Journal of Applied Informatics and Computing (JAIC)
Vol.10, No.2, April 2026, pp. 1347~1358
e-ISSN: 2548-6861

Comparison of K-Means and K-Medoids for PKH Priority Clustering

M. Ahsanul Fikri [1]*, Ifnu Wisma Dwi Prastya [2]*, Alif Yuanita Kartini [3]**
* Department of Informatics Engineering, Universitas Nahdlatul Ulama Sunan Giri
** Department of Statistics, Universitas Nahdlatul Ulama Sunan Giri
ahsanulfikri26@gmail.com [1], ifnudwi867@gmail.com [2], alifyuanita@unugiri.ac.id [3]

Article Info

Article history:

Received 2026-01-15
Revised 2026-02-23
Accepted 2026-04-08

Keyword:

Clustering,
Family Hope Programme,
K-Means,
K-Medoids,
Silhouette Score.

ABSTRACT

The Family Hope Program (PKH) is a government policy aimed at poverty alleviation and protecting vulnerable groups. Differences in the number and composition of PKH recipients between sub-districts indicate variations in the level of social vulnerability and potential regional inequality that require objective mapping. This study aims to identify the concentration of socially vulnerable groups and inequality between sub-districts in Bojonegoro Regency through a data-based cluster approach. The data used comes from the Bojonegoro Regency Social Service which includes PKH recipient components, namely early childhood, SD, SMP, SMA, people with severe disabilities, the elderly, and pregnant women. This study compares several clustering scenarios including the K-Means and K-Medoids algorithms, variations in the number of clusters (k = 2 and k = 3), Euclidean and Manhattan distance functions, and data conditions without normalization and with Min–Max normalization. The determination of the number of clusters is done using the Elbow method, while cluster quality is evaluated using the Silhouette Score. The results of the study show that the best scenario was obtained in K-Medoids with Euclidean distance and Min–Max normalization at k = 3 with a Silhouette Score value of 0,8115. The clustering results divided the sub-districts into three categories, namely high clusters (many) with 10 sub-districts, medium with 13 sub-districts, and low (few) with 5 sub-districts. These findings can be used as a basis for determining the priority of policy interventions, especially in high clusters, so that the distribution of PKH becomes more targeted and is able to reduce social inequality between regions.

I. Introduction

The Family Hope Program (PKH) is a form of conditional social assistance designed by the government to reduce poverty and improve the quality of life of beneficiary families [1]. Since its launch, the Family Hope Program (PKH) has proven to have a positive impact on improving access to education, healthcare, and the well-being of poor communities in Indonesia. However, its implementation still faces various challenges, particularly related to differences in characteristics and levels of social vulnerability across regions, which have not been comprehensively mapped [2]. Regional variations in socio-economic conditions, such as differences in the number of school-age children, elderly people, people with disabilities, and pregnant women in beneficiary families, indicate the existence of social inequalities that have the potential to affect the effectiveness of the program [3]. A similar situation also occurred in Bojonegoro Regency. Based on data from the Bojonegoro Regency Social Service for 2025, differences in the number and composition of PKH recipients across sub-districts are evident, indicating an uneven concentration of vulnerable social groups. Therefore, a more systematic, data-driven approach is needed to map social vulnerability and regional disparities between sub-districts as a basis for planning more equitable social assistance policies.

In the context of data analysis, clustering is an unsupervised learning technique used to group objects based on similar characteristics without requiring prior labels [4]. The K-Means algorithm, first introduced by MacQueen in

1967 [5], is a very popular method due to its conceptual simplicity and computational efficiency in handling large datasets [6][7]. On the other hand, K-Medoids was developed with a similar approach, but uses medoids as cluster centers so that it is more resistant to the influence of outliers and extreme variations in the data [8][9]. These two algorithms have high relevance when applied to complex socio-economic data, such as the grouping of priority sub-districts for social assistance recipients, where a balance between efficiency and accuracy is needed.

Several previous studies in the socio-economic field have used more predictive approaches such as linear regression and decision trees. For example, in the study conducted by D. E. Waluyo, H. W. Kinasih, C. Paramita, D. Pergiwati, R. Nohan, and F. A. Rafrastara [10] Studies have shown that linear regression has higher accuracy than Decision Tree, Random Forest, and k-Nearest Neighbor (kNN) in predicting quantitative variables, such as health index or income level. However, these predictive approaches generally require labeled data with a clear target variable. The characteristics of regional aggregate data, such as fluctuating and unlabeled sub-district socio-economic data, make predictive methods less suitable, so clustering-based approaches are considered more relevant.

Comparisons of various clustering methods have been widely conducted to examine the performance of each algorithm in both social and technical contexts. A study by R. Afifa, M. I. Mazdadi, T. H. Saragih, F. Indriani, and M. Muliadi [11] compared K-Means with PCA, and showed that K-Means produced better evaluation values. In the research of E. Setiawati, U. D. Fernanda, S. Agesti, M. Iqbal, and M. O. A. Herjho [12] compared K-Means, and DBSCAN and found that the choice of method is highly dependent on the structure and density of the data. However, a study by J. Heidari, N. Daneshpour, and A. Zangeneh [13] K-Means has a major weakness, namely high sensitivity to outliers and extreme values, because this method uses the average (mean) as the cluster center. The presence of outliers can significantly affect the position of the centroid, causing the clustering results to be less stable and less representative of the actual data pattern. This condition becomes increasingly important when the data has high variation or there are significant differences in values between objects. In contrast, K-Medoids is more robust against outliers because it uses the original data objects as cluster centers (medoids), not the average value. This approach makes the position of the cluster center less likely to shift due to extreme values. Therefore, K-Medoids is considered superior and more stable in handling datasets with high variation, uneven distribution, and the possibility of outliers, so the obtained clustering results are more representative and reliable. In a study by S. Fahad, F. Su, S. U. Khan, M. R. Naeem, and K. Wei [14] confirmed that Hierarchical Clustering is more stable for non-linear data, and A. Gupta, H. Sharma, and A. Akhtar [15] discovered the advantages of K-Means in terms of efficiency and speed. In addition, K-Means continues to be a relevant basic algorithm in various practical data analysis applications, as shown by K. P. Sinaga and M.-S. Yang [16] and W. Yang [17].

In the K-Means method, the cluster center is represented by the centroid, which is the average value of all the data in a cluster. The centroid is not always the original data, but rather an abstract calculated value [18]. This approach makes K-Means more sensitive to extreme values or outliers, since small changes in the data can significantly affect the position of the mean. In contrast, in the K-Medoids method, the cluster center is represented by a medoid, which is one of the original data objects that has the smallest total distance from the members of that cluster. Because medoids are real data, the position of the cluster center is more stable and less easily affected by extreme values [19]. This difference in representation shows that K-Medoids is more robust and more suitable for use on datasets that have high variation or potentially contain outliers [20]. Therefore, the selection of the K-Medoids method in this study has a strong methodological basis for producing more representative and stable clustering.

In addition to comparative research, several studies have also directly implemented K-Means and K-Medoids in real-world cases. K-Means was used to identify crime-prone areas in Indonesia in a study by R. H. Maharrani, P. D. Abda'u, and M. N. Faiz [21], on the other hand H. Wahyono, H. Setiaji, T. Hartati, and N. Wiliani [22] using a similar method to identify areas prone to traffic accidents. Another study in the research of Y. Candra, R. Goejantoro, and A. T. R. Dani [23] grouping provinces in Indonesia based on economic and crime indicators, while research by R. Khairuna, Nurdin, and Ar Razi [24] applied K-Medoids to handle social data with outliers. These results demonstrate that both algorithms have high flexibility for application to aggregate regional data, making them relevant for use in the context of mapping sub-districts receiving the Family Hope Program (PKH). Previous research has shown that both K-Means and K-Medoids methods perform well in data clustering and are capable of producing optimal clusters across various datasets. Therefore, this study uses these clustering methods to group data based on the level of similarity in their characteristics. The application of these methods is expected to produce more optimal clustering, be stable against data variations, and be able to describe data distribution patterns more clearly and informatively.

This study aims to identify the concentration of socially vulnerable groups and social inequality between sub-districts in Bojonegoro Regency through a data-driven regional clustering approach. The analysis was conducted by comparing the performance of the K-Means and K-Medoids algorithms in grouping sub-districts based on the characteristics of Family Hope Program (PKH) recipients. The number of clusters was determined using the Elbow method, considering $k = 2$ and $k = 3$ as a representation of variations in social conditions between regions [25]. The clustering process was implemented using two distance functions, namely Euclidean and Manhattan, and considering data conditions without normalization and with Min–Max Scaling normalization. Evaluation of the quality of clustering results was carried out using the Silhouette Score to measure the level of cohesion and separation between clusters. The results of this study are expected to illustrate patterns of social vulnerability and regional inequality between sub-districts, as well as contribute to academics in the application of data mining techniques to support regional socio-economic policies. Furthermore, the resulting clustering model can be implemented in a local government Decision Support System (DSS), where PKH recipient data is regularly updated and automatically processed to group sub-districts into high, medium, and low vulnerability categories. Through this integration, the government can utilize the clustering results as a basis for determining aid allocation priorities, monitoring changes in social conditions, and implementing policy effectiveness periodically. Thus, this study not only contributes theoretically but also applies in supporting more targeted, transparent, and data-driven decision-making, particularly in protecting vulnerable groups of Family Hope Program (PKH) recipients.

II. RESEARCH METHODOLOGY

At this stage, the steps taken to achieve the research objectives are explained, which must be carried out systematically, in detail and in an organized manner in order to achieve the research objectives as expected. This research framework is depicted in Figure 1.

This research begins with a literature study, data processing, and continues with data preprocessing, namely data reduction so that it can be processed using a comparison of K-Means and K-Medoids by calculating distances using Euclidean distance and Manhattan distance, then evaluated using the Silhouette Score. After the clusters are found, analysis of the clusters and their data is carried out to produce results and conclusions.

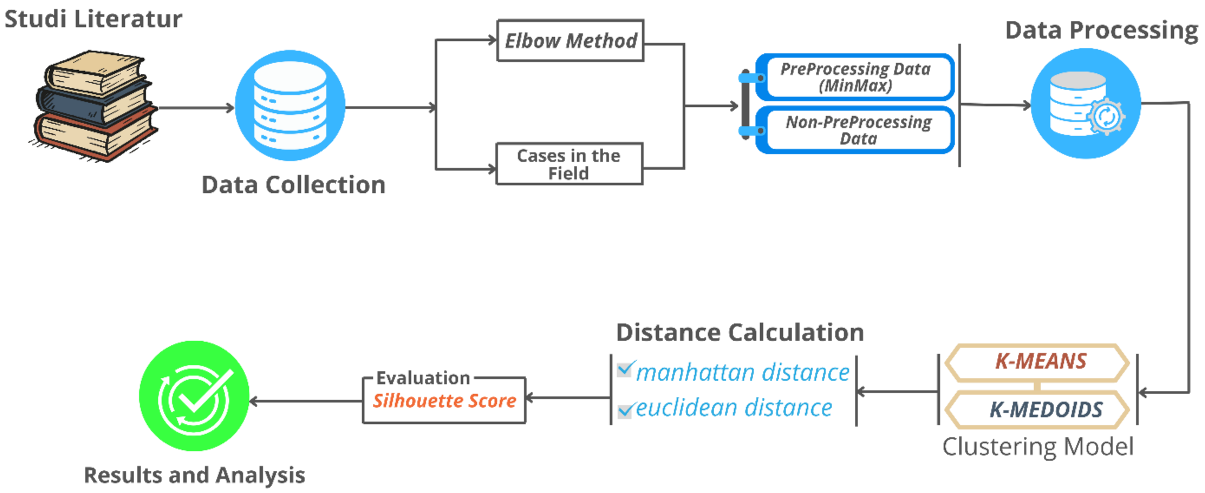

Figure 1.Research process

A. Dataset Collection

The dataset for this study comes from the Bojonegoro Regency Social Service with a data period of 2025. This dataset includes attributes such as AUD (number of Early Childhood), SD (number of elementary school-aged children), SMP (number of junior high school-aged children), SMA (number of high school-aged children), DB (number of people with disabilities), Lansia (number of elderly residents), and Hamil (number of pregnant women) relevant to the clustering process. Available data in the form of official reports and digital documents were collected through documentation methods. This study has several limitations. The data used are administrative data, thus potentially containing biases, such as possible delays in data updates, inaccurate reporting at the field level, or differences in recording standards between regions. Furthermore, the indicators used in this study are still limited to the demographic characteristics of vulnerable groups and do not include more comprehensive socioeconomic indicators. Therefore, the clustering results obtained need to be interpreted with these limitations in mind. A detailed explanation of each feature is shown in Table I.

TABLE I
FEATURE EXPLANATION

Name Features	Explanation
AUD	Shows the number of members in Family Card (KK) of PKH recipients who are included in the Early Childhood category (0–6 years).
SD	Shows the number of members in Family Card (KK) of PKH recipient which includes elementary school age children (±7–12 years).
SMP	Shows the number of members in Family Card (KK) of PKH recipient which includes children of junior high school age (±13–15 years).

SMA	Shows the number of members in Family Card (KK) of PKH recipient which includes high school aged children (±16–18 years).
DB	Shows the number of members in in Family Card (KK) of PKH recipient who are people with disabilities.
Lansia	Shows the number of members in Family Card (KK) of PKH recipient who are elderly (≥60 years).
Hamil	Shows the number of members in Family Card (KK) of PKH recipient who are pregnant.

B. MinMaxScaler Normalization

Before clustering is performed, the data is first processed by normalization using the Min–Max Normalization method. MinMaxScaler normalization has advantages in high-dimensional data because it ensures that all data in the database has the same range. This technique is especially important when the data is unstructured and contains highly divergent values [26]. The goal of this step is to equalize the scale of all numeric attributes so that they fall within the same value range, namely between 0 and 1. Because clustering algorithms such as K-Means and K-Medoids rely heavily on distance calculations, normalization is necessary. Differences in attribute scales can cause attributes with larger values to dominate the clustering process. As a result, the clustering results become less representative and biased.

$$x_1 = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

Explanation**:**

- x_1= normalized result value
- x= original data value
- x_{min}= minimum value of an attribute
- x_{max}= maximum value of an attribute

The minimum and maximum values of each attribute are used to transform the values. Therefore, the original data distribution is still maintained, but on the same scale. This method is particularly suitable for use with socioeconomic data that has a wide range of values between variables, such as the number of aid recipients, income levels, or regional welfare indicators.

C. Elbow Method

Before the data clustering process began, the Elbow method was used to determine the ideal number of clusters for this study. Experimental methods and quantitative approaches were used in this study. To find the best cluster structure, numerical data from the preprocessing stage was analyzed. The Elbow Method calculates the level of cluster compactness by calculating the Within Cluster Sum of Squares (WCSS) value for various numbers of clusters [27].

In the initial stage, the range of clusters to be tested is determined, namely K = 1 to the highest value adjusted to the research data. The K-Means algorithm is used to combine all data into clusters for each value of K. This process combines data into clusters based on the closest distance to the centroid, the center of the cluster. After this process is complete, the WCCS value is calculated, which shows the total squared distance between each data point and the cluster centroid [28].

$$WCCS = \sum_{k=1}^{K} \sum_{x_i \in C_k} \| x_i - \mu_k \|^2 \quad (2)$$

To determine the ideal number of clusters, we observed the inflection point, or the point at which the WCCS value began to decline significantly. In the next clustering stage of this study, the K value was considered to be the most ideal number of clusters.

D. Algorithm K-Means and K-Medoids

K-Means is a clustering method used to divide data into a number of k clusters based on similarities in characteristics between data points. Each cluster is represented by a single centroid (center point). After determining the distance between the centroid and the data, the data is grouped into the closest cluster. The advantages of this method are that it is fast, simple, and effective for large data sets. After all the data has been grouped, new centroids are calculated using the average of all cluster members [29][30].

$$C_j = \frac{1}{n_j} \sum_{i=1}^{n_j} x_i^{(j)} \quad (3)$$

Explanation:

- c_j= new centroid in cluster j
- n_j= amount of data in a cluster j
- $x_i^{(j)}$= the i-th data in the cluster j

Meanwhile, K-Medoids Clustering is a clustering method similar to K-Means, but the cluster centers used are medoids, which are actual data (not average values). Medoids are chosen because they have the minimum total distance to other cluster members [31]. This method is more resistant to outliers (extreme data) because it is not affected by average values that can change drastically due to extreme data. K-Medoids attempts to minimize the total absolute distance between all data and its cluster medoids. If the exchange of medoids causes a decrease in the J value, then a new medoid will be selected.

E. Euclidean Distance and Manhattan Distance

Euclidean Distance is a measure of the geometric distance between two points in n-dimensional space. This distance is calculated based on the Pythagorean theorem and describes the “straight” distance between two points. This method is commonly used with numerical data because it can show the realistic proximity between objects. Here is the formula for Euclidean Distance [32].

$$d(p,q) = \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2} \quad (4)$$

Explanation:

- p_i, q_i= the value of the i-th attribute of objects p and q,
- n= number of attributes used.

The smaller the distance value d(p,q), the greater the similarity between the data. Meanwhile, Manhattan Distance calculates the distance between two points by adding the absolute differences between their attributes. This distance is like calculating the distance traveled through a grid-shaped road (such as in Manhattan). This method is more resistant to outliers than Euclidean Distance. Here is the formula for Manhattan Distance [33] :

$$d(p,q) = \sum_{i=1}^{n} |p_i - q_i| \quad (5)$$

Explanation:

- p_i, q_i= the value of the i-th attribute of objects p and q,
- n= number of attributes used.

The smaller the distance value d(p,q), the greater the similarity between the data.

F. Evaluation

The evaluation of clustering results is performed using the Silhouette Score. This method is used to measure how well data is placed in its cluster. The Silhouette value indicates how close the data is to its own cluster compared to other clusters. Silhouette Score formula [34] :

$$S_i = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (6)$$

Explanation**:**

- $a(i)$= average distance between data i and all other data in the same cluster.
- $b(i)$= average distance between data i and all data in the nearest cluster.
- $S(i)$= silhouette value for the i-th data point.

Interpretation of Silhouette values:

- $S(i) \approx +1$→ The data are in the right cluster (correct).
- $S(i) \approx 0$→ data is on the boundary between clusters.
- $S(i) \approx -1$→ data on possible misplacement of clusters.

The average Silhouette Score is used to evaluate the overall quality of the clustering results. The higher the average score (closer to 1), the better the clustering results produced by the algorithm.

III. Result and Discussion

This stage explains the structure and attributes of the dataset used, as data characteristics are important factors influencing the quality of the resulting clusters. The dataset used is aggregate data at the sub-district level, with attributes including AUD (number of Early Childhood Children), SD (number of Elementary School-aged children), SMP (number of Middle School-aged children), SMA (number of High School-aged children), DB (number of people with disabilities), Lansia (number of elderly residents), and Hamil (number of pregnant women) in each sub-district. Distance-based clustering methods such as K-Means and K-Medoids are used. The following datasets used in this study are shown in Table II.

TABLE II
ORIGINAL DATASET

No	Subdistrict	AUD	SD	SMP	SMA	DB	Lansia	Hamil
1	Balen	409	944	669	743	185	2209	3
2	Baureno	306	940	689	740	222	1859	2
3	Bojonegoro	151	457	321	332	68	763	1
4	Bubulan	92	210	132	141	59	655	0
5	Dander	303	899	679	687	158	2183	5
6	Gayam	185	394	291	343	60	930	0
7	Gondang	116	355	293	258	69	1114	2
8	Kalitidu	227	601	458	466	133	1331	0
9	Kanor	322	811	566	609	158	1896	1
10	Kapas	202	547	411	392	129	1172	4
11	Kasiman	81	295	231	244	92	868	0
12	Kedewan	43	123	87	106	26	381	1
13	Kedungadem	347	998	728	815	213	3630	8
14	Kepohbaru	350	1028	756	810	171	2572	6
15	Malo	234	580	390	468	111	1207	4
16	Margomulyo	150	511	390	409	100	1163	3
17	Ngambon	47	115	78	121	21	424	0
18	Ngasem	312	772	600	763	173	2292	2
19	Ngraho	415	942	746	819	145	1843	11
20	Padangan	184	606	409	481	115	983	1
21	Purwosari	184	492	359	450	92	1174	2

22	Sekar	197	493	345	437	77	1146	8
23	Sugihwaras	224	565	399	400	126	1584	3
24	Sukosewu	183	436	332	379	91	1275	1
25	Sumberejo	345	896	646	699	201	2262	5
26	Tambakrejo	302	876	656	655	127	2054	0
27	Temayang	102	284	174	278	52	859	1
28	Trucuk	178	417	305	330	57	997	4

A. Preprocessing

The obtained dataset consists of 28 districts representing all sub-districts in Bojonegoro Regency based on the 2025 period. The preprocessing stage was used to improve data quality and readiness. The data was processed for analysis and grouping. The data collection first underwent a preprocessing process aimed at improving data quality and readiness. The data preprocessing stage, which included adjusting the scale and data consistency, is shown in Table III.

TABLE III
AFTER MINMAX SCALLER

No	Subdistrict	AUD	SD	SMP	SMA	DB	Lansia	Hamil
1	Balen	0,9838	0,9079	0,8716	0,8934	0,8159	0,5626	0,2727
2	Baureno	0,7069	0,9036	0,9011	0,8892	1,0000	0,4549	0,1818
3	Bojonegoro	0,2903	0,3745	0,3584	0,3169	0,2338	0,1175	0,0909
4	Bubulan	0,1317	0,1040	0,0796	0,0490	0,1890	0,0843	0,0000
5	Dander	0,6989	0,8587	0,8864	0,8148	0,6815	0,5546	0,4545
6	Gayam	0,3817	0,3055	0,3141	0,3323	0,1940	0,1689	0,0000
7	Gondang	0,1962	0,2628	0,3171	0,2131	0,2388	0,2256	0,1818
8	Kalitidu	0,4946	0,5323	0,5604	0,5049	0,5572	0,2923	0,0000
9	Kanor	0,7500	0,7623	0,7197	0,7054	0,6815	0,4662	0,0909
10	Kapas	0,4274	0,4731	0,4911	0,4011	0,5373	0,2434	0,3636
11	Kasiman	0,1021	0,1971	0,2256	0,1935	0,3532	0,1498	0,0000
12	Kedewan	0,0000	0,0087	0,0132	0,0000	0,0248	0,0000	0,0909
13	Kedungadem	0,8172	0,9671	0,9587	0,9943	0,9552	1,0000	0,7272
14	Kepohbaru	0,8252	1,0000	1,0000	0,9873	0,7462	0,6743	0,5454
15	Malo	0,5134	0,5093	0,4601	0,5077	0,4477	0,2542	0,3636
16	Margomulyo	0,2876	0,4337	0,4601	0,4249	0,3930	0,2406	0,2727
17	Ngambon	0,0107	0,0000	0,0000	0,0210	0,0000	0,0132	0,0000
18	Ngasem	0,7231	0,7196	0,7699	0,9214	0,7562	0,5881	0,1818
19	Ngraho	1,0000	0,9058	0,9852	1,0000	0,6169	0,4499	1,0000
20	Padangan	0,3790	0,5377	0,4882	0,5259	0,4676	0,1852	0,0909
21	Purwosari	0,3790	0,4129	0,4144	0,4824	0,3532	0,2440	0,1818
22	Sekar	0,4139	0,4140	0,3938	0,4642	0,2786	0,2354	0,7272
23	Sugihwaras	0,4865	0,4928	0,4734	0,4123	0,5223	0,3702	0,2727
24	Sukosewu	0,3763	0,3515	0,3746	0,3828	0,3482	0,2751	0,0909
25	Sumberejo	0,8118	0,8554	0,8377	0,8316	0,8955	0,5789	0,4545
26	Tambakrejo	0,6962	0,8335	0,8525	0,7699	0,5273	0,5149	0,0000
28	Trucuk	0,3629	0,3307	0,3348	0,3141	0,1791	0,1895	0,3636

The Min-Max normalization process is carried out by transforming each attribute value into a range of 0-1 based on the minimum and maximum values for each attribute. In the Balen District data, before normalization, the attribute values used were still on different scales, namely 409, 944, 669, 743, 185, and 2209. After the Min-Max normalization process was applied, these values changed to 0,9838; 0,9079; 0,8716; 0,8934; 0,8159; and 0,5626. This method works by scaling each value based on the minimum and maximum values for each attribute. Through this process, all attributes are on a uniform scale without eliminating differences or relative comparisons between the original values. The use of the Min-Max Scaler ensures that no attribute dominates the distance calculation process due to having a larger value range. Thus, the normalized data becomes more representative, proportional, and balanced for use in the analysis process. This condition allows the clustering method to produce groupings that are more accurate, objective, and reflect actual data patterns, because each attribute has an equal influence in determining the closeness between objects.

B. Determining the Number of Clusters (k Value)

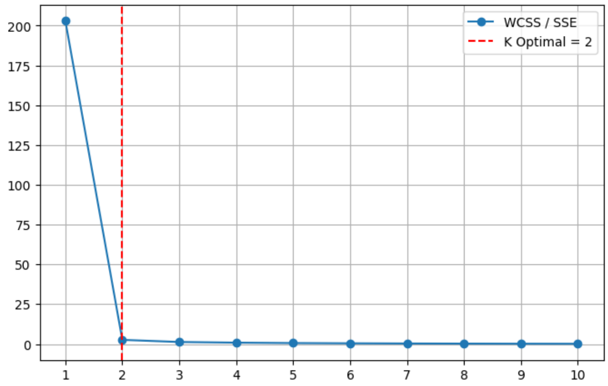

Figure 2 result elbow method

At this stage, the results of the Within Cluster Sum of Squares (WCSS) calculation using the Elbow method, elbow point, or elbow point, were obtained, which was most clearly found at k = 2. At this value, there was a significant decrease in WCSS compared to the previous k value, but increasing the number of clusters after k = 2 did not provide a significant decrease in WCSS. This shows that dividing the data into two clusters mathematically and structurally is sufficient to show data variations throughout the world.

However, observations and field needs indicate that the analyzed data are conceptually more relevant when grouped into three main categories: high, moderate, and low. These categories indicate varying levels of urgency and intervention needs, which cannot be comprehensively represented by just two clusters.

Therefore, this study not only relies on the results of the Elbow method in determining the number of clusters, but also considers the results of observations and field needs. The determination of the number of clusters is then carried out by comparing two scenarios: k = 2 based on the elbow point of the Elbow method and k = 3 based on conceptual considerations of the observation results. This comparison aims to assess the differences in the characteristics of the resulting clustering and determine the most representative number of clusters, both mathematically and practically, in describing the data conditions.

C. Clustering

In this stage, the clustering process for sub-districts receiving the Family Hope Program (PKH) in Bojonegoro Regency was conducted through eight test scenarios. Each scenario combines a clustering algorithm, a distance function, and a data normalization technique, aiming to obtain the most optimal clustering results for mapping social vulnerability and regional inequality between sub-districts. The eight clustering scenarios used in this study are as follows :

- K-Means uses Euclidean Distance without normalization.
- K-Means uses Euclidean Distance with Min–Max Scaler normalization.
- K-Means uses Manhattan Distance without normalization.
- K-Means uses Manhattan Distance with Min–Max Scaler normalization.
- K-Medoids uses Euclidean Distance without normalization.
- K-Medoids uses Euclidean Distance with Min–Max Scaler normalization.
- K-Medoids uses Manhattan Distance without normalization.
- K-Medoids uses Manhattan Distance with Min–Max Scaler normalization.

The following presents the results of the clustering calculations for sub-districts receiving the Family Hope Program (PKH) in Bojonegoro Regency based on two calculation mechanism scenarios. Details of these results are presented in Tables IV and V, which show the sub-district clustering results for each scenario.

TABLE IV
K-MEANS WITH MANHATTAN DISTANCE NON-NORMALIZATION

Subdistrict	C1	C2	C3	Shortest Distance	Cluster	a(i)	b(i)	s(i)
Balen	196,9000	2270,1111	1743,8833	196,9000	C1	196,9000	1743,8833	0,8871
Baureno	556,1000	1868,1111	1459,9948	556,1000	C1	556,1000	1459,9948	0,6191
Bojonegoro	3025,9000	798,8888	220,8775	220,8775	C3	220,8775	798,8888	0,7235
Bubulan	3827,9000	1600,8888	196,0431	196,0431	C3	196,0431	1600,8888	0,8775
Dander	217,1000	2022,1111	1675,8804	217,1000	C1	217,1000	1675,8804	0,8704
...	...	...	...	...	...	...	...	...
Trucuk	2830,9000	605,8888	291,3392	291,3392	C3	291,3392	605,8888	0,5191

Based on the clustering results using the K-Means algorithm with Manhattan Distance without Min–Max normalization, Balen District has the closest distance to Cluster 1 (C1) with a Shortest Distance value of 196,9000. This value is smaller than the distance of Balen to C2 2270,1111 and C3 1743,8833, so Balen District is objectively grouped into Cluster C1.

The a(i) value of 196,9000 indicates the level of closeness of Balen District to other members in the same cluster, which is relatively small, indicating that the socio-economic characteristics of Balen are quite homogeneous with other districts in Cluster C1. Meanwhile, the b(i) value of 1743,8833 represents the average closest distance of Balen to other clusters, which is much larger than a(i). The significant difference between a(i) and b(i) indicates a clear separation of the clusters.

This is reinforced by the Silhouette Score s(i) value of 0,8871, which is close to the maximum value of 1. This value indicates that the results of grouping Balen District into Cluster C1 are classified as very good, with a high level of internal cohesion and clear separation from other clusters. Thus, Balen District can be categorized as a district that has consistent and representative characteristics of PKH recipients for its group in the K-Means Manhattan scenario without normalization.

A similar pattern was also found in other sub-districts, where each sub-district was grouped based on its proximity to the center of the formed cluster. The values C1, C2, and C3 represent the distance of each sub-district to the center of Cluster 1, Cluster 2, and Cluster 3, respectively, reflecting the degree of similarity in the socio-economic characteristics of PKH recipients in each cluster. Shortest Distance indicates the minimum distance of a sub-district to one of the clusters, which is the basis for determining the final cluster or group in which the sub-district is classified. Furthermore, the value a(i) describes the average distance of a sub-district to other members of the same cluster as an indicator of internal cohesion, while the value b(i) represents the average distance of the nearest cluster to other clusters as an indicator of separation between clusters. The combination of these two values produces the Silhouette Score s(i), which measures the accuracy of sub-district placement within its cluster. Positive s(i) values close to 1 in most sub-districts indicate that the resulting grouping is able to clearly and consistently differentiate socio-economic characteristics between sub-districts in the K-Means scenario with Manhattan Distance without normalization.

TABLE V
K-MEDOIDS WITH EUCLIDEAN DISTANCE NOMALISASI MINMAX

Subdistrict	C1	C2	C3	Shortest Distance	Cluster	a(i)	b(i)	s(i)
Balen	0,1411	1,3126	3,4087	0,1411	C1	0,1411	1,3126	0,8924
Baureno	0,1935	1,2134	3,1789	0,1935	C1	0,1935	1,2134	0,8405
Bojonegoro	1,4517	0,0783	0,2591	0,0783	C2	0,0783	0,2591	0,6975
Bubulan	2,7989	0,5420	0,0000	0,0000	C3	0,0000	0,5420	1,0000
Dander	0,0000	0,9129	2,7989	0,0000	C1	0,0000	0,9129	1,0000
...	...	...	...	...	...	...	...	...
Trucuk	1,3406	0,1080	0,3836	1080	C2	0,1080	0,3836	0,7184

Based on the clustering results using the K-Medoids algorithm with Euclidean Distance on data that has been normalized using Min–Max Scaler, Balen District shows the smallest distance to Cluster 1 (C1) with a Shortest Distance value of 0,1411. This value is lower than the distance of Balen to Cluster 2 (1,3126) and Cluster 3 (3,4087), so Balen District is quantitatively classified into Cluster C1. This indicates that the socio-economic characteristics of PKH recipients in Balen District have a high level of similarity with other districts included in the cluster.

The a(i) value of 0,1411 reflects the relatively small average distance between Balen District and other members of Cluster C1, indicating strong internal cluster cohesion. Conversely, the b(i) value of 1,3126 reflects the closest average distance between Balen District and other clusters, which is much larger than a(i). The striking difference between the two values indicates that the separation between Cluster C1 and the other clusters is clear in this scenario.

This finding is supported by the Silhouette Score s(i) value of 0,8924, which is close to the maximum value of 1. This value indicates that Balen District is placed very precisely in Cluster C1, with optimal levels of cluster cohesion and separation. Thus, the application of the K-Medoids algorithm with Euclidean Distance and Min–Max normalization is able to represent the characteristics of PKH recipients in Balen District in a stable and consistent manner compared to other districts.

A similar clustering pattern was also found in other sub-districts, where each sub-district was grouped into Cluster 1 (C1), Cluster 2 (C2), or Cluster 3 (C3) based on its proximity to the medoid of each cluster. These three clusters represent groups of sub-districts with relatively homogeneous socio-economic characteristics of PKH recipients within each cluster. Shortest Distance indicates the minimum distance of a sub-district to the nearest cluster medoid, which is the basis for determining cluster membership. Furthermore, the value a(i) represents the average distance of a sub-district to other members of the same cluster as an indicator of internal cohesion, while the value b(i) indicates the average distance of the nearest cluster to other clusters as an indicator of separation between clusters. The combination of these two values produces the Silhouette Score s(i), which is used to assess the accuracy of sub-district placement within its cluster. The s(i) value which is generally positive and close to 1 indicates that the clustering results obtained through the K-Medoids algorithm with Euclidean Distance and Min–Max normalization are able to differentiate the characteristics of PKH recipients between sub-districts clearly, stably, and representatively.

To ensure that the clustering results are able to represent the concentration of socially vulnerable groups and social inequality between sub-districts, a clustering quality evaluation was conducted using the Silhouette Score. This evaluation is used to assess the extent to which the formed clusters have a high level of internal similarity (cohesion) and clear differences between clusters (separation), thus objectively depicting variations in socio-economic conditions between sub-districts. The Silhouette Score value is calculated for each combination of k values, clustering algorithms, distance functions, and the application of data normalization. The results of the comparison of Silhouette Score values for all test scenarios are presented in Table VI, which serves as the basis for determining the most optimal clustering configuration for identifying patterns of concentration of socially vulnerable groups and social inequality between sub-districts in Bojonegoro Regency.

TABLE VI
EVALUATION USING SILHOUETTE SCORE

Value K	Algorithm			
	K-Means with Euclidean Distance	***K-Means with Manhattan Distance***	***K-Medoids with Euclidean Distance***	***K-Medoids with Manhattan Distance***
K=2	0,7430	0,7485	0,7688	0,2560
K=2 Normalization	0,6144	0,7281	0,6144	0,7089
K=3	0,6698	0,6679	0,6604	0,6809
K=3 Normalization	0,5968	0,6453	**0,8115**	0,6213

Based on the evaluation results using the Silhouette Score presented in Table VI, the differences in the evaluation results for each combination of algorithms and distance metrics indicate that the choice of distance calculation method has a significant influence on the quality of the resulting clusters. Euclidean distance calculates the distance based on a straight line between points using the squared difference approach for each attribute, making it more sensitive to large value differences in an attribute. Meanwhile, Manhattan distance calculates the distance based on the sum of the absolute differences between attributes, making it more stable against value variations and less affected by extreme differences in one dimension.

Based on the test results, it is seen that in some conditions the evaluation value is higher when using certain metrics, indicating that the data structure is more suited to the characteristics of the distance calculation. For example, at K = 3 with normalization, the K-Medoids method with Euclidean Distance produces the highest value, indicating that once the data is on a uniform scale, the geometric distance differences between objects become more optimal in forming clear clusters. Conversely, in conditions without normalization, some combinations with Manhattan Distance show more stable results because this method is more tolerant of scale differences between attributes.

These differences in results confirm that clustering quality is influenced not only by the algorithm used, but also by the type of distance metric and the data normalization process. Therefore, the combination of algorithms, distance metrics, and preprocessing techniques must be tailored to the characteristics of the dataset to produce more optimal and representative clustering.

In the k = 2 scenario without normalization, the K-Medoids algorithm with Euclidean Distance produces the highest Silhouette Score value of 0,7688, followed by K-Means with Manhattan Distance of 0,7485. This indicates that in the division of two clusters, the K-Medoids method tends to be more stable in distinguishing groups of sub-districts with different levels of social vulnerability. However, in the same scenario, K-Medoids with Manhattan Distance produces a low Silhouette Score value of 0,2560, which indicates that the method is less able to optimally separate clusters in this configuration.

When Min–Max normalization is applied at k = 2, the Silhouette Score values experience varying changes. Some methods experience a decrease in value, such as K-Means with Euclidean Distance, while other methods such as K-Means with Manhattan Distance and K-Medoids with Manhattan Distance show relatively better performance. This indicates that the effect of normalization on clustering quality is highly dependent on the combination of algorithms and distance functions used.

In the k = 3 scenario without normalization, the Silhouette Score values for the four methods are relatively balanced, with a range of values between 0,6604 and 0,6809. This condition indicates that the division of sub-districts into three clusters is starting to be able to describe variations in socio-economic conditions in more detail, although the differences in quality between the methods are not yet very significant.

The best results were obtained in the k = 3 scenario with Min–Max normalization, where K-Medoids with Euclidean Distance produced the highest Silhouette Score of 0,8115. This value indicates that this configuration has the most optimal ability to form compact and clearly separated clusters. With three clusters formed, the sub-district grouping can represent the concentration of socially vulnerable groups and social inequality between sub-districts in more detail and conceptually, compared to the division into two clusters.

Overall, the Silhouette Score evaluation results indicate that the K-Medoids method with Euclidean Distance and Min–Max normalization at k = 3 is the most optimal configuration in this study. This configuration is considered most capable of producing a stable, representative cluster structure with a clear level of separation between groups. Therefore, this model is suitable for use as a basis for policy analysis in determining the priority of Family Hope Program (PKH) recipients in Bojonegoro Regency.

Furthermore, the results of this study are not only descriptive but can also be implemented in the form of a Decision Support System (DSS) in local governments. In the implementation scenario, social indicator data such as the number of early childhood, school-age children (elementary, junior high, and senior high), people with severe disabilities, the elderly, and pregnant women are inputted periodically by relevant agencies into the system. The updated data is then automatically processed through the Min–Max normalization stage and clustering using the K-Medoids algorithm to classify each sub-district into three vulnerability categories: high (high), medium, and low (low). The system then displays the clustering results in the form of a list of sub-district categories, a recapitulation of the number of members in each cluster, and graphic visualizations or thematic maps that facilitate interpretation by policymakers. This information can be used by local governments as an objective basis for determining priorities for social assistance budget distribution, developing more targeted intervention programs, and periodically monitoring changes in social conditions. Furthermore, the clustering results can also be used as an evaluation tool to assess the effectiveness of implemented policies and identify areas experiencing increased or decreased vulnerability. Thus, implementing this model in the DSS enables a more transparent, measurable, and data-driven decision-making process. Integrating data-driven analysis into the regional planning system is expected to improve program targeting accuracy, optimize resource utilization, and support efforts to reduce social inequality between sub-districts in Bojonegoro Regency in a sustainable manner.

D. Best Clustering Model

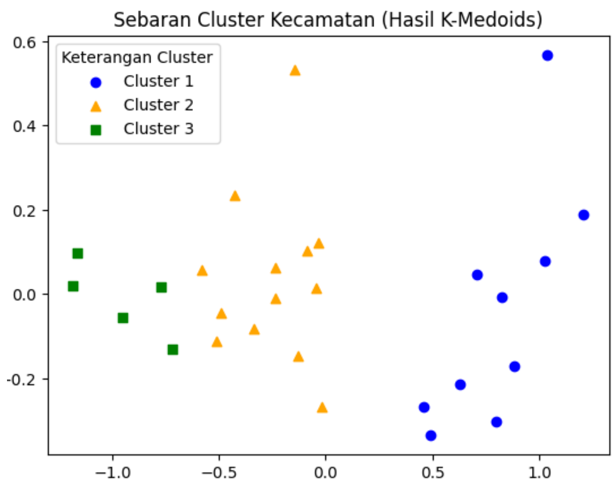

Figure 3 cluster distribution

Based on the Silhouette Score evaluation, the best scenario in this study was the K-Medoids method with Euclidean Distance at k = 3 using Min–Max normalization. This scenario produced the most optimal clustering quality and was therefore used as the basis for further analysis.

The clustering results divided the sub-districts in Bojonegoro Regency into three clusters with distinct characteristics. Visualization of the cluster distribution in two-dimensional space using Principal Component Analysis (PCA) results shows that each cluster forms a relatively separate group, indicating that the clustering process worked well and was able to differentiate characteristics between sub-districts.

Cluster 1 (C1) represents the high PKH recipient category (a large cluster). In the PCA visualization, this cluster appears to be clustered on one side of the PCA space and has a relatively large number of members. The sub-districts included in Cluster 1 are Balen, Baureno, Dander, Kanor, Kedungadem, Kepohbaru, Ngasem, Ngraho, Sumberejo, and Tambakrejo. The sub-districts in this cluster have relatively similar characteristics and demonstrate a high intensity of PKH recipients compared to other clusters.

Cluster 2 (C2) represents the moderate category of PKH recipients (the moderate cluster). Visually, this cluster is located in the middle of the PCA space and serves as a bridge between the high- and low-intensity clusters. The sub-districts included in Cluster 2 include Bojonegoro, Gayam, Gondang, Kalitidu, Kapas, Malo, Margomulyo, Padangan, Purwosari, Sekar, Sugihwaras, Sukosewu, and Trucuk. This cluster reflects a medium-level socio-economic condition.

Meanwhile, Cluster 3 (C3) is the cluster with the fewest sub-districts and represents the low PKH recipient category (a small cluster). In the PCA visualization, this cluster is clearly separated from the other two clusters, indicating significant differences in characteristics. The sub-districts included in Cluster 3 are Bubulan, Kasiman, Kedewan, Ngambon, and Temayang, which have the lowest PKH recipient rates based on their proximity to the medoid.

Thus, the results of this clustering provide an overview of the distribution of sub-districts in Bojonegoro Regency based on the level of PKH recipients, which can be used as a supporting basis in determining the priority of social assistance distribution so that it is more targeted and equitable between regions.

Based on the results of these clustering, local governments are expected to implement more targeted policies for each cluster. Districts in Cluster 1 should be prioritized for aid distribution and community empowerment programs. Districts in Cluster 2 require monitoring and preventive interventions to prevent an increase in the number of vulnerable communities. Meanwhile, districts in Cluster 3 still require attention through economic strengthening and welfare improvement programs to maintain their relatively good conditions. In addition to being a result of academic analysis, the K-Medoids clustering model can be implemented in the form of a local government Decision

Support System (DSS). In this system, social indicator data, such as the number of PKH recipients by category (early childhood, elementary school, junior high school, high school, elderly, disabled, pregnant women), is input periodically by the relevant agencies.

IV. CONCLUSION

This study successfully identified the concentration of socially vulnerable groups and social inequality between sub-districts in Bojonegoro Regency through a data-based clustering approach on recipients of the Family Hope Program (PKH). The clustering results indicate that the distribution of socially vulnerable groups is uneven and tends to be concentrated in certain sub-districts.

Based on the clustering quality evaluation, the K-Medoids algorithm with Euclidean Distance on data normalized using Min–Max and the number of clusters $k = 3$ produced the best model. This model was able to group sub-districts into three clusters representing sub-districts with a large, medium, and small number of socially vulnerable groups. Differences in the number of sub-districts and the characteristics of PKH recipients in each cluster indicate social inequality between sub-districts, both in terms of intensity and distribution of vulnerable groups.

Thus, the results of this study provide a clear picture of sub-districts with a high concentration of socially vulnerable groups as areas requiring greater attention, as well as sub-districts with moderate and low levels of social inequality as benchmarks for assessing regional social inequality. These findings can be used as a supporting basis for determining Family Hope Program (PKH) distribution priorities to ensure more targeted and equitable distribution across sub-districts in Bojonegoro Regency. For further research, comparisons with other clustering methods, such as hierarchical clustering, can be conducted to evaluate the consistency and stability of the clustering results. This approach can provide additional perspectives in understanding the data structure and strengthen the validity of the research findings.

REFERENCES

[1] A. Sofianto, "Implementasi Program Keluarga Harapan (PKH) di Provinsi Jawa Tengah," *Sosio Konsepsia J. Penelit. dan Pengemb. Kesejaht. Sos.*, vol. 10, no. 1 SE-Articles, pp. 14–31, Dec. 2020, doi: 10.33007/ska.v10i1.2091.

[2] N. Hidayatingsih and A. R. Sofa, "Implementasi pendidikan karakter Islami dalam Program Keluarga Harapan (PKH) untuk masyarakat pedesaan: Studi kasus di Desa Dawuhan," *Karakter J. Ris. Ilmu Pendidik. Islam*, vol. 2, no. 2, pp. 11–25, 2025.

[3] M. K. Saidi, Z. Abdussamad, and R. Tantu, "Dampak Sosial Masyarakat pada Program PKH di Desa Dulupi Kecamatan Dulupi Kabupaten Boalemo," *Res. Rev. J. Ilm. Multidisiplin*, vol. 4, no. 1, pp. 585–591, 2025.

[4] Y. Long *et al.*, "Spatially informed clustering, integration, and deconvolution of spatial transcriptomics with GraphST," *Nat. Commun.*, vol. 14, 2023, [Online]. Available: https://api.semanticscholar.org/CorpusID:255478092

[5] J. Macqueen, "Some Methods For Classification And Analysis Of Multivariate Observations," vol. 233, no. 233, pp. 281–297.

[6] Z. Huang, H. Zheng, C. Li, and C. Che, "Application of Machine Learning-Based K-means Clustering for Financial Fraud Detection," *Acad. J. Sci. Technol.*, 2024, [Online]. Available: https://api.semanticscholar.org/CorpusID:269451472

[7] A. Y. Kartini, "Algoritma K-Means Dan K- Medoids Untuk Pengelompokan Kecamatan Penerima Bantuan Sosial di Kabupaten Bojonegoro," 2023.

[8] "Comparison of K-Means and K-Medoids Clustering for Grouping The Sub-Districts In Bojonegoro Regency Based On Educational Supporting Factors ," *J Stat. J. Ilm. Teor. dan Apl. Stat.*, vol. 16, no. 2, pp. 514–523, doi: 10.36456/jstat.vol16.no2.a8079.

[9] A. S. C. Martins, L. R. de Araújo, and D. R. R. Penido, "K-Medoids clustering applications for high-dimensionality multiphase probabilistic power flow," *Int. J. Electr. Power \& Energy Syst.*, 2024, [Online]. Available: https://api.semanticscholar.org/CorpusID:267708052

[10] D. E. Waluyo, H. W. Kinasih, C. Paramita, D. Pergiwati, R. Nohan, and F. A. Rafrastara, "Komparasi dan Implementasi Algoritma Regresi Machine Learning untuk Prediksi Indeks Harga Saham Gabungan," *J. Inform. J. Pengemb. IT*, 2024, [Online]. Available: https://api.semanticscholar.org/CorpusID:268991837

[11] R. Afifa, M. I. Mazdadi, T. H. Saragih, F. Indriani, and M. Muliadi, "Implementasi Principal Component Analysis (PCA) dan Gap Statistic untuk Clustering Kanker Payudara pada Algoritma K-Means," *Sist. J. Sist. Inf.*, vol. 13, no. 5, pp. 1852–1864, 2024.

[12] E. Setiawati, U. D. Fernanda, S. Agesti, M. Iqbal, and M. O. A. Herjho, "Implementation of K-Means, K-Medoid and DBSCAN Algorithms In Obesity Data Clustering," *IJATIS Indones. J. Appl. Technol. Innov. Sci.*, 2024, [Online]. Available: https://api.semanticscholar.org/CorpusID:270060204

[13] J. Heidari, N. Daneshpour, and A. Zangeneh, "A novel K-means and K-medoids algorithms for clustering non-spherical-shape clusters non-sensitive to outliers," *Pattern Recognit.*, vol. 155, p. 110639, 2024, [Online]. Available: https://api.semanticscholar.org/CorpusID:270171548

[14] S. Fahad, F. Su, S. U. Khan, M. R. Naeem, and K. Wei, "Implementing a novel deep learning technique for rainfall forecasting via climatic variables: An approach via hierarchical clustering analysis," *Sci. Total Environ.*, vol. 854, p. 158760, 2023, doi: https://doi.org/10.1016/j.scitotenv.2022.158760.

[15] A. Gupta, H. Sharma, and A. Akhtar, "A comparative analysis of k-means and hierarchical clustering," *EPRA Int. J. Multidiscip. Res.*, vol. 7, no. 8, 2021.

[16] K. P. Sinaga and M.-S. Yang, "Unsupervised K-Means Clustering Algorithm," *IEEE Access*, vol. 8, pp. 80716–80727, 2020, [Online]. Available: https://api.semanticscholar.org/CorpusID:218597833

[17] W. Yang, "Analysis and Application of Big Data Feature Extraction Based on Improved k-means Algorithm," *Scalable Comput. Pr. Exp.*, vol. 25, pp. 137–145, 2024, [Online]. Available: https://api.semanticscholar.org/CorpusID:266781771

[18] M. Mursalim, P. Purwanto, and M. A. Soeleman, "Penentuan Centroid Awal Pada Algoritma K-Means Dengan Dynamic Artificial Chromosomes Genetic Algorithm Untuk Tuberculosis Dataset," *Techno. Com*, vol. 20, no. 1, pp. 97–108, 2021.

[19] D. Anggraini and M. S. Hasibuan, "Initial centroid pada algoritma K-means dan K-medoids," *J. Multimed. dan Teknol. Inf.*, vol. 7, no. 03, pp. 487–500, 2025.

[20] A. Meiriza and E. Ali, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Program BPJS Ketenagakerjaan," *Indones. J. Comput. Sci.*, vol. 12, no. 2, 2023.

[21] R. H. Maharrani, P. D. Abda'u, and M. N. Faiz, "Clustering method for criminal crime acts using K-means and principal component analysis," *Indones. J. Electr. Eng. Comput. Sci.*, 2024, [Online]. Available: https://api.semanticscholar.org/CorpusID:268217000

[22] H. Wahyono, H. Setiaji, T. Hartati, and N. Wiliani, "K-Means Clustering for Identifying Traffic Accident Hotspots in Depok City," *J. Appl. Res. Comput. Sci. Inf. Syst.*, 2024, [Online]. Available: https://api.semanticscholar.org/CorpusID:278552068

[23] Y. Candra, R. Goejantoro, and A. T. R. Dani, "Pengelompokan

Provinsi Berdasarkan Indikator Ekonomi, Pendidikan, Kesehatan, dan Kriminalitas di Indonesia Menggunakan Algoritma Centroid Linkage," *Euler J. Ilm. Mat. Sains dan Teknol.*, vol. 12, no. 1, pp. 9–15, 2024.

[24] R. Khairuna, Nurdin, and Ar Razi, "Analisis Perbandingan Algoritma K-Means Dan K-Medoids Untuk Klusterisasi Teks Ulasan Pada Aplikasi Cookpad: Comparative Analysis Of K-Means And K-Medoids Algorithms For Clusterization Of Review Texts On The Cookpad Application," *Rabit J. Teknol. dan Sist. Inf. Univrab*, vol. 10, no. 2 SE-Articles, pp. 445–458, Jul. 2025, doi: 10.36341/rabit.v10i2.6131.

[25] S. Amaliyah and S. Rianti, "Penerapan Data Mining Untuk Menentukan Kelompok Prioritas Penerima Bantuan PKH Menggunakan Metode Clustering K-Means Pada Desa Kuala Dendang," *J. Inform. Dan Rekayasa Komput.*, vol. 3, no. 1, pp. 453–458, 2023.

[26] B. Deepa and K. Ramesh, "Epileptic Seizure Detection Using Deep Learning Through Min Max Scaler Normalization," *Int. J. Health Sci. (Qassim).*, no. I, pp. 10981–10996, 2022, doi: 10.53730/ijhs.v6nS1.7801.

[27] M. Al Ghifari and W. T. Harsanti Putri, "Clustering Courses Based On Student Grades Using K-Means Algorithm With Elbow Method For Centroid Determination," *Inf. J. Ilm. Bid. Teknol. Inf. dan Komun.*, vol. 8, no. 1 SE-Articles, pp. 42–46, Jan. 2023, doi: 10.25139/inform.v8i1.4519.

[28] D. Anggreani, N. Nurmisba, D. Setiawan, and L. Lukman, "Optimization of K-Means Clustering Method by Using Elbow Method in Predicting Blood Requirement of Pelamonia Hospital Makassar," *Internet Things Artif. Intell. J.*, 2024, [Online]. Available: https://api.semanticscholar.org/CorpusID:272982268

[29] D. T. Yuliana, M. I. A. Fathoni, and N. Kurniawati, "Penentuan Penerima Kartu Indonesia Pintar KIP Kuliah Dengan Menggunakan Metode K-Means Clustering," *J. Focus Action Res. Math. (Factor M)*, vol. 5, no. 1 SE-Articles, pp. 127–141, Dec. 2022, doi: 10.30762/f_m.v5i1.570.

[30] N. Puspitasari, G. Lempas, H. Hamdani, H. Haviuddin, and A. Septiarini, "Perbandingan Algoritma K-Means dan Algoritma K-Medoids Pada Kasus Covid-19 di Indonesia," *Build. Informatics, Technol. Sci.*, vol. 4, no. 4, p. 2015−2027, 2023, doi: 10.47065/bits.v4i4.2994.

[31] N. Rohman and A. Wibowo, "Perbandingan Metode K-Medoids dan Metode K-Means Dalam Analisis Segmentasi Pelanggan Mall," *SINTECH (Science Inf. Technol. J.*, vol. 7, no. 1, pp. 49–58, 2024, doi: 10.31598/sintechjournal.v7i1.1507.

[32] I. M. Nur, U. G. Mada, U. M. Semarang, A. Info, and S. Vulnerability, "Analysis of Social Vulnerability in Java Island using K-Medoids Algorithm with Variation of Distance Measurements (Euclidean , Manhattan , Minkowski)," vol. 7, no. 2, pp. 467–475, 2024.

[33] N. A. Siagian, A. Rikki, and P. B. N. Simangunsong, "Clustering Menggunakan Metode K-Medoids dengan pendekatan Manhattan Distance," *KAKIFIKOM (Kumpulan Artik. Karya Ilm. Fak. Ilmu Komputer)*, vol. 6, no. 2, pp. 169–175, 2024, doi: 10.54367/kakifikom.v6i2.4608.

[34] F. R. Senduk, I. Indwiarti, and F. Nhita, "Clustering of Earthquake Prone Areas in Indonesia Using K-Medoids Algorithm," *Indones. J. Comput.*, vol. 4, no. 3, pp. 65–76, 2020, doi: 10.34818/INDOJC.2019.4.3.359.