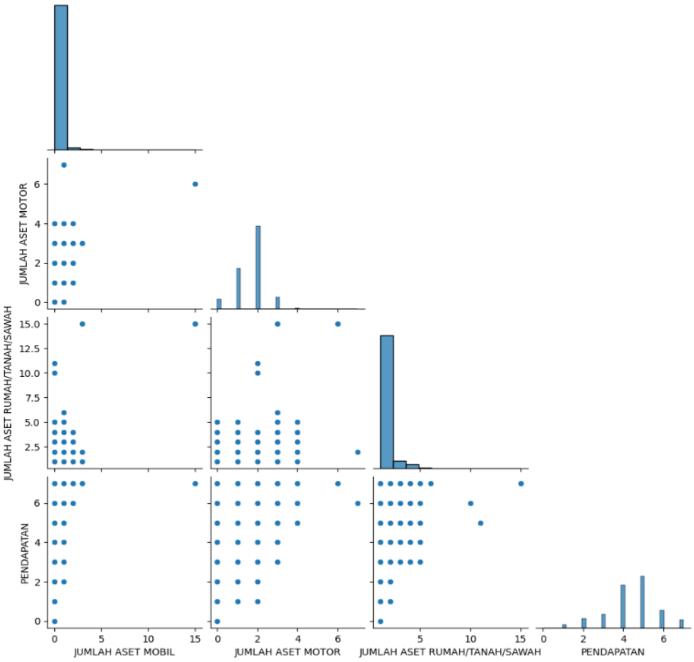

Figure 2. Pair plot showing the relationships between total assets

Figure 2 shows a pair plot of assets and income, highlighting that most incomes range from 1–3, with few reaching 6. House/land assets range from 1–5, motorbike assets center around 2, and car assets around 0–2. Outliers are also visible. Table 3 complements this by detailing the number of household heads by occupation, offering insight into the population's economic structure.

Table 3. Employment Data of Village Residents

No	Occupation Type	Head of Household Population
1	Private Employee	427
2	Self-employed	331
3	Farm Labourer / Plantation	84
4	Farmer/Grower	75
5	Housewife	26
6	Civil Servant	21
7	Teacher	16
8	Pensioners	7
9	Entrepreneur	7
10	The Indonesian National Armed Forces	6
11	Village Apparatus	5
12	Nurse	2
13	Indonesian National Police	2
14	Doctor	1
15	Head of Village	1
16	Honorary Employee	1

- **Correlation Analysis Between Variables**

  After analyzing data distribution, correlation analysis in Figure 3 shows moderate relationships, with the strongest between motorbike assets and income (0.59), followed by car-house/land (0.46), and car-income (0.38); others range from 0.29 to 0.34.

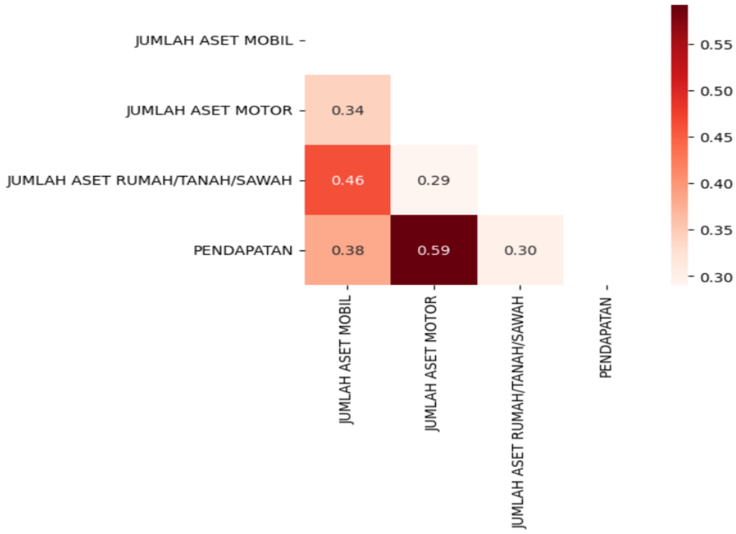

Figure 3. Heatmap Correlation

- **Descriptive Statistical Analysis**

  The results of the descriptive statistical analysis are shown in Table 4, which illustrates the distribution and characteristics of the data, including the mean, distribution, and minimum and maximum value.

Table 4. Statistical Descriptive

	Total Car Assets	Total Motorcycle Assets	Total Assets House / Land/ Rice Field	Revenue Category
Count	1012.000	1012.000	1012.000	1012.000
Mean	0.235178	1.696640	1.414032	4.460474
Std	0.655214	0.776681	1.057927	1.297269
Min	0.000000	0.000000	1.000000	0.000000
25%	0.000000	1.000000	1.000000	4.000000
50%	0.000000	2.000000	1.000000	5.000000
75%	0.000000	2.000000	1.000000	5.000000
Max	15.000000	7.000000	15.000000	7.000000

Table 4 shows descriptive stats for 1,012 samples. On average, people own 0.23 cars, 1.69 motorbikes, and 1.41 house/land units. Income averages 4.46. Each variable shows wide variation, with max values up to 15 and visible spread in percentile data.

3. Data Preprocessing

- **Missing Value Detection**

  Table 5 shows the results of missing value detection on the analysed variables, which indicates that there are no missing values in the data.

Table 5. Missing Value Detection

Variable	Value
ID	**0**
Occupation	**0**
Total Car Assets	**0**
Total Motorcycle Assets	**0**
Variable	**Value**
Total Assets House / Land/ Rice Field	**0**
Revenue Category	**0**

The detection results show that all variables analysed have complete data without any missing values. Thus, the data used in this analysis can be confirmed to be intact and ready for further processing.

- **Outlier Identification**

Table 6. Outlier Identification

Variable	Value
Total Car Assets	206
Total Motorcycle Assets	10
Total Assets House / Land/ Rice Field	246
Revenue Category	143

Table 6 shows detected outliers for each variable. However, the data was kept unchanged, as K-Medoids is robust to outliers and still yields effective results. This preserves data integrity and ensures the clustering reflects original patterns.

- **Data Normalization**

Table 7. Data Normalization

	ID	Total Car Assets	Total Motorcycle Assets	Total Assets House / Land/ Rice Field	Revenue Category
0	K001	0.000	0.285	0.357	5.142
1	K002	0.000	0.142	0.000	3.428
2	K003	0.000	0.285	0.000	4.285
…	…	…	…	…	…
1007	S262	0.266	0.428	0.000	4.285
1008	S263	0.000	0.285	0.000	5.142
1009	S264	0.000	0.285	0.000	4.285

Table 7 shows normalized data for assets and income using Min-Max Scaling. This ensures all variables are on a 0–1 scale, reducing bias from differing value ranges. Weighted normalization keeps important variables like income more influential. Higher values reflect greater income or asset ownership.

4. Implementation PSO and K-Medoids Clustering

The implementation of K-Medoids clustering optimised by Particle Swarm Optimization (PSO) was performed using Python programming language. The Scikit-learn library was used to implement K-Medoids, while PySwarms was applied to optimise the selection of medoids. An initial set of medoids was randomly chosen, and clustering performance was evaluated using the Silhouette Score.

Table 8. Cluster Centre Point Result

The Best Medoid	0	609	1011	578	195
The best cost	268.1676231473807				

Table 8 shows the results of the Particle Swarm Optimization (PSO) algorithm implementation run with a maximum of 100 iterations to find the optimal medoid combination. During the iteration process, PSO managed to find the best solution at the 18th iteration with a combination of medoids [0, 609, 1011, 578, 195] which resulted in a minimum total distance cost of 268.1676231473807. Although the best solution has been found at the 18th iteration, the algorithm still continues the process until the 100th iteration to ensure there is no better solution.

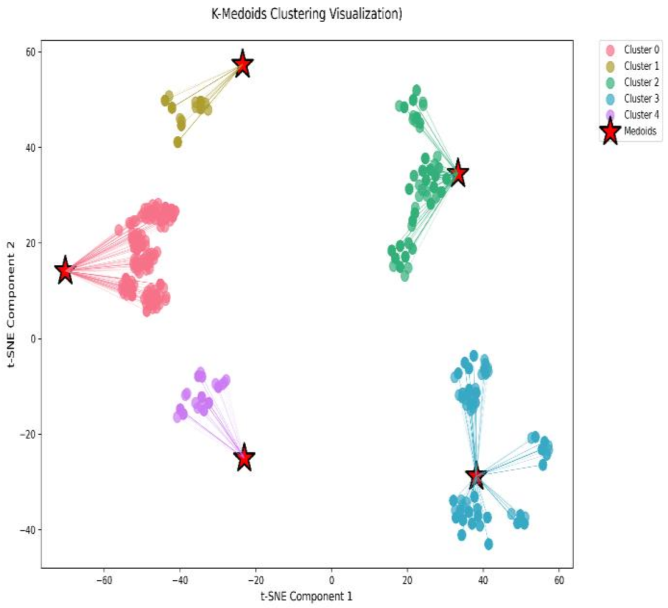

Figure 4. Cluster Visualization

Figure 4 shows the t-SNE visualization of the K-Medoids clustering (with PSO initialization), dividing data into 5 color-coded clusters. Red stars mark the medoids as cluster centers. The clear separation supports better decision-making in aid distribution.

Table 9. Cluster Distribution

Cluster	Cluster Total
0	179
1	89
2	296
3	354
4	94

Table 9 shows the distribution of households across five clusters: Cluster 0 consists of 179 households, Cluster 1 includes 89, Cluster 2 has 296, Cluster 3 contains 354, and Cluster 4 holds 94 households. These clusters represent groups of aid beneficiaries who share similar characteristics based on variables such as income and asset ownership. The clustering was generated using the Particle Swarm Optimization (PSO) and K-Medoids algorithms, which work together to ensure accurate grouping by optimizing cluster centers and minimizing intra-cluster differences. This classification supports more targeted and equitable aid distribution by clearly distinguishing population segments with differing levels of need.

5. Evaluation Model

Table 10 Evaluation Model

Index	Value
Silhouette Score	0.6531

Table 10 shows a Silhouette Score of 0.6531, indicating that the clustering model has a fairly good separation between groups. While not perfect, the score suggests the clusters are distinct, with some room for improvement.

CONCLUSION

The clustering results using the Particle Swarm Optimization (PSO) and K-Medoids methods show that the data is divided into five clusters with the following distribution: Cluster 0 consists of 179 households, Cluster 1 of 89 households, Cluster 2 of 296 households, Cluster 3 of 354 households, and Cluster 4 contains 94 households. The Silhouette Score value of 0.6531 indicates that the cluster structure has good cohesion and a fairly clear separation between groups.

Based on further analysis, Cluster 1 was identified as the highest-priority group for aid recipients, followed by Clusters 4, 2, 3, and 0, respectively. This means that while technically labeled as "Cluster 1" by the tool, it represents the first-priority group in interpretation.

The application of the K-Medoids algorithm optimised with PSO effectively grouped the data, supporting a more accurate and targeted aid distribution strategy. Moreover, the results demonstrate that this method improves decision-making efficiency in resource allocation.

SUGGESTION

Based on the clustering analysis results that show the effectiveness of the Particle Swarm Optimization and K-Medoids methods, it is recommended that future researchers use other evaluation methods to improve understanding of the characteristics of beneficiaries. In addition, sensitivity analysis needs to be carried out to evaluate the impact of changes in clustering parameters on the results. Regular monitoring and evaluation of the effectiveness of aid distribution is very important, in addition to providing training and socialisation to relevant officers regarding the use of clustering analysis results in the aid distribution process.

REFERENCES

Alghifari, F., & Juardi, D. (2021). Penerapan Data Mining Pada Penjualan Makanan Dan Minuman Menggunakan Metode Algoritma Naïve Bayes. *Jurnal Ilmiah Informatika, 9*(02), 75–81. https://doi.org/10.33884/jif.v9i02.3755

BPS. (2024, July). *Persentase Penduduk Miskin Maret 2024 turun menjadi 9,03 persen. - Badan Pusat Statistik Indonesia*.

https://www.bps.go.id/id/pressrelease/2024/07/01/2370/persentase-penduduk-miskin-maret-2024-turun-menjadi-9-03-persen-.html

Buulolo, E., Syahputra, R., & Fau, A. (2020). Algoritma K-Medoids Untuk Menentukan Calon Mahasiswa Yang Layak Mendapatkan Beasiswa Bidikmisi di Universitas Budi Darma. *Jurnal Media Informatika Budidarma*, *4*(3), 797. https://doi.org/10.30865/mib.v4i3.2240

EMedia RI. (2024). *Banyak Penyaluran Bansos Tak Tepat Sasaran, Legislator Minta Kemesos Segera Perbaiki DTKS - EMedia DPR RI*. https://emedia.dpr.go.id/2024/05/21/banyak-penyaluran-bansos-tak-tepat-sasaran-legislator-minta-kemesos-segera-perbaiki-dtks/

Hu, Z. (2024). Comparison of K-Means, K-Medoids and K-Means++ algorithms based on the Calinski-Harabasz index for COVID-19 epidemic in China. *Applied and Computational Engineering*, *49*(1), 11–20. https://doi.org/10.54254/2755-2721/49/20241046

Kemenko Perekonomian. (2018). *Program Bantuan Sosial Untuk Rakyat - Kementerian Koordinator Bidang Perekonomian Republik Indonesia*. https://ekon.go.id/publikasi/detail/1281/program-bantuan-sosial-untuk-rakyat

Muhima, R. R., Kurniawan, M., Wardhana, S. R., Yudhana, A., Sunardi, Rahmawati, W. M., & YuliastuΘ, G. E. (2018). *Kupas Tuntas Algoritma Clustering Konsep, Perhitungan Manual dan Program*.

Musaddad, A. A., & Kriswibowo, A. (2021). Optimalizing Big Data in Reducing Miss-Targeting Family Hope Program (PKH) in Sidoarjo Disctrict with Approach Machine Learning. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, *15*(1), 99. https://doi.org/10.22146/ijccs.62589

Novianti, A., Afnan, I. M., Utama, R. I. B., & Widodo, E. (2021). Grouping of Districts Based on Poverty Factors in Papua Province Uses The K-Medoids Algorithm. *Enthusiastic : International Journal of Applied Statistics and Data Science*, *1*(2), 94–102. https://doi.org/10.20885/enthusiastic.vol1.iss2.art6

Pratiwi, S. H., Witanti, W., Hendro, T., Achmad, U. J., & Abstract, Y. (2024). Optimasi Penentuan Vendor Untuk Material Pesawat Menggunakan Algoritma Particle Swarm Optimization. *Jurnal Ilmiah Wahana Pendidikan, Februari*, *2024*(4), 825–837. https://doi.org/10.5281/zenodo.10537168

Radhi, M., Amalia, A., Sitompul, D. R. H., Sinurat, S. H., & Indra, E. (2022). Analisis Big Data Dengan Metode Exploratory Data Analysis (Eda) Dan Metode Visualisasi Menggunakan Jupyter Notebook. *Jurnal Sistem Informasi Dan Ilmu Komputer Prima (JUSIKOM PRIMA)*, *4*(2), 23–27. https://doi.org/10.34012/jurnalsisteminformasidanilmukomputer.v4i2.2475

Rispandi. (2023). Implementasi Metode K-Medoids Untuk Clustring Penerima Bantuan Berdasarkan Normalisasi Data Masyarakat Miskin Dengan Metode Desimal Scaling. *Journal of Computing and Informatics Research*, *3*(1), 141–152. https://doi.org/10.47065/comforch.v3i1.1063

Shahapure, K. R., & Nicholas, C. (2020). Cluster quality analysis using silhouette score. *Proceedings - 2020 IEEE 7th International Conference on Data Science and Advanced Analytics, DSAA 2020*, 747–748. https://doi.org/10.1109/DSAA49011.2020.00096

Trimono, Farid, N., & Habibi, M. (2023). Statistika Deskriptif Pada Analisis Ketimpangan Kemiskinan (Studi Kasus Data Kemiskinan Wilayah Perkotaan dan Pedesaan Seluruh Provinsi di Indonesia periode tahun 2021-2022). *Seminar Nasional Sains Data*, *2023*(Senada), 253–259.

Wijaya, D. B., Noersasongko, E., & Purwanto, P. (2024). Optimasi Centroid Awal Algoritma K-Medoids Menggunakan Particle Swarm Optimization Untuk Segmentasi Customer. *Techno.Com*, *23*(1), 221–232. https://doi.org/10.62411/tc.v23i1.9516