CLUSTER ANALYSIS OF CITIES/DISTRICTS IN WEST KALIMANTAN BASED ON STUNTING RESPONSE INDICATORS USING THE CALINSKI HARABASZ INDEX

TEGAR RAMA PRIYATNA
NIM H1011211014

FINAL YEAR PROJECT

DEPARTMENT OF MATHEMATICS
FACULTY OF MATHEMATICS AND NATURAL SCIENCES
UNIVERSITAS TANJUNGPURA
PONTIANAK
2025

CLUSTER ANALYSIS OF CITIES/DISTRICTS IN WEST KALIMANTAN BASED ON STUNTING RESPONSE INDICATORS USING THE CALINSKI HARABASZ INDEX

TEGAR RAMA PRIYATNA
NIM H1011211014

Final year project
as one of the requirements to obtain a degree
in Bachelor of Mathematics at the Department of Mathematics

DEPARTMENT OF MATHEMATICS
FACULTY OF MATHEMATICS AND NATURAL SCIENCES
UNIVERSITAS TANJUNGPURA
PONTIANAK
2025

APPROVAL SHEET

Title of the final project	:	**Cluster Analysis Of Cities/Districts In West Kalimantan Based On Stunting Response Indicators Using The Calinski-Harabasz Index**
Student Name	:	Tegar Rama Priyatna
NIM	:	H1011211014
Department/ Study Program	:	Matematika/Matematika
Date of Pass	:	27 March 2025
Supervisor's Decree	:	No.2550/UN22.8/TD.O6/2024/2 September 2024
Examiner's Decree	:	No.949/UN22.8/TD.O6/2025/20 March 2025

Supervisor

Supervisor I	Supervisor II
Nur'ainul Miftahul Huda, S.Si.,M.Si.	Dr. Yundari, S.Si.,M.Sc.
NIP 199411142020122014	NIP 198310202008012012

Examiner

Chief Examiner	Member of Examiner
Dr. Bayu Prihandono, M.Sc.	Meliana Pasaribu, M.Sc.
NIP 197911152005011003	NIP 198710192019032006
Session Leader (concurrently examiner member)	Session Secretary (concurrently examiner member)
Nur'ainul Miftahul Huda, S.Si.,M.Si.	Dr. Yundari, S.Si.,M.Sc.
NIP 199411142020122014	NIP 198310202008012012

Authorize
Dean of the Faculty of Mathematics and Natural Sciences
Universitas Tanjungpura

Prof. Dr. Gusrizal, M.Si.
NIP 197108022000031001

ACADEMIC INTEGRITY STATEMENT

I am the undersigned:

Name	: Tegar Rama Priyatna
NIM	: H1011211014
Department/ Study Program	: Mathematics/ Mathematics
Faculty	: Faculty of Mathematics and Natural Sciences

Now declare that the Final Project scientific document presented does not contain elements of a violation of academic integrity by the Regulation of the Minister of Education, Culture, Research and Technology of the Republic of Indonesia Number 39 of 2021. Suppose in the future this Final Project scientific document contains elements of a violation of academic integrity per the provisions of the legislation, I am willing to accept academic sanctions and/or applicable legal sanctions.

Thus, this statement is to be used as appropriate.

Pontianak, March 27th, 2025

Tegar Rama Priyatna

NIM H1011211014

Analisis Klaster Kota/Kabupaten Di Kalimantan Barat Berdasarkan Variabel Indikator Penanganan Stunting Dengancalinski Harabasz Index

Abstrak

Stunting di provinsi Kalimantan Barat (Kalbar) pada 2023 mencapai 27% atau lebih tinggi 10% dari target nasional. Permasalahan terjadi tidak tepat sasaran dalam memprioritaskan daerah prioritas untuk ditangani karena indikator hanya berdasarkan jumlah penduduk dan jumlah ibu melahirkan. Tujuan penelitian yakni menjelaskan peran metode modifed K-Mean dan CHI dalam pembentukan klaster terbaik dan menafsirkan kondisi tiap klaster. Objek Penelitian menggunakan delapan variabel indikator yang diperoleh dari website sistem informasi keluarga dan BPS pada 2023 yang meliputi jumlah remaja hadir konseling, jumlah pemberian informed consent, jumlah kasus komplikasi, jumlah pengeluaran alokon, jumlah persediaan alokon, tingkat laju pertumbuhan penduduk,tingkat kepadatan penduduk, dan angka harapan hidup. Peran metode modified yakni pembentukan klaster dengan menganalisis satuan data tiap variabel, tingkat karakteristik tiap objek melalui jarak Euclidean, mencari nilai pusat klaster (centroid) kemudian dilakukan iterasi hingga hasil konsisten sebagai hasil pembentukan klaster sedangkan peran CHI melakukan evaluasi dari tiap jumlah klaster dari jumlah minimum yakni satu hingga maksimum yakni tujuh untuk memperoleh jumlah klaster terbaik. Diperoleh $K = 5$ klaster terbaik dengan status klaster 1 cukup buruk beranggotakan tiga objek, klaster 2 sangat buruk beranggotakan empat objek, klaster 3 cukup beranggotakan tiga objek, klaster 4 sangat baik beranggotakan tiga objek, dan klaster 5 baik beranggotakan satu objek.

Kata kunci: *Centroid, Modified K-Mean Cluster,* Jarak Euclidean.

Cluster Analysis Of Cities/Regencies In West Kalimantan Based On Stunting Handling Indicator Variables With Calinski-Harabasz Index

Abstract

Stunting in West Kalimantan (Kalbar) province in 2023 reached 27% or 10% higher than the national target. The problem is that there is a lack of targeting in prioritizing priority areas to be addressed because the indicators are only based on population and the number of mothers giving birth. The purpose of the study was to explain the role of the modified K-Means and CHI methods in forming the best cluster and interpreting the condition of each cluster. The research object uses eight indicator variables obtained from the family information system website and BPS-Statistics Indonesia in 2023 which include the number of adolescents attending counseling, the number of informed consent, the number of complication cases, the number of aslocon expenditures, the number of alocon supplies, the population growth rate, the population density level, and life expectancy. The eight variables used were selected because they had been analyzed using PCA with the Kaiser approach and PVE, which represented 76.76% of the information divided into three main dimensions. 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 the number of clusters from the minimum number of one to the maximum of seven to obtain the best number of clusters. The best $K = 5$ clusters are obtained, with the status of cluster 1 being quite bad, with three objects, cluster 2 is very bad, with four objects, cluster 3 is good, with three objects, cluster 4 is very good, with three objects, and cluster 5 is good, with one object.

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

ACKNOWLEDGEMENTS

All praise and gratitude to Allah SWT for blessings and grace, enabling the author to complete this final year project titled “Cluster Analysis Of Cities/Districts In West Kalimantan Based On Stunting Response Indicators Using The Calinski Harabasz Index” to obtain a Bachelor of Mathematics degree from the Mathematics Study Program, Faculty of Mathematics and Natural Sciences, Tanjungpura University.

The author realizes that the completion of this final year project would not have been possible without the support of the beloved family and the assistance of various parties. Therefore, the author would like to express gratitude to:

Nur’ainul Miftahul Huda, S.Si.,M.Si., as the first supervisor, and Dr. Yundari, M.Sc., as the second supervisor, who have dedicated their time, thoughts, and efforts to provide guidance and solutions to the author during the preparation and completion of this final year project.

Dr. Bayu Prihandon, M,Sc., as the first examiner, and Meliana Pasaribu, M.Sc., as the second examiner, who have provided guidance and motivation to the author in completing the final year project.

The family of mathematics students of the 2021 intake, who cannot be mentioned one by one, have provided motivation and knowledge to the author. The author is aware that this final year project still needs some suggestions, input, and constructive criticism. It is hoped that this final year project can make a significant contribution to the development of knowledge, particularly in the field of mathematics, and be beneficial for readers and utilized optimally.

Pontianak, March 27th ,2025

Tegar Rama Priyatna

CONTENS

APPROVAL SHEET ii
ACADEMIC INTEGRITY STATEMENT iii
Abstrak iv
ACKNOWLEDGEMENTS vi
CONTENS vii
LIST OF FIGURES ix
LIST OF TABLES x
LIST OF APPENDICES xi
LIST OF SYMBOLS xii
CHAPTER I INTRODUCTION 1
1.1 Research Background 1
1.2 Problem Statements 5
1.3 Research Aims 5
1.4 The Scope of the Study 5
1.5 Literatur Review 6
1.6 Research Methodology 6
CHAPTER II THEORITICAL BACKGROUND 11
2.1 Multivariat Analysis 11
2.2 Clustering Method 12
2.3 *Cluster Non Hierarki* 13
2.4 Descriptive Statistics 13
2.5 Data Standardization Using Z-Score 14
2.6 *Euclidean Distance* 15
2.7 *Centroid* 15
2.8 *K-Mean Cluster* 15
2.9 Cluster Labeling 17
CHAPTER III PENENTUAN JUMLAH *CLUSTER* OPTIMAL 18
3.1 Modified K-Means Clustering Method 18

3.2 ***Calinski Harabasz Index*** 19
3.3 **Data Process** 21
3.3.1 **Deskripsi Data** 21
3.3.2 **Data Standardization** 26
3.4 **Calculating Euclidean Distance** 28
3.5 **Initial Determination of *K*** 28
3.6 **Cluster Formation Using the Modified K-Means Method** 29
3.6.1 **Initial Cluster Formation** 29
3.6.2 **Finding the initial centroid value of the cluster** 30
3.6.3 **Calculating the Distance of Each Object to the Centroid** 32
3.6.4 **Object Placement in Clusters** 33
3.6.5 **Evaluating Centroid Values and Object Placement** 35
3.7 **Sum of Squared Within Clusters (SSW) Value** 37
3.8 **Sum of Squared Between Clusters (SSB) Value** 38
3.9 **Calinski-Harabasz Index (CHI) Value** 38
3.10 **Cluster Interpretation** 41
CHAPTER IV CONCLUSIONS 45
4.1 **Conclusion** 45
4.2 **Recommendations** 45
REFERENCES 47
APPENDICES 54

LIST OF FIGURES

Figure 1. 1 Stunting Achievements and Targets.....3
Figure 1. 2 Chart of Targets and Indicators.....3
Figure 1. 3 Flowchart for Determining the Best Number of Clusters 10
Figure 2. 1 Characteristics of Cluster Formation.....16
Figure 3. 1 Map of West Kalimantan province 21
Figure 3. 2 Initial Cluster Formation 30
Figure 3. 3 Object Placement in Clusters.....34
Figure 3. 4 Map of Best Clusters for Stunting Management.....41

LIST OF TABLES

Table 3. 1 Stunting Handling Indicator Variable Data 22
Table 3. 2 Data on Stunting Handling Indicator Variables 23
Table 3. 3 Descriptive Statistics of Data 26
Table 3. 4 Standardize data 27
Table 3. 5 Euclidean Distance of each Member 28
Table 3. 6 Initial Centroid Values for Each K 31
Table 3. 7 Initial Initialization 33
Table 3. 8 New Centroids from Process 1 36
Table 3. 9 Determining the Optimal Number of Clusters 38
Table 3. 10 Cluster Members for K=5 (Final Results) 39
Table 3. 11 Final Centroid Values and Cluster Statuses 39
Table 3. 12 Cluster Interpretation 40

LIST OF APPENDICES

Appendix 1 PCA process results......54
Appendix 2 Indicator Variable Data of Stunting Handling.....55
Appendix 3 Initial Centroid Values.....56
Appendix 4 Initial Initialization Value.....57
Appendix 5 Final Centroid Value.....58

LIST OF SYMBOLS

Z_{ij}	:	Standardization of data for the $i-$th object on the $j-$th variable
x_{ij}	:	represents the value of the$-i$ object observed on the$-j$ indicator variable, where $i = 1,2,3,\dots,N$ dan $j = 1,2,3,\dots,n$
N	:	Number of Objects
n	:	Number of Variables
$\bar{x}_j$	:	The average value of the j-indicator variable, where $j = 1,2,3 \dots,n$
s_j	:	Standard deviation of j-indicator variables, where $j = 1,2,3 \dots,n$
$d(x_i, x_t)$	:	Distance value between $i-$th object and $t-$th object, where $i = 1,2,3,\dots,N$ dan $t = 1,2,3,\dots,N$
$d(c_l, x_j)$	:	The distance value between the $l-$th centroid and the $t-$th object, where $l = 1,2,3,\dots,n$ dan $t = 1,2,3,\dots,N$
c_{lj}	:	The l $-$th centroid on the j $-$th variable, where $l = 1,2,3,\dots,n$ dan $j = 1,2,3,\dots,n$
m_k	:	Number of objects in the $k-$th cluster, *where* $k = 1,2,3,\dots,K$
K	:	Number of Clusters
$C_k X_j$	:	Interpretation on the $k-$th cluster and $j-$th variable, where $k = 1,2,3,\dots,K$ dan $j = 1,2,3,\dots,n$
N_k	:	The number of objects belonging to the $k-$th cluster, dengan $k = 1,2,3,\dots,K$

CHAPTER I
INTRODUCTION

1.1 Research Background

Human development and the improvement of healthcare service quality have become focal points for realizing high-quality human resources. Superior human resource quality is a key national agenda of the Negara Kesatuan Republik Indonesia (NKRI) to achieve a Golden Generation by 2045, and this aligns with international objectives through the Sustainable Development Goals (SDGs). SDGs are a program for the sustainable development of human quality, with health and humanitarian issues addressed in the third main goal, namely 'good health and well-being.' The third SDG aims to ensure the achievement of equitable health and the improvement of well-being for all individuals across all ages.

The problem faced in implementing the third point of the SDGs is regarding the prevalence of stunting. The challenge faced in implementing the third point of the SDGs is related to the prevalence of stunting. Stunting is a condition related to an individual's height not in accordance with the expected growth conditions based on age and gender, indicating prolonged undernutrition when stunting occurs, indicating a long-term malnutrition status (Aryu, 2020). Stunting occurs in children during their growth and development period, which in fact can be anticipated by monitoring their nutritional needs.

Fulfilling nutritional needs plays an important role in children's growth and development. Growth affects changes in the body caused by the increase in physical body size (the increase in the size and number of cells), which influences a child's height. Meanwhile, development affects the size and structure of the body, either partially or as a whole, which will impact the child's weight. (Melianti dkk., 2023)

Fulfilling children's nutritional needs not only affects their physical growth and development but also influences brain development, which directly impacts the development of cognitive the capacity to think, explore, and solve problems, including both simple and complex issues (Mahayuna dkk., 2021). Addressing the problem of stunting is crucial because prevention can only occur during the first 1,000 days of life, starting from pregnancy. If a child is already stunted, the resulting permanent problems-particularly in cognitive abilities-cannot be corrected even with improved nutrition after this period, as the brain tends not to develop optimally (Nakita, 2025).

The problem of stunting continues to occur due to the lack of information and parental awareness regarding children's nutritional status, which generally leads to malnutrition. If a child's nutritional deficiencies are left unaddressed and become chronic, this can affect their mental condition, psychomotor development, and intelligence, which are all correlated with the third SDG goal and the vision for a Golden Generation in 2045.

The prevalence of stunting in Indonesia remains relatively high because it does not meet the World Health Organization (WHO) standard, which is a maximum of less than 20%, while Indonesia, based on survey status gizi Indonesia (SSGI) which was delivered by Badan Kependudukan dan Keluarga Berencana Nasional (BKKBN) Indonesia In 2022, the stunting rate was still recorded at 21.6%, a decrease of 2.8% from the previous year. To address this issue, the Indonesian government has set a target to reduce the stunting rate to 14% by 2024 (Kesehatan, 2023). To achieve this target, the government is coordinating with all stakeholders and has designated 12 priority provinces for accelerated stunting reduction based on the 2021 SSGI, with West Kalimantan (Kalbar) being one of the focus provinces.

According to data from Badan Pusat Statistik (BPS) of West Kalimantan Province, Kalbar recorded economic growth of 4.98% in the first quarter of 2024, indicating a positive trend. However, the economy is closely related to stunting; the 2023 Semester I Report on the Acceleration of Stunting Reduction states that stunting contributes to a 2%–3% loss in gross domestic product (GDP) each year (Wardoyo, 2023).

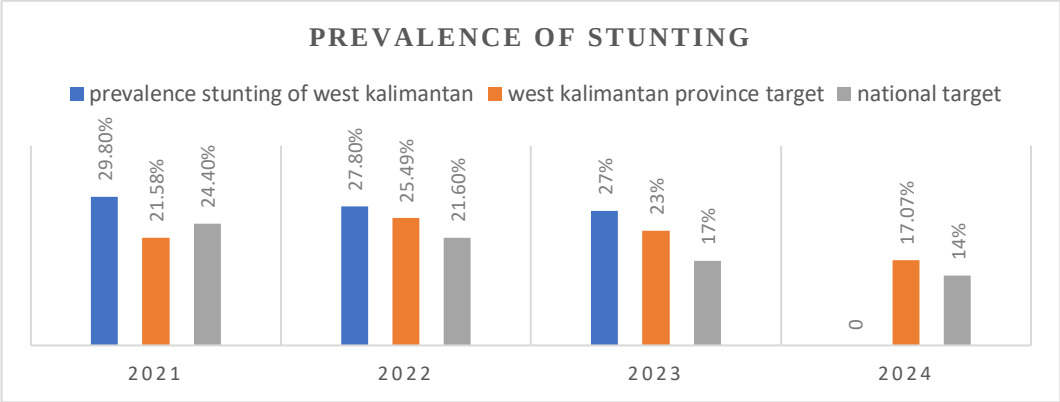


Figure 1. 1 Stunting Achievements and Targets

Based on Figure 1.1 This indicates a serious commitment to addressing stunting, as from 2021 to 2023, neither the national nor provincial targets were achieved, and for 2024, the value is recorded as zero because the stunting prevalence data is not yet available. The failure to meet the stunting prevalence targets should actually refer to the data on target Pasangan Usia Subur (PUS) from BKKBN Kalbar in 2023, as presented in Figure 1.2. This data shows that the selection of priority areas was based solely on the number of births and population, whereas there are many other, more relevant indicators that could better represent the selection of priority regions for intervention.

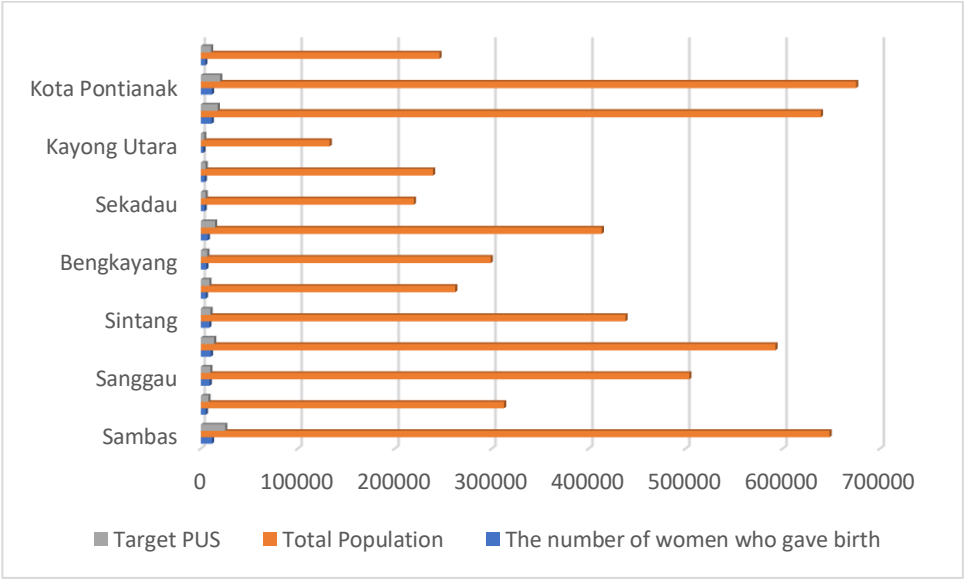


Figure 1. 2 Chart of Targets and Indicators

Based on Figure 1.2, it can be seen that several areas prioritized in stunting intervention include Pontianak, Kubu Raya, and Sambas. This indicates the need for evaluation to improve and accelerate stunting reduction efforts, as these three priority regions have better accessibility and infrastructure compared to other cities/regencies such as Kapuas Hulu, which is located far from the capital of West Kalimantan Province.

To optimize stunting prevalence reduction so that it is more effective, it is necessary to examine the specific characteristics of stunting issues in all 14 cities/regencies. Once the characteristics of stunting problems are identified, the government can determine priority areas for the Kalbar Zero Stunting program to ensure the projected targets are achieved. Cluster analysis methods can assist in prioritizing regions by grouping cities or regencies in West Kalimantan based on similar characteristics into clusters. The Modified K-Means Cluster is a clustering analysis that emphasizes heterogeneity between clusters and the similarity of each member within a cluster, using Euclidean distance calculations. Therefore, in forming clusters, the initial distance used as a reference is not determined by the researcher's preference but by the similarity of characteristics among the cities/regencies, which can improve the quality and efficiency of cluster formation. However, cluster analysis has a limitation in that it cannot independently determine the optimal number of clusters to be formed.

To determine the optimal number of clusters, the Calinski-Harabasz Index (CHI) method is used, as it evaluates the feasibility of the number of clusters formed by using the Sum of Squared Within-cluster (SSW) value to measure the compactness, or the similarity of objects within the same cluster, and the Sum of Squared Between-cluster (SSB) value to measure the separation, or the level of homogeneity between clusters. Thus, a higher CHI value indicates the formation of the optimal number of clusters, where each cluster consists of members with increasingly similar characteristics and the clusters themselves are more homogeneous. (Mubarak, 2023).

1.2 Problem Statements

Based on the background presented, the problem statements in this study include the following:

1. What is the role of Modified K-Mean Cluster and Calinski Harabasz Index in forming the best cluster?
2. What is the condition of each cluster based on the number of best clusters obtained?

1.3 Research Aims

Based on the problem statements that have been presented, the objectives of the research conducted are:

1. Explain the role of modifed K-Mean and CHI methods in the formation of the best cluster.
2. Interpretation the condition of each cluster based on the number of best clusters obtained.

1.4 The Scope of the Study

To limit the problem with the aim that the discussion is focused and does not expand, it is necessary to limit the problem, which includes:

1. 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, the number of severe complication and failure cases, the availability of contraceptive devices and drugs, the distribution of alokon, the number of adolescents attending counseling at youth and student information and counseling centers, life expectancy, population growth rate, and population density 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.

2. The methods used for cluster formation are the Modified K-Means clustering method and the Calinski-Harabasz Index to determine the optimal number of clusters.

1.5 Literatur Review

Choldun and Rachmatullah (2022) found that the accuracy achieved using the Modified K-Means Cluster method was higher than that of the K-Means Cluster method, with the Modified K-Means Cluster reaching 89.33% accuracy, while the K-Means Cluster achieved 82.67%.

Antonius, Chrismanto, and Sebastian (2020) found that the Modified K-Means method was able to produce more consistent clusters and better results in terms of accuracy and cluster quality. The application of the Modified K-Means method resulted in better clusters, as indicated by a purity value of 0.42 (42%) for Modified K-Means, compared to 0.391 (39.1%) for K-Means.

Suraya and Wijayanto (2022) found that the two best methods were Average Linkage, with the optimal number of clusters being four-performing well in evaluations using the DI, DBI, and XBI methods-and K-Means, as the second-best method, with the optimal number of clusters being two, performing well in evaluations using the SC and CHI methods.

Mardhotillah, Fadli, Elisa, and Zurweni (2023) found that the K-Means Cluster method was better than the Fuzzy C-Means method, as indicated by the Calinski-Harabasz Index value of 16.308 compared to 12.008, with the optimal number of clusters being two

1.6 Research Methodology

The first step in determining the number of clusters to be formed (K) for grouping cities/regencies in West Kalimantan Province based on stunting management factors is to conduct a descriptive analysis of the data using eight indicators: the number of adolescents attending counseling (X_1), the number of informed consent forms given during family planning services (X_2), the number of severe complication and failure cases (X_3), the number of contraceptive devices distributed (X_4), the number of contraceptive devices available (X_5), the annual

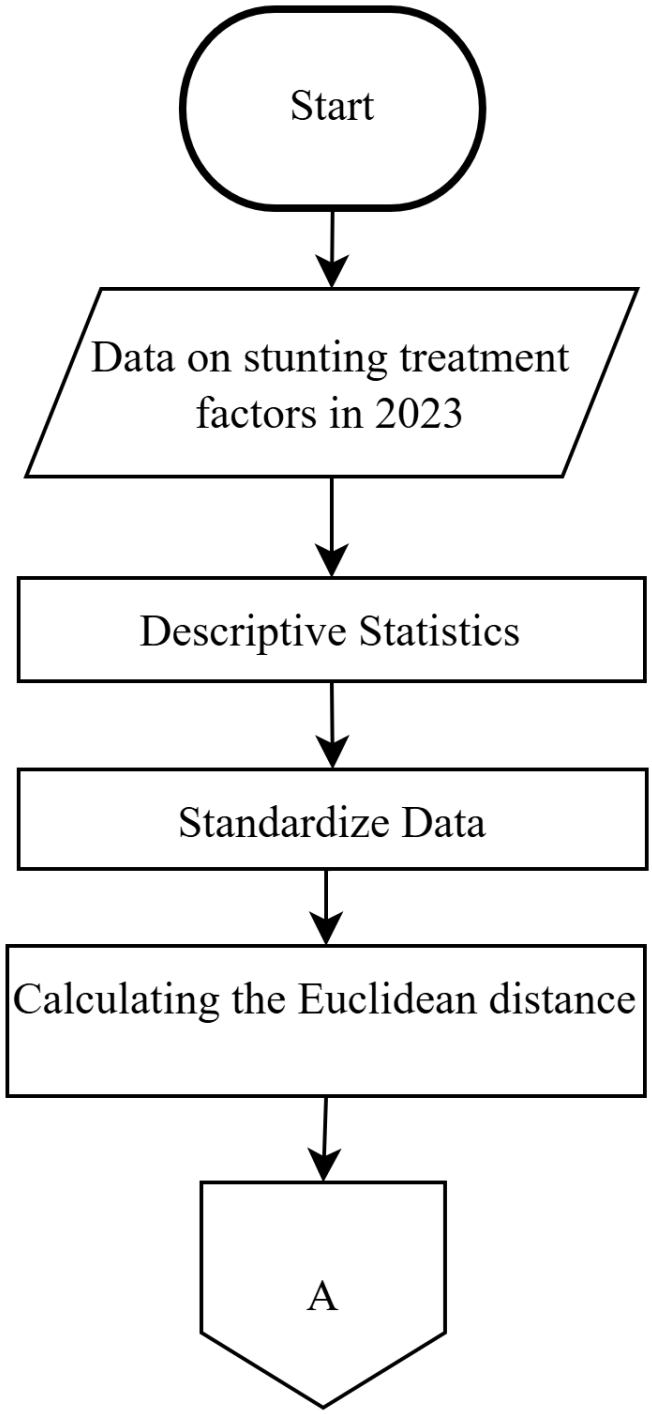
population growth rate (X_6), population density per square kilometer (X_7), and life expectancy (X_8).

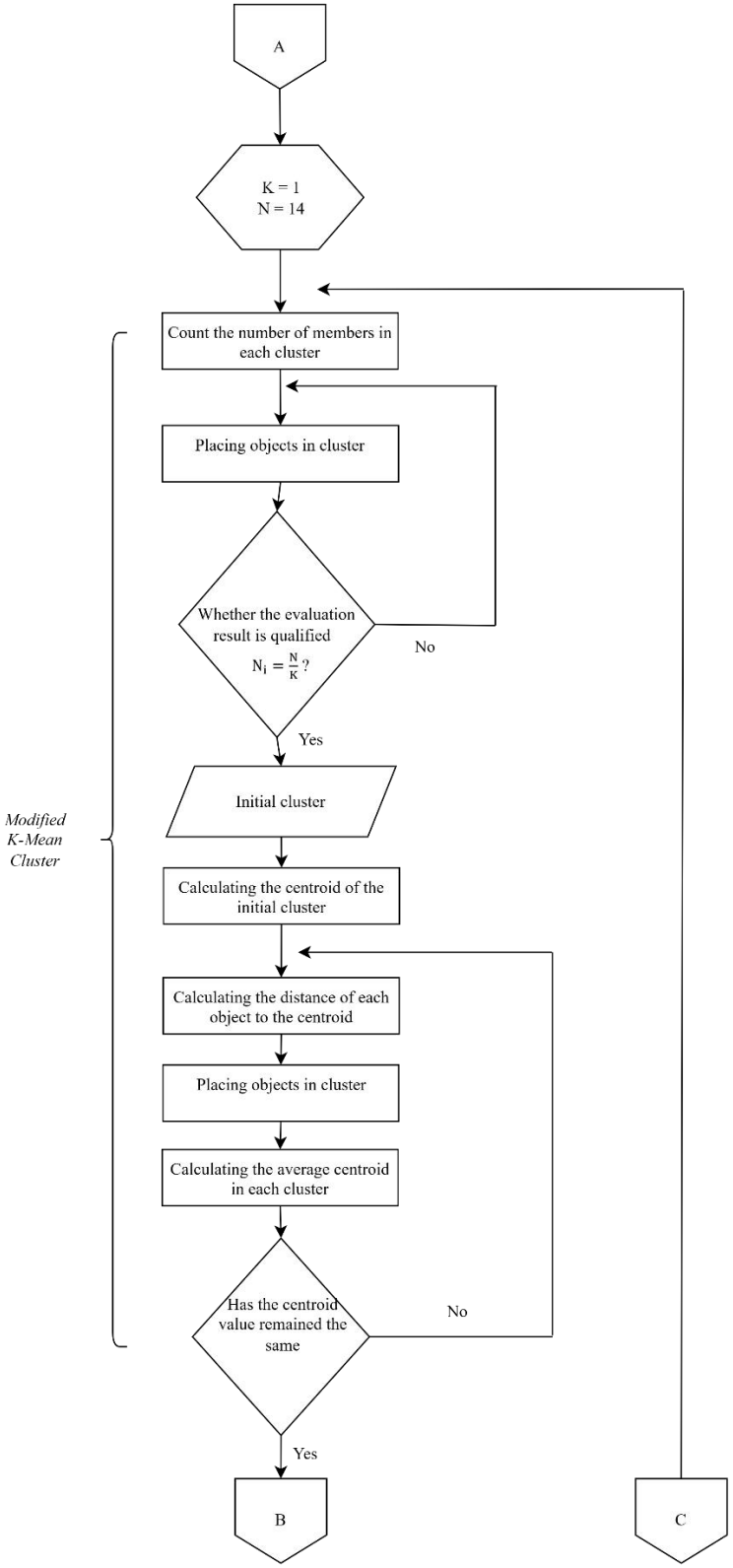
ata normalization is performed using the z-score method because the data used varies considerably in units. Additionally, the purpose of data normalization is to avoid bias and to improve proportional performance, resulting in more stable calculations. After the data has been normalized, distance calculation is carried out using the Euclidean distance to measure the similarity between data objects. Initial clusters are then formed based on the minimum value of the Euclidean distance, ensuring that each cluster contains N/K members. The process begins from K_min = 1 to K_max = 7, and the threshold value is determined through a trial system by ensuring that each member of the initial cluster remains within the N/K range. Once the initial clusters are formed, the initial centroid is calculated using the centroid formula, which is the mean of the Euclidean distances within each initial cluster.

The next step is to segment the data based on the initial centroid values by calculating the Euclidean distance of each object to each of the K centroid values. For each object, the smallest distance is selected for segmentation. Then, new centroids are calculated for each segment or cluster that has been formed, and the process is repeated until the centroid values do not change or the cluster membership of the objects remains unchanged from the previous step. At this point, these centroid values become the final centroids.

For clusters numbering two, the Sum of Squares Within Cluster (SSW) is calculated for each cluster formed. Next, the Sum of Squares Between Clusters (SSB) is calculated between each cluster. The SSW and SSB values obtained are then used in the next step to calculate the CHI value. The CHI value is determined by considering the final centroid values, SSW, and SSB from $K_{\min}$ to K_{max} to identify the optimal number of clusters based on the highest CHI value. Once the optimal number of clusters is determined, each cluster is interpreted to understand its characteristics by examining the final centroid values. The final step is to visualize the clusters and present recommended actions that can be taken for each

cluster. The analysis steps in this study can be seen in the flowchart shown in Figure 1.3.




A
K = 1
N = 14
Count the number of members in each cluster
Placing objects in cluster
Whether the evaluation result is qualified
$N_i = \frac{N}{K}$?
No
Yes
Initial cluster
Calculating the centroid of the initial cluster
Calculating the distance of each object to the centroid
Placing objects in cluster
Calculating the average centroid in each cluster
Has the centroid value remained the same
No
Yes
B
C
Modified K-Mean Cluster

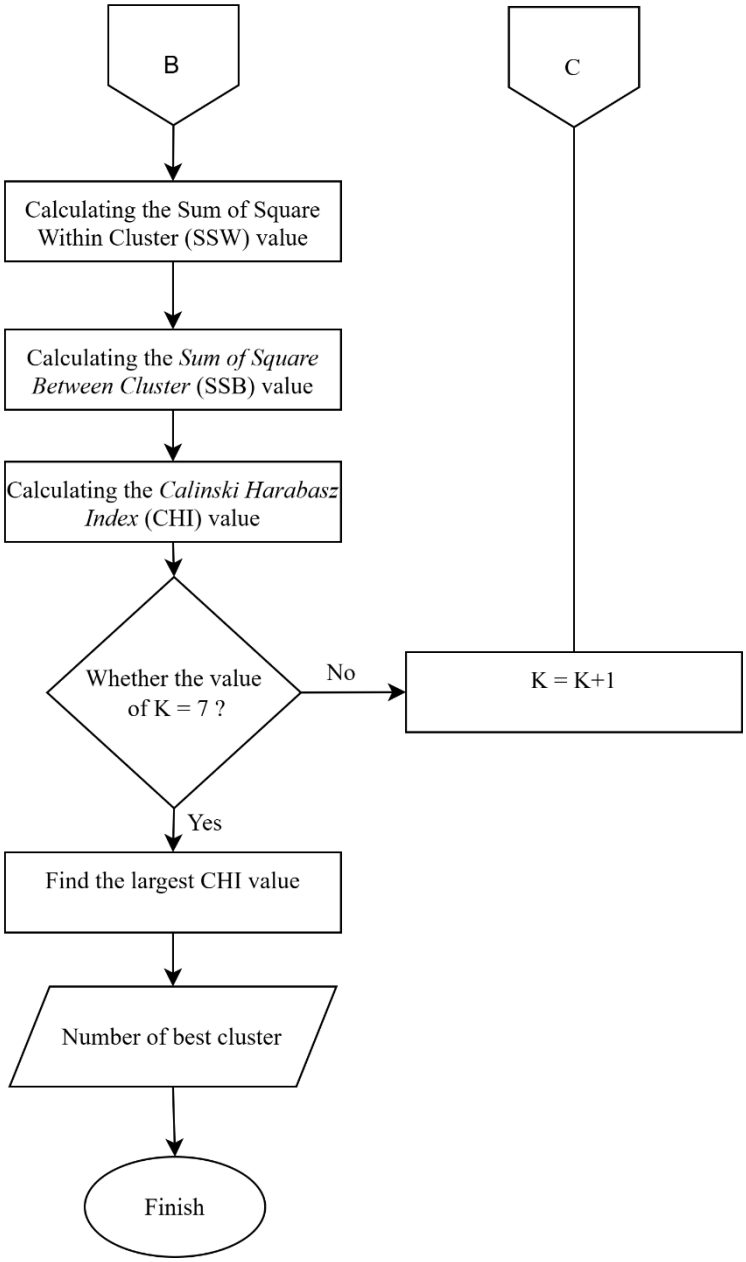


Figure 1. 3 Flowchart for Determining the Best Number of Clusters