

DOC-20250613-WA0006.docx

by akunpremiumcumlaude@gmail.com 1

Submission date: 13-Jun-2025 01:13PM (UTC+0900)

Submission ID: 2695390209

File name: DOC-20250613-WA0006.docx (1.05M)

Word count: 5491

Character count: 28171



The Title Must Be Brief, Concise, and Informative (Maximum 15 Words)

First Author¹, Second Author^{2,*}, Third Author³, and Fourth Author⁴

^{1,3}Affiliation One and Three

²Affiliation Two

⁴Affiliation Four

*Corresponding author's email: xxxxx@xxxxx.com

Abstract

The stunting rate in West Kalimantan has reached 27%, largely due to the government's failure to prioritize regions for essential services and education, especially for adolescents and pregnant women. This study aims to explain the role of modified K-Means and CHI methods in forming optimal clusters and interpreting their conditions. Eight research variables were used, sourced from BPS and SIGA in 2023: number of adolescents receiving counseling, number of informed consents, complication cases, aslokon expenditure, aslokon stock, population growth rate, population density, and life expectancy. These variables were validated using PCA with the Kaiser and PVE approaches. Clustering was conducted by analyzing each variable's data, object characteristics using Euclidean distance, determining centroid values, and iterating until results stabilized. The number of clusters was evaluated from one to seven to find the optimal amount using CHI. Results identified five clusters: cluster 1 (fairly poor, three objects), cluster 2 (very poor, four objects), cluster 3 (good, three objects), cluster 4 (very good, three objects), and cluster 5 (good, one object).

Keywords: Centroid, Modified K-Mean Cluster, and Euclidean Distance.

Copyright © 2025 by Authors, Published by CAUCHY Group. This is an open access article under the CC BY-SA License (<https://creativecommons.org/licenses/by-sa/4.0>)

1 Introduction

Non-hierarchical clustering is a part of cluster analysis. The non-hierarchical method aims to group N research objects into K clusters, where the value of K is predetermined according to the needs of the study and $K < N$ [1]. The number of clusters (K) is determined in advance in order to analyze the temporary cluster centers, which will be updated through an iterative process until the criteria are met-specifically, when there is no movement of objects from one cluster to another and no change in the mean value of the centroids or cluster centers [2].

In non-hierarchical clustering processes, the Modified K-Means Clustering method is generally used for cluster formation. Clusters are formed starting from the minimum number of clusters (K_{min}), which is two clusters, up to the maximum number of clusters (K_{max}), which is seven clusters. This modified method addresses inconsistent clustering results from standard K-Means by improving cluster quality evaluation metrics, reducing the number of iterations required, and optimizing processing time efficiency [3]. Certainly! Here is your paragraph translated into English with attention to grammar and clarity:

After forming clusters for (K_{min}) to (K_{max}), an analysis is conducted to determine the optimal number of clusters within the interval $[K_{min}, K_{max}]$ where $K \in R$. Determining the optimal number of clusters is important, as it helps ensure that the grouping is done with the appropriate number of clusters, allowing for treatments based on the similarity levels among cluster members. The Calinski-Harabasz Index (CHI) method is chosen because it uses a hybrid criterion, evaluating both the compactness of cluster centroids to ensure sufficient separation between clusters and the similarity among members within each

cluster. Additionally, the CHI method incorporates the values of K and N in the evaluation process [4].

Research on the Modified K-Means Clustering method has been conducted by several researchers, including Choldun and Rachmatullah (2022), the aim of their research was to compare the K-Means and Modified K-Means methods in order to assess the impact of improved distance calculation using correlation-based transformation versus Euclidean distance in clustering sepal length, sepal width, petal length, and petal width. The results showed that the Modified K-Means Clustering method achieved a higher accuracy rate compared to the standard K-Means Clustering method, with Modified K-Means achieving 89.33% accuracy and K-Means Clustering achieving 82.67%.

Research on the CHI method was conducted by Suraya and Wijayanto (2022), the objective of this research was to determine the best method for forming clusters based on the Special Stunting Management Index (IKPS) for 34 provinces in 2018, with a minimum number of clusters $K_{min} = 2$ and a maximum number of clusters $K_{max} = 5$. The methods used included Average Linkage, K-Means, K-Medoids, and Fuzzy C-Means, and the evaluation was carried out using the Connectivity Coefficient (CC), Dunn Index (DI), Silhouette Coefficient (SC), Davies Bouldin Index (DBI), Silhouette Index (SI), Calinski-Harabasz Index (CHI), and Xie and Beni Index (XBI). The results indicated that the two best methods were Average Linkage, with the optimal number of clusters being four (as supported by the DI, DBI, and XBI evaluation methods), and K-Means, with the optimal number of clusters being two (as supported by the SC and CHI evaluation methods).

The study aims to explain the role of the modified K-Mean and CHI methods in forming the best cluster and interpreting the condition of each cluster. In this study, the modification of the k-means method lies in the selection of the initial centroids based on the characteristics among the research objects. This modification has also been carried out by Cahyanto et al. The difference between this study and previous research is that this study aims to improve the quality of cluster formation using the modified k-means clustering method and applies the Euclidean method to determine the level of similarity between objects. Furthermore, the results of each cluster formation are analyzed using the CHI method. introduction should begin by presenting the general context of the research area, explaining

METHODS

Data

The data used in this study consists of stunting management indicators, comprising eight indicators: the number of informed consent forms given during services to new and repeat family planning participants (X_1), the number of severe complication and failure cases (X_2), the availability of contraceptive devices and drugs (X_3), the distribution of alokon (X_4), the number of adolescents attending counseling at youth and student information and counseling centers (X_5), life expectancy (X_6), population growth rate (X_7), and population density (X_8) for each city/regency in West Kalimantan Province in 2023, which were obtained from sistem informasi keluarga (SIGA), badan pusat statistik (BPS), dan Laporan Semester I Penyelenggaraan Percepatan Penurunan Stunting Kalimantan Barat Tahun 2023. The variables used are sufficient to explain the issues in stunting management, referring to Appendix 1 and employing the Principal Component Analysis (PCA) method with the Proportion of Variance Explained (PVE) approach and the Kaiser criterion. Using these two approaches, it was found that the eight indicator variables for stunting management can be further processed because, according to the Kaiser criterion, three Principal Components (PC1, PC2, and PC3) with eigenvalues ≥ 1 were formed. Meanwhile, the PVE approach showed that the three main components (PC1, PC2, and PC3) explain 76.770% of the total data variance, indicating that the data sufficiently represents the dataset.

Table 1 Data on Stunting Handling Indicator Variables

²⁷ Kabupaten	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
Sambas	15	6535	1	41687	64899	1,20	109	69,805
Mempawah	24	9273	2	19132	40234	1,32	162	71,785
Sanggau	437	16556	3	26534	42369	1,39	40	71,815
Ketapang	583	28922	15	39214	99373	1,35	20	71,505
Sintang	0	16974	8	29201	71327	1,40	20	72,460
...								
Singkawang	930	9741	4	29029	45803	1,67	447	72,865

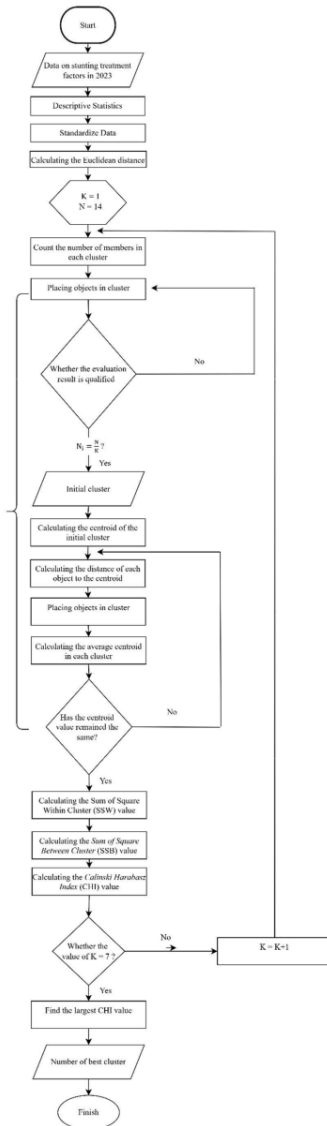


Figure 1 Research Model

The research method contains 5 methods that are used in the research and can be seen in the following explanations.

First Author

Standardization of data

Data standardization using the Z-Score method, this method is commonly used because it standardizes the dataset by measuring data deviation based on a review of the average value to see the concentration of the dataset and using the standard deviation value to see the data distribution [5]. The following is the formula for finding the value of data standardization [5]:

$$Z_{i,j} = \frac{x_{i,j} - (\bar{x}_j)}{s_j} \quad (1)$$

Where:

$Z_{i,j}$	:	Standardization of data for object to on- i variable to- j , where $i = 1,2, \dots, 14$ and $j = 1,2, \dots, 8$
$x_{i,j}$	:	object to on- i variable to- j , where $i = 1,2, \dots, 14$ and $j = 1,2, \dots, 8$
$\bar{x}_j$	:	the average value of variable to- j , where $j = 1,2, \dots, 8$
s_j	:	standard deviation value of variable to- j , where $j = 1,2, \dots, 8$

Distance calculations

The calculation uses the Euclidean distance method because it is able to improve the lack of random initialization at the center of the data which serves to increase accuracy in the cluster formation process [6]. For the calculation of Euclidean distance, it is divided into two calculations, namely for calculating the distance between objects and objects and calculating the distance between centroids and objects. The following is the formula for finding the distance value using the Euclidean method [7].

$$d(x_i, x_k) = \sqrt{\sum_{j=1}^n (x_{i,j} - x_{k,j})^2} \quad (2)$$

$$d(c_l, x_i) = \sqrt{\sum_{j=1}^n (c_{l,j} - x_{i,j})^2} \quad (3)$$

Where:

- $d(x_i, x_k)$: Distance between object to- i and object to- k
- $d(c_l, x_i)$: Distance between centroid to- l and object to- i
- n : Total number of variables
- j : data from variable to- j , where $j = 1$
- x_i : object to- i
- x_k : object to- k
- $c_{l,j}$: centroid to- l on variable to- j

Equation (2) is used to find the Euclidean distance between two objects whose calculation results are used for initial cluster formation to find the initial centroid value, while Equation (3) is used to find the distance value between the centroid and the object to- i the variable to- j the iteration process.

Modified K-Means Cluster

The Modified K-Means Cluster method is a modification made to improve the shortcomings of the K-Means Cluster method in the process of finding the initial cluster center value (centroid) is done randomly which causes inconsistent cluster results. Modifications are made to the initial cluster formation by adding a condition that the number of members of each cluster must meet $\frac{\text{number of objects } (N)}{\text{number of clusters } (K)}$, this aims to improve the quality of cluster formation and time efficiency at the iteration stage [8]. Cluster formation is carried out

The Title Must Be Brief, Concise, and Informative, and...

in stages by first completing the formation at the minimum number of clusters (K_{min}), namely one after completing the formation at the number of clusters of two and so on, then the $K = K + 1$ process can be carried out up to the maximum number of clusters (K_{max}), namely up to seven clusters. The stages with the Modified K-Means Clustering method are as follows:

1. Determine the number of clusters to be formed (K).

Determining the number of clusters starts for $K_{min} = 1$ aims to see the effect in the clustering process, then sets for $K_{max} = \frac{N}{2} = \frac{14}{2} = 7$. The value of $K_{max} = 7$ because it meets the requirements of non-hierarchical clusters, namely $K < N$. In addition, the value of $K > \frac{N}{2}$ or the value of K in the range of 8 to 13 will cause a bias that does not represent the actual conditions or namely overfitting because the number of clusters is too much than necessary, causing the model to be too complex [9].

2. Count the number of members in each cluster

The process of calculating the number of members in each cluster is done $\frac{N}{K}$. This process aims to make each cluster have the same number of members, so that the iteration process is more efficient and to optimize the role of each object in the process of forming the initial centroid value.

3. Places objects in the cluster.

Next, find the Euclidean distance value used to analyze the level of object similarity, the smaller the Euclidean value of the two objects the more similar they are. Once obtained, the two objects enter the first two members in the first cluster. Compare the object distance value of the remaining 12 objects with the two objects at the beginning then select the smallest object value object until it meets $\frac{N}{K}$. Do the same steps to place objects in the second cluster until the next cluster.

4. Evaluation of initial cluster members.

The results in Step 3 are then evaluated by seeing whether the $\frac{N}{K}$ requirement has been met, and the threshold value is appropriate based on the closest distance. if it has not met the $\frac{N}{K}$ requirement or there is an error in placing the smallest value as a cluster member, then do Step 3 again [10].

5. Obtaining the initial cluster.

The final result in Step 4 is the initial cluster, which is the cluster used to obtain the initial centroid value based on the members of each cluster, which in the next stage is used as the initial capital in the iteration process.

6. Initial centroid formation.

Initial centroid formation is used for the iteration process in cluster formation. The following is the formula for finding the centroid value [3].

$$c_{l,j} = \left(\frac{1}{m_k} \right) \sum_{x \in C_i} Z_{l,j} \quad (4)$$

Where m_k is the number of observations in the cluster to $-k$ and $c_{l,j}$ is the centroid value to $-l$ on variable to $-j$, where $t = 1, 2, \dots, K$ and $j = 1, 2, \dots, 8$.

7. Calculate the distance value of each object with each centroid.

The calculation of the distance between the $c_{l,j}$ value and x_{ij} aims to place the object in the cluster. The

The Title Must Be Brief, Concise, and Informative, and...
calculation uses Equation (3).

8. Place objects into clusters that have the closest centroid distance.

The results in Step 7 have been obtained, namely the value of the Euclidean distance between each object and the centroid, then next look for the minimum value between each object and the centroid. If the minimum value of the first object is located at the first centroid then the first object is placed in the first cluster, as well as if the minimum value of the second object is located at the fourth centroid then the second object is placed in cluster four, do so until the 14th object.

9. Iteration.

The results of placing objects in clusters are used to re-conduct Stage 6, then repeat Stage 6 to Stage 8 to iterate, the iteration stops when the centroid value at iteration $n + 1$ is the same as the centroid value at iteration n , and the object placed into the cluster at iteration $n + 1$ does not change position from iteration n , so the centroid and object placement are declared consistent.

Calinski Harabasz Index (CHI)

Determination of the optimal number of clusters to help grouping with the right amount of use based on the level of similarity of cluster members, the CHI method evaluates the homogeneity and heterogeneity of the clusters formed [4]. CHI is used to evaluate the cluster model that has been formed, this method refers to the ratio of the amount of disperse (sum of squared distances) between clusters and in clusters for all clusters [11]. The stages with the Calinski Harabasz Index method are as follows:

1. Calculating SSW and SSB values.

The Sum of Squared Within cluster (SSW) process is carried out to evaluate the level of similarity between each member in the same cluster, while the Sum of Squared Bithwin cluster (SSB) process is carried out to evaluate the level of similarity between each cluster. The following is the formula for finding the SSW and SSB values:

$$SSW = \sum_{i=1}^N d^2(c_l, Z_{l,i}) \quad (5)$$

$$SSB = \sum_{i=1}^k N_k * d^2(c_l, \bar{x}_l) \quad (6)$$

Where:

- $d^2(c_l, x_j)$: The square of the distance between centroids to $-l$ with object to $-i$
- $d^2(c_l, \bar{x}_j)$: The square of the distance between centroids to $-l$ with average of each variable
- N_k : The number of objects belonging to the cluster $-k$
- N : The number of objects researched

2. Finding the CHI value.

The process of finding the CHI value aims to evaluate the number of clusters formed based on a review of the SSW, SSB and review values involving the number of objects (N) and the number of clusters formed (K). The following is the formula for finding the CHI value [12].

$$CHI = \frac{SSB}{SSW} \times \frac{N - K}{K - 1} \quad (7)$$

Cluster Labeling

Cluster labeling is done to make it easier to know the condition of each cluster that has been formed, so that it helps to make it easier to choose priority areas. The following is the formula for finding values for cluster labeling [13].

$$C_k x_j = (\bar{x}_j) + c_{ij} \times s_j \quad (8)$$

Where $C_k X_j$ is the labeling of the cluster to $-k$ and variable to $-j$.

RESULT AND DISCUSSION

Descriptive Statistic

The data in Table 1 were used because of the processing results using Principle Component Analysis (PCA) with the Kaiser approach and Proportion of Variance Explained (PVE). The PCA method is used to explain the structure of the relationship between variables in the analysis, while the Kaiser approach is used to reduce the variables into several main dimensions with the condition that the number of dimension formation depends on the number of factor values that meet the eigenvalue > 1 , while the PVE approach is used to measure the contribution of information to the main dimensions obtained by meeting the target, namely the $PVE \geq 70\%$. The results of data processing with the PCA method are presented in Table 2.

Table 2. PCA process results.

Faktor	Variabel	Keterangan	Nilai Eigen	Proportion of Variance Explained (PVE)
PC1	X1	Banyak remaja hadir konseling	2.85803	0.3573
PC2	X2	Banyak pemberian informasi consent pada pelayanan peserta KB	2.22732	0.2784
PC3	X3	Banyak kasus komplikasi berat dan kegagalan	1.05563	0.132
PC4	X4	Banyak pengeluaran alat dan obat kontrasepsi	0.93276	0.1166
PC5	X5	Banyak persediaan alat dan obat kontrasepsi	0.39514	0.04744
PC6	X6	Laju pertumbuhan penduduk per tahun	0.26695	0.03337
PC7	X7	Kepadatan penduduk per km persegi	0.22563	0.0282
PC8	X8	Angka harapan hidup	0.05417	0.00677

Based on Table 2, based on the Kaiser approach, the variables were reduced to 3 main dimensions and for the PVE approach, it was found that the variables were sufficiently reduced or represented three main dimensions because the sum of the PVE values on the three factors was 0.7677 or represented 76.77%. Table 3 shows the abbreviations for the names of cities/districts.

Table 3. Abbreviation of Each Object

Districts/Cities	City/District abbreviation	Districts/Cities	City/District abbreviation
7 Sambas	SBS	Landak	LDK
Mempawah	MPW	Sekadai	SKD
Sanggau	SGU	Melawi	MLW
Ketapang	KTP	Kayong Utara	KYU
Sintang	STG	Kubu Raya	KBR
Kapuas Hulu	KPH	Pontianak	PTK
Bengkayang	BKY	Singawang	SKW

The data in Table 1 is then presented in the form of descriptive statistics to make it easier to understand the data for each variable. Descriptive statistics present the mean and standard deviation values that will be

used for data standardization and interpretation of the condition of each cluster at the best number of clusters. Descriptive statistics are shown in Table 4.

Table 4. Descriptive Statistics of Data

Variable	N	Minimum		Maximum		Average	Standard Deviation
		Value	Districts/Cities	Value	Districts/Cities		
X_1	14	0	STG, BKY, SKD, MLW	930	SKW	213,143	281,690
X_2	14	3623	MLW	28922	KTP	12229,785	7848,068
X_3	14	1	LDK, SBS	27	KBR	7,500	6,832
X_4	14	10684	KYU	47519	KBR	26271,929	10601,631
X_5	14	22680	MLW	196707	SKD	66778,286	44240,906
X_6	14	0,870	PTK	1,770	KYU	1,436	0,231
X_7	14	8	KPH	5714	PTK	485,071	1454,304
X_8	14	69,270	KYU	74,575	BKY	72,302	1,460

The data presented in Table 4 represents some variables that are good enough. The variables X_1, X_3 and X_6 recorded good values because the difference between the mean value and the standard deviation is still within the unit value, this means that the value is spread around the average. Other variables must be improved in dealing with stunting conditions in the community. The standard deviation of each variable is still relatively high, meaning that the value is spread away from the average so that the data has outlier data, therefore the data needs to be standardized. Based on Table 4, the data units are quite different so that they are more susceptible to the influence of outliers, so the units need to be equalized with the *Z-Score* method using Equation (1).

Modified K-Means Cluster

Data on the quality of stunting handling that has been standardized, then the calculation of Euclidean distance is carried out based on Equation (2) with $n = 8$ because there are eight indicator variables, it should be noted that for the results of the Euclidean distance, the characteristics between regions are said to be more similar if the distance value is smaller. The next stage for the value of N obtained in Table 4 is 14 and K is the number of clusters to be formed. The process of determining the value of K starts for $K_{\min} = 1$, because to prove homogeneity and heterogeneity in cluster formation, then determine $K_{\max} = \frac{N}{2} = \frac{14}{2} = 7$, because it meets the requirements of non-hierarchical clusters and avoids overfitting due to the number of clusters being too much than necessary [9].

Initial cluster formation based on the shortest distance obtained from the calculation of the Euclidean distance as a threshold value, then placing objects in clusters by fulfilling the members of each cluster as many as $\frac{N}{K}$. The value of K starting from $K = 1$ to $K = 7$ is processed until it is obtained that the initial cluster has no change in the value of the objective function [10]. The results of the initial cluster formation are presented in Figure 2.

The Title Must Be Brief, Concise, and Informative, and . . .

K=1	K=2		K=3			K=4				K=5				
	C1	C2	C1	C2	C3	C1	C2	C3	C4	C1	C2	C3	C4	C5
SBS	MLW	SGU	MLW	MPW	KTP	MLW	MPW	KTP	SKD	MLW	MPW	SGU	KTP	KYU
MPW	BKY	SKW	BKY	STG	KBR	BKY	STG	KBR	KYU	BKY	KPH	SKW	KBR	PTK
SGU	MPW	SBS	LDK	SGU	SKW	KPH	SGU	SKW	PTK	LDK	STG	SBS	SKD	
KTP	STG	KTP	KPH	SBS	SKD	LDK	SBS			(1.22; 1.01)	(1.86; 2.66)	(2.81; 3.92)	(5.40; 5.78)	
STG	KPH	SKD	KYU	PTK										
KPH	LDK	KBR												
BKY	KYU	PTK												
LDK	(1.22; 2.92)	(2.44; 5.13)												
SKD														
MLW														
KYU														
KBR														
PTK														
SKW														
0.000	(1.22; 1.61)	(1.547; 1.66)	2.44	2.81	4.06	5.78								

Figure 2. Initial Cluster Formation Results for Each K

Based on Figure 2, the number of members of each cluster has met $\frac{N}{K}$ then it has met the objective function, for the value located at the bottom position of each cluster is a range of threshold values that indicate the similarity between members in each cluster. This stage is the difference between the Modified K-Means Cluster method and the K-Means Cluster method because the placement of objects in the initial cluster will be used to form the initial centroid so that it will streamline the iteration process.

Based on Figure 2, average value of each variable for each cluster is calculated based on Equation (4) using standardized values to obtain the initial centroid value [14]. The centroid values are presented in Table 5.

Table 5. Initial Centroid Value

		Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8
K = 1	C1	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	C2	-0,515	-0,610	-0,282	-0,780	-0,554	0,287	-0,299	0,207
K = 2	C1	0,515	0,610	0,282	0,780	0,554	-0,287	0,299	-0,207
	C2	-0,515	-0,610	-0,282	-0,780	-0,554	0,287	-0,299	0,207
K = 7	C1	-0,757	-1,011	-0,146	-0,980	-0,792	0,816	-0,307	1,148
	C2	-0,448	-0,474	-0,220	-0,630	-0,435	-0,395	-0,275	0,036
	C3	0,019	0,578	-0,293	0,151	-0,224	-0,179	-0,313	-0,113
	C4	1,215	1,592	1,976	1,612	0,922	0,470	-0,301	-0,615
	C5	1,095	-0,606	-0,732	-0,393	-0,509	0,448	-0,163	0,716
	C6	-0,393	-0,868	-0,659	-0,008	-0,513	0,210	-0,285	-1,894
	C7	-0,732	0,789	0,073	0,248	1,552	-1,369	1,644	0,721

The calculation uses Equation (3) to obtain the distance value of each object with the initial centroid. The calculation results are presented in Table 6.

Table 6. Value of Object-Centroid Distance

	K = 2		K = 7						
	C1	C2	C1	C2	C3	C4	C5	C6	C7
SBS	3,338	2,952	4,333	2,934	2,767	4,691	3,871	2,029	4,123
MPW	1,145	2,711	2,231	0,774	1,626	4,872	2,306	1,917	3,606
SGU	2,094	1,775	3,067	1,846	0,953	3,664	1,776	2,456	3,698
KTP	4,406	2,082	5,210	4,092	2,881	1,404	4,161	4,494	3,884
STG	1,848	1,618	2,658	1,587	0,953	3,443	2,669	2,728	2,799
KPH	1,002	2,301	1,704	0,774	1,632	4,155	1,922	2,683	3,370
BKY	1,474	3,465	0,609	1,903	2,713	5,171	2,257	3,653	4,091
LDK	1,221	3,323	1,328	1,416	2,371	5,435	1,762	3,149	3,943
SKD	3,840	3,038	4,316	3,637	3,526	4,000	4,342	4,475	2,891
MLW	1,253	3,774	0,609	1,951	2,835	5,487	2,247	3,084	4,702
KYU	2,751	4,208	3,403	3,049	3,498	5,493	3,448	2,029	5,727
KBR	5,257	3,480	5,736	5,196	4,371	1,404	5,168	5,262	5,233
PTK	5,549	4,581	6,016	5,162	4,897	6,264	5,672	6,186	2,891
SKW	3,349	3,013	3,722	3,467	3,008	4,093	1,762	3,864	4,997

The minimum value of the initial centroid value is bolded in orange in Table 6 which aims to place objects in clusters, this result is temporary because the iteration process continues. The members in each cluster for each K are presented in Figure 3.

K=1	K=2		K=3			K=4				K=5				
C1	C1	C2	C1	C2	C3	C1	C2	C3	C4	C1	C2	C3	C4	C5
SRL	MPW	SGU	MLW	MPW	KTP	MLW	STG	KTP	SKD	MLW	MPW	SGU	KTP	PTK
MPW	KPH	SKW	BKY	STG	KBR	BKY	SGU	KBR	PTK	BKY	KPH	SKW	KBR	PTK
SGU	BKY	SBS	LDK	SGU	SKW	KPH	SBS	SKD		LDK	STG	SBS	SKD	
KTP	LDK	KTP	KPH	SBS	SKD	LDK				KYU				
STG	MLW	SKD		KYU	PTK	KYU								
KPH	KYU	KBR				MPW								
BKY		PTK												
LDK		STG												
SKD														
MLW														
KYU														
KBR														
PTK														
SKW														

Figure 3. Placement of Objects on Clusters

The next stage is to iterate by calculating the average centroid to form a new centroid. The purpose of forming a new centroid is to see changes in the centroid value. If there is a change, the iteration process is carried out, but if the average centroid value as the new centroid value does not change, this value is the final centroid value with the location of the object in the cluster is final or does not need to be changed. Then based on Figure 3, the average centroid value is calculated using Equation (4). The new centroid value in the first process obtained the same value as the previous centroid value, so the initial centroid value is the final centroid value, so the iteration process stops because it is consistent.

The final centroid value has been obtained, the next step is to find the SSW value by referring to Equation (5). The formation of the number of clusters is better if the SSW value is smaller. Furthermore, calculating the SSB value using Equation (6), the best number of clusters is obtained when the SSB value is getting bigger. The SSB and SSW values are presented in Table 7.

CHI Value

Table 20. Determination of the Best Number of Clusters

	K = 2	K = 3	K = 4	K = 5	K = 6	K = 7
Nilai SSW	84,442	73,714	53,298	39,311	36,671	33,827
Nilai SSB	27,558	38,286	58,881	72,689	75,330	78,173
Nilai CHI	3,916	2,857	3,682	4,160	3,287	2,696

Based on Table 7, the black bolded value is the best value in each calculation and it is obtained that the best number of clusters $K = 5$, meaning that there are 5 categories of regional groups based on the quality of stunting handling, which include very good, good, sufficient, less, and very less handling groups.

Cluster Labeling

The labeling of each cluster is used to sort the cluster positions from best to worst, this aims to determine the cluster priority standards that must be focused on so that the stunting index target in 2024 can be achieved. Based on Table 7, the best number of clusters is obtained when $K = 5$ with a value of 4,160. The members of each cluster at $K = 5$ are presented in Table 8.

Table 8. Final Results of Each Cluster Member

C1	C2	C3	C4	C5
BKY	MPW	SBS	KTP	PTK
LDK	STG	SGU	SKD	
MLW	KPH	SKW	KBR	
	KYU			

To find out the priority areas, each cluster will be labeled, this is important to do because it is used to analyze the characteristics of each cluster based on the problem of handling stunting by referring to indicator variables. Labeling begins by calculating the average value of the final centroid which will be presented in Table 9 using Equation (4), then performing calculations using Equation (8).

Table 9. Final Centroid Value and Status of Each Cluster

Indicator Variable	Final Centroid Value				
	C1	C2	C3	C4	C5
X ₁	-0.623	-0.434	0.879	0.558	-0.707
X ₂	-0.972	-0.339	-0.164	1.039	1.645
X ₃	-0.415	-0.183	-0.707	1.586	-0.659
X ₄	-1.002	-0.614	0.580	1.089	0.454
X ₅	-0.710	-0.438	-0.356	1.594	0.167
X ₆	0.506	0.124	-0.071	0.217	-2.450
X ₇	-0.304	-0.295	-0.197	-0.303	3.595
X ₈	1.114	-0.474	-0.553	-0.300	1.112
Average	-0.301	-0.332	-0.074	0.685	0.395
Status	Bad	Very Bad	Enough	Very Good	Good

Based on Table 9 for the status of the labeling results based on the average value of the final centroid, the higher or positive value, the better the cluster status. The next step is to perform calculations with Equation (8) to obtain the characteristics of each indicator variable in each cluster. The calculation results are presented in Table 10.

Table 10. Cluster Labeling

	interpretation				
	C1	C2	C3	C4	C5
X ₁	37,667	91,000	460,667	370,333	14,000
X ₂	4601,000	9570,250	10944,000	20386,330	25142,000
X ₃	4,667	6,250	2,667	18,333	3,000
X ₄	15653,330	19765,750	32416,670	37817,000	31083,000
X ₅	35388,670	47411,000	51023,670	137280,700	74173,000
X ₆	1,553	1,465	1,420	1,487	0,870
X ₇	42,333	55,500	198,667	44,000	5714,000
X ₈	73,928	71,610	71,495	71,865	73,925
Hasil					

Table 10 is the cluster labeling result of 2 iterations. These results are compared with the initial average value data in Table 3. There are several interpretations for black values that are bolded meaning above average, for not bolded below average, and for indicator variables that are bolded meaning values above average meaning

The Title Must Be Brief, Concise, and Informative, and...

less good. The results in Table 8 to Table 10 are presented visually through Figure 4 to explain the condition of stunting management in each region.

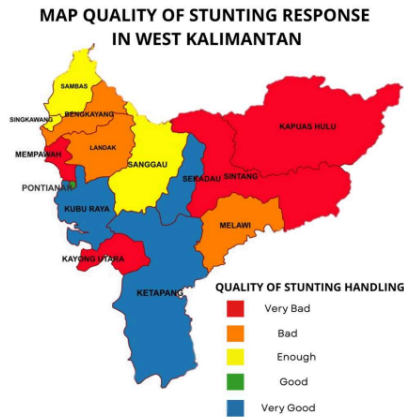


Figure 4. Map Best Cluster

Based on Table 8 and Figure 4, it is further interpreted that cluster 1 is a cluster with poor stunting conditions, but there are some positive notes that have been shown, namely the value of life expectancy (AHH) which is above average, which means that public awareness about health is quite high. Cluster 2 shows the complex problems that occur so that cluster 2 is a cluster with a very poor level of quality in handling stunting, AHH is below average, public awareness to utilize health facilities is low [15]. Cluster 3 is a moderate condition cluster, has problems with the low awareness of health workers to provide informed consent when providing services to family planning participants, the next problem is that although the alocon expenditure is above average, the alocon supply is below average so it is necessary to increase public awareness of the importance of consistent use of alocon, the last problem is that AHH is below average. in cluster 4 and cluster 5 are clusters with very good and good conditions.

Conclusions

The role of the modified method is the formation of clusters by analyzing the data units of each variable, the level of characteristics of each object through Euclidean distance, finding the value of the cluster center (centroid) then iterating until the results are consistent as a result of cluster formation while the role of CHI evaluates each number of clusters from a minimum number of one to a maximum of seven to obtain the best number of clusters. The best clusters from the modified K-Mean cluster process ranging from K_{min} to K_{max} are then reviewed based on the SSW and SSB methods and then the CHI value is calculated. The best cluster formed is at $K = 5$ with a CHI value of 4.160 which is the largest CHI value of the other K. Cluster 1 is a cluster with poor stunting handling quality consisting of three cities/districts namely Bengkayang, Landak, and Melawi. Cluster 2 is a cluster with very poor stunting management quality consisting of four cities/districts namely Mempawah, Sintang, Kapuas Hulu, and North Kayong. Cluster 3 is a cluster with sufficient quality of stunting management consisting of three cities/districts namely Sambas, Sanggau, Singkawang. Cluster 4 is a

cluster with very good stunting management quality consisting of three cities/districts namely Ketapang, Sekadau, and Kubu Raya. Cluster 5 is a cluster with good stunting handling quality, which is only filled by Pontianak.

References

- [1] S. T. M. Mukti dan A. Wibowo, "Pemetaan Wilayah Kesehatan Ibu Dan Anak Di Provinsi Jawa Timur Tahun 2014," *Jurnal Ilmiah Kebidanan*, 2017.
- [2] R. A. H. Imanni, E. Sulistianingsih, dan H. Perdana, "Analisis Cluster Menggunakan Algoritma K-Means Berdasarkan Faktor Penyebab Stunting Pada Provinsi Kalimantan Barat," *Bimaster : Buletin Ilmiah Matematika, Statistika dan Terapannya*, 2023.
- [3] R. Cahyanto, A. R. Chrismanto, dan D. Danny Sebastian, "Pengelompokan Komentar Dataset Sentipol dengan Modified K-Means Clustering," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 6(3), 2020.
- [4] D. R. S. Saputro dan others, "Algoritme Partitioning Around Medoid (Pam) Dengan Calinski-Harabasz Index Untuk Clustering Data Outlier," *UNEJ e-Proceeding*, hlm. 22–29, 2022.
- [5] I. M. K. Karo dan H. Hendriyana, "Klasifikasi Penderita Diabetes menggunakan Algoritma Machine Learning dan Z-Score," *Jurnal Teknologi Terpadu*, vol. 8, no. 2, hlm. 94–99, 2022.
- [6] S. Setyaningtyas, B. I. Nugroho, dan Z. Arif, "Tinjauan Pustaka Sistematis: Penerapan Data Mining Teknik Clustering Algoritma K-Means," *Jurnal Teknoif Teknik Informatika Institut Teknologi Padang*, vol. 10, no. 2, hlm. 52–61, 2022.
- [7] R. T. D. Kurniawati, R. Rahmawati, dan Y. Wilandari, "Pengelompokan Kualitas Udara Ambien Menurut Kabupaten/Kota di Jawa Tengah Menggunakan Analisis Klaster," *Jurnal Gaussian*, vol. 4, no. 2, hlm. 393–402, 2015.
- [8] H. Aulawi, W. A. Kurniawan, dan F. A. Rachman, "Analisis Sentimen Kepuasan Driver Terhadap Kebijakan Baru Sistem Order Gojek," *Jurnal Sains dan Teknologi ISTP*, vol. 14, no. 1, hlm. 86–94, 2020.
- [9] M. I. Hutagalung dan S. Sriani, "Pengelompokan Data Penyakit THT Menggunakan Algoritma K-Means Clustering: Grouping of ENT Disease Data Using K-Means Clustering Algorithm," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 4, hlm. 1568–1577, 2024.
- [10] D. Arkham dan D. Swanjaya, "K-Means Method For Clustering Public Service Assessment of Government Organization In Kediri City," dalam *Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi)*, 2020, hlm. 155–160.
- [11] A. M. Sikana dan A. W. Wijayanto, "Analisis Perbandingan Pengelompokan Indeks Pembangunan Manusia Indonesia Tahun 2019 dengan Metode Partitioning dan Hierarchical Clustering," *J. Ilmu Komput.*, vol. 14, no. 2, hlm. 66–78, 2021.
- [12] X. Wang dan Y. Xu, "An improved index for clustering validation based on Silhouette index and Calinski-Harabasz index," dalam *IOP Conference Series: Materials Science and Engineering*, IOP Publishing, 2019, hlm. 052024.
- [13] D. A. Rahmah, "Analisis klaster berdasarkan indikator kesejahteraan rakyat menggunakan metode Self Organizing Maps (SOM)," PhD Thesis, Universitas Islam Negeri Maulana Malik Ibrahim, 2022.
- [14] M. R. Nugroho, I. E. Hendrawan, dan P. P. Purwantoro, "Penerapan Algoritma K-Means Untuk Klasterisasi Data Obat Pada Rumah Sakit ASRI," *Nuansa Informatika*, vol. 16, no. 1, hlm. 125–133, 2022.
- [15] E. Ramadhani, N. Salwa, dan M. S. Mazaya, "Identifikasi Faktor-Faktor yang Memengaruhi Angka Harapan Hidup di Sumatera Tahun 2018 Menggunakan Analisis Regresi Spasial Pendekatan Area," *Journal of Data Analysis*, vol. 3, no. 2, hlm. 62–75, 2020.

²<https://www.kaggle.com/datasets/hendratno/covid19-indonesia>

ORIGINALITY REPORT

9%

SIMILARITY INDEX

7%

INTERNET SOURCES

5%

PUBLICATIONS

0%

STUDENT PAPERS

PRIMARY SOURCES

1	www.researchgate.net Internet Source	1 %
2	garuda.kemdikbud.go.id Internet Source	<1 %
3	perpustakaan.poltekkes-malang.ac.id Internet Source	<1 %
4	nbi.ku.dk Internet Source	<1 %
5	hdl.handle.net Internet Source	<1 %
6	jurnal.stis.ac.id Internet Source	<1 %
7	jurnal.untan.ac.id Internet Source	<1 %
8	magestic.unej.ac.id Internet Source	<1 %
9	pubs.ascee.org Internet Source	<1 %
10	sciencedocbox.com Internet Source	<1 %
11	Indah Fahmiah, Ratih Ardiati Ningrum. "Human Development Clustering in Indonesia: Using K-Means Method and Based on Human Development Index Categories", Journal of Advanced Technology and Multidiscipline, 2023	<1 %

-
- 12 Khandare, Anand D.. "Modified K-means algorithm for emotional intelligence mining", 2015 International Conference on Computer Communication and Informatics (ICCCI), 2015. <1 %

Publication

-
- 13 Zhao-Yu Wang, Chen-Yu Wu, Yan-Ting Lin, Shie-Jue Lee. "Weighted z-Distance-Based Clustering and Its Application to Time-Series Data", Applied Sciences, 2019 <1 %

Publication

-
- 14 www.abuseipdb.com <1 %

Internet Source

-
- 15 Cucuk Nur Rosyidi, Susy Susmartini, Lu'lu' Purwaningrum, Satoshi Muraki. "Mismatch Analysis of Elementary School Desk and Chair Key Characteristics in Indonesia", Applied Mechanics and Materials, 2014 <1 %

Publication

-
- 16 link.springer.com <1 %

Internet Source

-
- 17 Zahid Halim, Muhammad Atif, Ahmar Rashid, Cedric A. Edwin. "Profiling Players Using Real-World Datasets: Clustering the Data and Correlating the Results with the Big-Five Personality Traits", IEEE Transactions on Affective Computing, 2019 <1 %

Publication

-
- 18 journals.ametsoc.org <1 %

Internet Source

-
- 19 www.coursehero.com <1 %

Internet Source

-
- 20 www.harappadna.org <1 %

Internet Source

21 Andrea Cerioli. "Comparing three partitions: an inferential approach based on multi-way contingency tables", Communications in Statistics - Theory and Methods, 1997
Publication

22 core.ac.uk
Internet Source

23 discovery.researcher.life
Internet Source

24 journal2.unusa.ac.id
Internet Source

25 www.mdpi.com
Internet Source

26 www.researchpublish.com
Internet Source

27 www.slideshare.net
Internet Source

28 Corné Hoekstra, Sander Martens, Niels A. Taatgen. "Testing the skill-based approach: Consolidation strategy impacts attentional blink performance", PLOS ONE, 2022
Publication

29 Minshi Peng, Brie Wamsley, Andrew G Elkins, Daniel H Geschwind, Yuting Wei, Kathryn Roeder. "Cell type hierarchy reconstruction via reconciliation of multi-resolution cluster tree", Nucleic Acids Research, 2021
Publication

30 Xiaoyong Bian, Tianxu Zhang, Xiaolong Zhang, LuXin Yan, Bo Li. "Clustering-Based Extraction of Near Border Data Samples for Remote Sensing Image Classification", Cognitive Computation, 2012
Publication

31	Yue Shen, Cao Lei. "Research on evaluation of university education informatization level based on clustering technique", Heliyon, 2024 Publication	<1 %
32	doaj.org Internet Source	<1 %
33	ijccse.iasv.ru Internet Source	<1 %
34	journal.ubb.ac.id Internet Source	<1 %
35	journalinstal.cattleyadf.org Internet Source	<1 %
36	journals.scholarpublishing.org Internet Source	<1 %
37	statacumen.com Internet Source	<1 %
38	www.semanticscholar.org Internet Source	<1 %
39	Kofi Obeng. "Bus ride times of exceptional school children", Transportation Planning and Technology, 1993 Publication	<1 %
40	Wahyu Nur Fadillah, Yulita Molliq Rangkuti, Ichwanul Muslim Karo Karo. "ANALYZING THE EFFECT OF SIMILARITY FUNCTIONS ON PARTITIONING AROUND MEDOIDS ALGORITHM FOR MAPPING DHF DISEASE IN NORTH SUMATRA", BAREKENG: Jurnal Ilmu Matematika dan Terapan, 2024 Publication	<1 %
41	Isma Muthahharah, Sitti Masyitah Meliyana, Ansari Saleh Ahmar, Abdul Rahman. "K-Means Cluster for Grouping Regencies/Cities in South	<1 %

42

Son, Il-Young, Markus Guhe, Wayne D. Gray,
Birsen Yazici, Michael J. Schoelles, and Nancy
Jo Wesensten. "", Biomonitoring for
Physiological and Cognitive Performance
during Military Operations, 2005.

Publication

<1 %

Exclude quotes Off

Exclude matches Off

Exclude bibliography On