





## Stunting Incidence Segmentation: A Cluster Analysis Approach and Targeted Intervention Strategies



Sri Mulyati<sup>\*</sup>, Delvindra Faiz Noorhadi<sup>1</sup>, Hanuga Fathur Chaerulisma<sup>1</sup>, Novi Setiani<sup>1</sup>

Department of Informatics, Faculty of Industrial Technology, Universitas Islam Indonesia, Yogyakarta 55281, Indonesia

Corresponding Author Email: [mulya@uii.ac.id](mailto:mulya@uii.ac.id)

Copyright: ©2025 The author. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ijcmem.130113>

### ABSTRACT

**Received:** 11 January 2025

**Revised:** 12 March 2025

**Accepted:** 16 March 2025

**Available online:** 31 March 2025

#### Keywords:

*K-Means clustering, elbow method, cluster profiling, data-driven strategies, intervention prioritization*

This study develops a data-driven strategy for stunting prevention using the K-Means clustering method, validated through the Elbow Method and Cluster Profiling. The high prevalence of stunting in the research area highlights the need for precise health condition mapping to prioritize effective interventions. Data collected from toddlers in the region were grouped into three distinct clusters, each representing varying levels of risk and requiring tailored prevention strategies. These interventions include contextualized preventive education, optimized based on the specific characteristics and needs of each cluster. The results demonstrate that this method accurately maps health conditions, facilitates targeted interventions, and enhances resource allocation. Additionally, the clustering approach serves as a foundation for creating impactful and relevant health counseling materials to strengthen community education. The study's main contribution lies in providing a data-driven framework that supports evidence-based public health policy and localized stunting prevention strategies, ensuring adaptability to the unique needs of the research area.

## 1. INTRODUCTION

Childhood stunting remains a pressing public health issue in many developing countries, requiring multifaceted strategies for effective mitigation [1]. Stunting, defined as a height-for-age measurement falling below two standard deviations from the Child Growth Standards set by the World Health Organization (WHO), arises from a combination of factors, including malnutrition, inadequate sanitation, and underlying health conditions [2, 3]. Its causes are multifactorial, encompassing chronic malnutrition, inadequate sanitation, and poor health conditions. Effective mitigation demands precise health condition mapping to identify priority areas for intervention. Accurate regional data enables tailored prevention strategies, aligning resources with community-specific needs to optimize outcomes [4].

The WHO's 2025 global nutrition target and Sustainable Development Goal (SDG) 2.2.1 aim to reduce the prevalence of stunting among children under five by 2025. Our findings suggest that targeting stunting at an earlier age, such as 3 to 6 months, can maximize the impact of interventions [5].

Stunting remains a significant public health challenge, particularly in countries with high levels of poverty and hunger, such as Indonesia. Characterized by impaired growth and development due to chronic malnutrition, stunting severely impacts children's cognitive and physical development, potentially leading to long-term social and economic consequences [6]. According to the 2022 Indonesia Nutrition Status Survey (SSGI), the national stunting rate has decreased

to 21.6%. Despite this decline, the figure remains above the 20% threshold set by the WHO, indicating the need for continued attention to stunting as a serious public health issue.

In District X, Central Java, the prevalence of stunting showed an alarming increase from 15.8% to 18.2% in 2022. This condition emphasizes the need for more targeted and effective interventions. Traditional methods to overcome stunting [7], such as growth monitoring and nutritional supplementation, have proven less effective in adapting to dynamic field conditions. Therefore, integrating digital-based solutions and data-driven approaches into public health strategies has become increasingly important [8].

The stunting measurement report received by the health office is used to determine strategic decisions in handling conditions in the community [9-12]. Although this aids in decision-making to meet community needs, the health office faces significant challenges in delivering effective health counseling. One major issue is the limited availability of human resources and the lack of speakers familiar with local conditions, which reduces the effectiveness of extension material delivery [13]. Additionally, there is currently no specific, targeted content tailored to the unique needs and conditions of each region.

This issue is further exacerbated by the difficulties in effectively mapping the conditions of regional conditions, making it difficult to determine the process of developing appropriate extension materials. To address this problem, a more structured and data-driven approach, as well as grouping, is needed to ensure that counseling is not only on target but

also delivered by competent experts with relevant and specific materials. UNICEF's report on stunting in South Asia highlights the importance of interventions that prioritize child feeding, women's nutrition, and household sanitation as priorities to prevent stunting [14, 15].

Several studies have also highlighted the potential of data-driven approaches in improving our understanding of stunting patterns and guiding intervention strategies [15, 16]. These findings emphasize the need for localized strategies in stunting prevention, which serve as a foundation for developing more relevant health counseling materials.

This study addresses these gaps by integrating behavioral and environmental factors, such as maternal health, family smoking habits, and access to clean water, into the K-Means clustering framework. Unlike prior research that primarily focuses on anthropometric measures, this approach provides a more localized and comprehensive perspective on stunting prevention strategies. Community-based interventions, particularly those targeting maternal health and nutrition, have proven effective in reducing stunting rates.

Stunting remains a critical public health challenge in District X, Indonesia, where its prevalence increased from 15.8% to 18.2% in 2022. Prevention efforts currently face significant challenges, including limited human resources, ineffective material delivery, and inadequate regional health mapping. This study aims to develop a data-driven approach using K-Means grouping and Elbow method to classify regions based on health conditions and risk factors. The study seeks to create targeted intervention strategies and locally relevant health counselling materials, ultimately supporting resource allocation and more effective policy decisions in stunting prevention efforts.

By utilizing K-Means clustering and the Elbow method, we seek to identify different groups of health conditions in toddlers, providing input for more targeted and effective intervention strategies. This input takes the form of outreach material that is appropriate to regional conditions. The application of data collection and machine learning techniques in the field of public health, particularly in the context of child malnutrition and stunting, has gained attention in recent years. Previous research highlights the potential of machine learning in analyzing public health issues, including stunting.

The K-Means algorithm, combined with the Elbow Method, has been recognized in healthcare studies as an effective tool for segmenting patient groups and predicting health conditions. A recent study [16] evaluates the inertia-based index to determine the optimal number of clusters in K-Means, while other research demonstrates the effectiveness of the Elbow method in optimizing KNN for stroke prediction [17]. Some studies have also identified four key risk factors for stunting in Indonesia and classified malnutrition status in children [18].

Nevertheless, there is still a gap in the application of clustering techniques for mapping health conditions specifically for stunting prevention at the local level. This research addresses this gap by applying K-Means clustering and the Elbow method to a unique dataset from the District X, covering various health indicators relevant to stunting. By focusing on the local context and integrating multiple health factors, this study aims to provide insights for more targeted and effective stunting prevention strategies at the community level [9].

While these studies have made significant contributions, there remains a gap in the literature regarding the application of clustering techniques to map health conditions specifically

for stunting prevention at the local level. Moreover, the integration of multiple health indicators beyond traditional anthropometric measures in clustering analyses warrants further exploration.

Our study builds on this body of work by applying K-Means clustering and the Elbow method to a unique dataset from District X, incorporating a range of health indicators relevant to stunting. By focusing on a specific local context and considering multiple health factors, our research aims to provide insights for more targeted and effective stunting prevention strategies at the community level.

Although previous studies have explored clustering methods in healthcare, few have utilized localized data to develop stunting prevention strategies tailored to the unique health conditions of local communities. This study introduces a new approach by applying K-Means clustering and Elbow methods to specific local datasets from District X, Indonesia, to identify health risk groups that require more contextual stunting interventions.

This study combines indicators such as access to clean water, worm infections, and health history of pregnant women that have not been widely applied in clustering analysis for the context of stunting prevention. The resulting clusters provide a more accurate picture of the distribution of health risks and enable the development of more relevant and data-driven counseling strategies for each risk group in the local community. This approach enhances the effectiveness of counseling and resource allocation at the community level and has the potential to serve as a model for data-driven health interventions in other areas with similar characteristics.

This research aims to contribute to stunting prevention by applying data clustering techniques to map the health conditions of toddlers in District X. By utilizing the K-Means algorithm and the Elbow method, this research identified different health condition groups to inform more targeted intervention strategies. The outputs include extension materials tailored to regional conditions, providing a more effective approach to prevent stunting. The application of data collection and machine learning techniques in public health, particularly in addressing child malnutrition and stunting, has gained significant attention in recent years. Previous studies have demonstrated the potential of machine learning in analyzing public health issues, including stunting.

## 2. METHOD

This study uses a combination of the K-Means Clustering and the Elbow method. The Elbow method determines the optimal number of clusters for clustering using the K-Means method. The research steps are outlined in Figure 1.

### 2.1 Health data collection

The health data for this study were obtained from the Ministry of Health, District X, Central Java, Indonesia. The dataset consists of 231 toddlers with malnutrition in District X. Although the dataset size is relatively small (231 samples), the use of machine learning techniques, particularly K-Means clustering, is justified for several reasons. Firstly, this dataset represents a specific local context in District X, offering highly relevant insights for targeted interventions in stunting prevention. Secondly, the exploratory nature of this research aims to identify patterns and generate hypotheses rather than

produce broad generalizations.

In public health contexts where comprehensive health data is difficult to obtain, utilizing available data, however limited, is crucial for initiating data-driven approaches. Moreover, there are precedents in public health literature where machine learning techniques have been effectively applied to relatively small datasets, yielding meaningful results [18, 19].

K-Means clustering is computationally efficient and performs well with small datasets, especially when combined with the Elbow method to determine the optimal number of clusters [20, 21]. This study serves as a proof of concept and foundation for further research, potentially encouraging more comprehensive data collection in the future and offering the potential for incremental learning as additional data become available.

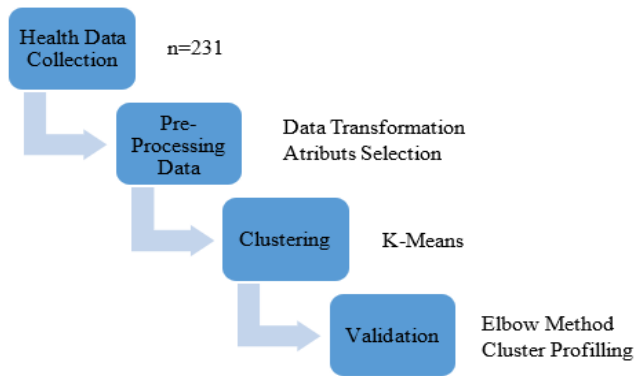


Figure 1. Research steps

## 2.2 Data preprocessing

Data preprocessing is essential to achieve accurate results. The preprocessing steps in this study include Data Transformation, Attributes Selection, and Data Grouping. Data Transformation is used to convert categorical data into nominal data. Attributes Selection removes unnecessary attributes. Data grouping aims to organize the toddler's data by sub-district, allowing for more localized analysis.

Table 1. Relevance to stunting

Attribute	Relevance to Stunting
Access to clean water	Contaminated water can cause gastrointestinal infections, contributing to chronic malnutrition and stunting
Worm Infection	Children infected with worms often experience nutrient absorption disorders that hinder growth
Healthy toilet	Poor sanitation increases the risk of infections that can impede a child's growth
Immunization	Incomplete immunization can increase the risk of infectious diseases, which negatively impact nutrition and child growth.
Smoking habits in the family	Exposure to cigarette smoke increases the risk of respiratory health disorders, which can hinder a child's growth.
Maternal health history	A mother's health condition before and during pregnancy significantly affects fetal development and the risk of stunting in children

The attributes used in this are clean water, worm infection, healthy toilet, immunization, smoking (family), and maternal history. In the context of this study, data-driven strategies are

applied through clustering techniques to identify areas with high-risk levels, enabling more effective and targeted interventions. The reasons for selecting the attributes used in clustering are: Each attribute is chosen based on its relationship with key risk factors contributing to stunting. Below is a more detailed explanation of how each attribute relates to the risk of stunting (refer to Table 1).

## 2.3 K-Means clustering

The K-Means clustering algorithm is implemented to partition the health data into optimal groups, aiming to minimize the within-cluster sum of squares through an iterative process. The mathematical formulation is as follows:

Given a set of observations  $(x_1, x_2, \dots, x_n)$ , where each observation represents a toddler's health indicators, K-Means clustering partitions the  $n$  observations into  $k$  sets  $S = \{S_1, S_2, \dots, S_k\}$  by minimizing:

$$\arg \min \sum_{i=1}^k \sum_{x \in S_i} \|x - \mu_i\|^2 \quad (1)$$

where,

$\mu_i$  is the mean of points in  $S_i$

$\|x - \mu_i\|^2$  represents the squared Euclidean distance

The algorithm executes in two primary steps:

Assignment step:  $S_i^{(t)} = \{x: \|x - \mu_i^{(t)}\|^2 \leq \|x - \mu_j^{(t)}\|^2 \forall j, 1 \leq j \leq k\}$

Update step:  $\mu_i^{(t+1)} = 1/|S_i^{(t)}| \sum_{x \in S_i^{(t)}} x$

## 2.4 Validation

This study builds upon findings from A Comparative Study of Clustering Algorithms in Healthcare Analytics [15, 22]. In the K-Means method, each data point was placed in a high-dimensional space where each dimension represents a data attribute. An iterative process was then performed to place the centroids (cluster center points) to minimize the distance between the data points and their respective centroids [23]. Data validation in this study was conducted using the Elbow Method and Cluster Profiling with the following explanation:

The Elbow method [4, 16, 23] was used to determine the optimal number of clusters by plotting the Sum of Squared Errors (SSE) against the number of clusters and identifying the "elbow point," where adding more clusters no longer significantly reduces the SSE. This study employs K-Means clustering as the algorithm, with the Elbow method used to optimize the clustering process. The Sum of Squared Errors (SSE) is calculated to evaluate clustering quality:

$$SSE = \sum_{i=1}^k \sum_{x \in S_i} (x - \mu_i)^2 \quad (2)$$

The interpretation of SSE is: i) lower SSE indicates that the data points are closely packed around their centroids, signifying well-defined clusters, and ii) a higher SSE suggests greater variability within clusters. The best number of clusters on K-Means was evaluated based on SSE from the elbow method. K-Means clustering was tested 4 times, using  $k=2$  until  $k=5$ .

Cluster profiling was performed to analyze the characteristics and distribution of variables in each cluster. In this study, clusters were profiled by calculating the mean, median, and range of attributes such as Worm Infection, Immunization, Smoking (Family), Maternal History and Maternal History. This process provides insight into

behavioral patterns and health characteristics in each cluster, guiding the development of tailored counseling materials

### 3. RESULT AND DISCUSSION

#### 3.1 Data before preprocessing

This study utilizes a dataset consisting of 231 records, each containing several key features. The attributes in the dataset include name, sex, date of birth, birth weight, birth height, parent's name, province, district/city, subdistrict, health center, village, address, age at measurement, date of measurement, weight, height, upper arm circumference, weight-for-age, height-for-age, weight-for-height, and weight gain. Additional critical attributes include access to clean water, the occurrence of worm infection, the availability of healthy toilets,

immunization status, and family smoking habits. The dataset also includes the medical history of pregnant women.

#### 3.2 Data after preprocessing

The raw data cannot be directly analyzed and must first undergo preprocessing to ensure accurate results. The preprocessing steps include several stages. Data Transformation involves converting categorical attributes into nominal data. For instance, binary values such as "Yes" and "No" are transformed into numerical values, 1 and 0, respectively. Attribute Selection follows, where irrelevant attributes are removed. In this study, the selected attributes include clean water, worm infection, healthy toilet, immunization, family smoking habits, and maternal history. Finally, data grouping organizes the data by subdistrict, as presented in the Table 2.

**Table 2.** Sampel data after preprocessing

Subdistrict	A	B	C	D	E	F	G	H	I	J	K	L
SUB1	0	22	20	1	0	22	1	21	4	16	11	1
SUB2	0	22	18	2	0	22	0	22	6	16	7	6
SUB3	0	25	24	0	0	25	0	25	7	17	20	0
SUB4	0	13	8	1	0	13	0	13	5	8	8	0
SUB5	0	6	5	1	0	6	0	6	1	5	3	1
SUB6	0	13	11	1	0	13	0	13	7	6	6	2
SUB7	0	5	4	0	0	5	0	5	2	2	2	0
SUB8	0	6	5	0	0	6	0	6	1	5	5	0
SUB9	0	7	7	0	0	7	0	7	1	6	4	1
SUB10	0	14	13	0	0	14	0	14	2	11	10	1

Notes: A: Clean Water (No), B: Clean Water (Yes), C: Worm Infection (No), D: Worm Infection (Yes), E: Healthy Toilet (No), F: Healthy Toilet (Yes), G: Immunization (No), H: Immunization (Yes), I: Family Smoking (No), J: Family Smoking (Yes), K: Maternal History (No), L: Maternal History (Yes)

The difference between the amount of data before and after preprocessing in this study is due to several main factors. The data preprocessing process involves stages such as data transformation, attribute selection, and data grouping to ensure data's consistency and relevance to the purpose of analysis. During preprocessing, incomplete data, irrelevant attributes, or data with extreme values (outliers) are removed to avoid skewing the results of the analysis. In addition, the transformation of data from categories to numerical values, such as converting binary data like "Yes" and "No" to 1 and 0, also affects the data structure. The attribute selection process further eliminates variables irrelevant to the research focus, such as demographic details that do not contribute directly to the results of the analysis. As a result, the final dataset used for analysis is smaller than the initial sample. This reduction is necessary to increase the validity and accuracy of the generated results

The major health issues in the District X are related to access to clean water, the prevalence of infectious diseases, and the overall effectiveness of immunization efforts. According to the data in Table 2 Sampel Data after Preprocessing, clean water and healthy toilets are present in 100% of cases. While 99.57% of toddlers have been immunized, only 0.43% have not. Considering the data distribution with a good tendency, the features selected for clustering include worm infection, immunization, family smoking habits, and maternal history.

#### 3.3 Using the K-Means algorithm

From the processed data in Table 2, clustering was performed using the K-Means technique. The results are

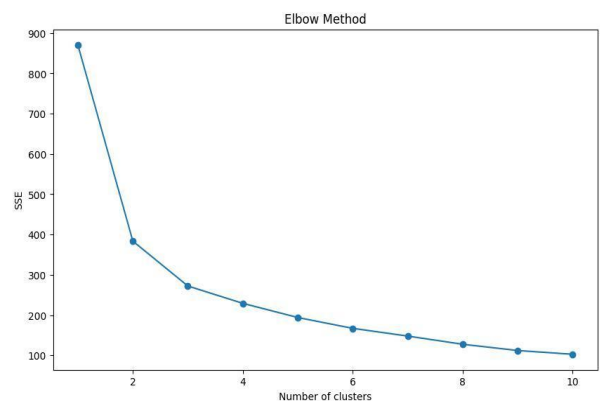
summarized in Table 3.

**Table 3.** Clustering results using K-Means

K	2	3	4	5
Cluster 1	18 Data	16 Data	9 Data	8 Data
Cluster 2	7 Data	5 Data	7 Data	8 Data
Cluster 3		4 Data	5 Data	5 Data
Cluster 4			4 Data	3 Data
Cluster 5				1 Data

#### 3.4 SSE result each cluster

The Elbow method was applied to determine the optimal number of clusters. The SSE scores for clusters with k values ranging from 2 to 5 are presented in Table 3.



**Figure 2.** Line graph SSE score

The data presented in Figure 2 and Table 4 indicate that the highest SSE value is observed at  $k = 2$ . Therefore, it can be concluded that the optimum number of clusters to be implemented is 3. The result of Clustering using the Elbow Method and K-Means algorithm

**Table 4.** SSE score each cluster

k	SSE Score	Difference
2	996.951	0
3	552.613	444.338
4	362.867	189.746
5	250.508	112.359

From Table 5, Cluster 1 displays the result of clustering data in Cluster 1, which consists of 16 subdistricts. This cluster is characterized by the presence of Worm Infection (No), Worm Immunization (Yes), Smoking (Family) (Yes), and Maternal History (No), with a low number of case findings.

From Table 6, Cluster 2 presents the clustering result of data in Cluster 2 for five subdistricts. The cluster is classified based on the presence of Worm Infection No, Immunization (Yes), Smoking (Family) Yes, and Maternal History (No) with a moderate number of case findings.

**Table 5.** Cluster 1

Subdistrict	C	D	G	H	I	J	K	L
SUB5	5	1	0	6	1	5	3	1
SUB7	4	0	0	5	2	2	2	0
SUB8	5	0	0	6	1	5	5	0
SUB9	7	0	0	7	1	6	4	1
SUB11	5	1	0	6	0	6	4	0
SUB12	9	0	0	9	0	8	1	4
SUB13	1	1	0	3	0	3	2	1
SUB14	3	0	0	3	0	3	1	0
SUB15	1	0	0	1	1	0	1	0
SUB16	2	0	0	2	1	1	2	0
SUB17	2	0	0	2	0	1	1	0
SUB18	2	0	0	2	0	2	1	0
SUB21	3	2	0	5	1	4	3	0
SUB23	6	1	0	8	2	5	3	0
SUB24	6	0	0	6	3	3	3	1
SUB25	2	0	0	2	1	1	2	0

Notes: C= Worm Infection (No), D = Worm Infection (Yes), G= Immunization (No), H= Immunization (Yes), I = Smoking (Family) (No), J= Smoking (Family) (Yes), K= Maternal History (No), L = Maternal History (Yes)

**Table 6.** Cluster 2

Subdistrict	C	D	G	H	I	J	K	L
SUB4	8	1	0	13	5	8	8	0
SUB6	11	1	0	13	7	6	6	2
SUB10	13	0	0	13	2	11	10	1
SUB19	15	0	0	14	5	11	7	3
SUB20	13	0	0	16	3	10	8	3

Notes: C = Worm Infection (No), D = Worm Infection (Yes), G = Immunization (No), H = Immunization (Yes), I = Smoking (Family) (No), J = Smoking (Family) (Yes), K = Maternal History (No), L = Maternal History (Yes).

From Table 7, Cluster 3 presents the clustering outcome in Cluster 3 for four subdistricts. This cluster is characterized by the presence of Worm Infection (No), Worm Immunization (Yes), Smoking (Family) (Yes), and Maternal History (No) with a high number of case findings. Cluster-Based Strategies for Stunting Prevention

**Table 7.** Cluster 3

Subdistrict	C	D	G	H	I	J	K	L
SUB1	20	1	1	21	4	16	11	1
SUB2	18	2	0	22	6	16	7	6
SUB3	24	0	0	25	7	17	20	0
SUB22	19	0	0	19	5	14	14	3

Notes: C = Worm Infection (No), D = Worm Infection (Yes), G = Immunization (No), H = Immunization (Yes), I = Smoking (Family) (No), J = Smoking (Family) (Yes), K = Maternal History (No), L = Maternal History (Yes).

Cluster-based strategies for stunting prevention address the unique characteristics and needs of each identified risk group. Using K-Means Clustering validated by the Elbow Method, three distinct clusters were identified based on public health risk levels. Significant variations among the clusters were observed, including differences in worm infection rates, immunization coverage, family smoking habits, and maternal health history. These variations highlight the diverse health conditions across sub-districts, emphasizing the necessity for tailored intervention strategies to ensure effective implementation.

This study uniquely applies K-Means clustering using local health indicators, such as maternal health, clean water access, and community-specific stunting prevention strategies. By incorporating these tailored indicators, the proposed clustering framework provides actionable insights for targeted interventions, enhancing the effectiveness of stunting prevention efforts in District X.

Cluster 1 (Low-Risk Areas) includes 16 sub-districts characterized by relatively good health conditions. The worm infection rate is low, with an average of 4-9 cases per sub-district. Immunization coverage is very high, nearly reaching 100%, with 8-9 babies in each sub-district receiving immunization. However, smoking habits in the family environment are present in 25%-35% of households (2-5 families per sub-district). Maternal health history shows very few cases, typically 0-1 cases per sub-district. While this cluster demonstrates stable health conditions, attention is still needed to address family smoking habits.

Unlike previous studies that primarily relied on anthropometric measures for clustering [18], this research incorporates behavioral and environmental factors, such as family smoking habits, maternal health history, and access to clean water. These additional parameters provide a more comprehensive understanding of stunting risks, thus allowing the development of more context-specific interventions. Furthermore, the inclusion of these indicators highlights gaps in previous methodologies, which often overlooked the influence of environmental and behavioral factors, which are critical for public health strategies.

Cluster 2 consists of five sub-districts with a moderate risk level. Worm infections remain relatively high, with 11-15 cases per sub-district. Immunization coverage in this region is in the range of 70%-85%, covering 9-13 babies per sub-district. Family smoking habits are more prevalent than in Cluster 1, reaching 40%-50% of households (6-11 families per sub-district). In addition, maternal health history showed a moderate prevalence, namely 2-3 cases per sub-district. This region requires increased immunization coverage and targeted interventions to reduce smoking habits in households.

Cluster 3 includes four sub-districts with health conditions that require intensive attention. Worm infections are very high, reaching 18-24 cases per sub-district. Immunization coverage

is relatively high at 90%-100%, with 19-25 babies per sub-district having received immunization. However, smoking habits in families are alarmingly prevalent, affecting 60%-80% of households (14-20 families per sub-district). Maternal health history also shows high numbers, with 3-6 cases per sub-district. In addition, approximately 50% of the areas in this cluster have inadequate sanitation, requiring comprehensive interventions such as sanitation improvements and mass deworming programs.

### 3.5 Cluster based strategies for stunting prevention

Based on the clustering results, stunting prevention strategies are tailored to the three to the specific characteristics and needs of each cluster. Cluster 1, which includes 16 low-risk sub-districts, is characterized by a low incidence of worm infections, high immunization coverage, and moderate smoking prevalence. The main strategies for this cluster include preventive education about worm infections, immunization, and the dangers of smoking in households, as well as the implementation of routine monitoring programs. Preventive education is a cost-effective approach that supports positive health behaviors, while routine monitoring programs align with WHO recommendations for early detection of health issues [3, 24]

For Cluster 2, which consists of five sub-districts with medium risk, the proposed strategy involves the intensification of worm control programs and the improvement of immunization coverage. Additionally, targeted campaigns about the dangers of smoking and the importance of maternal health are essential, considering the significant prevalence of smoking in this cluster. A more intensive worm control program can significantly improve children's health and

development, while the anti-smoking campaign focuses on reducing exposure to cigarette smoke, a known risk factor for stunting [24].

In Cluster 3, which consists of four high-risk sub-districts, comprehensive interventions are needed, including sanitation improvements, mass deworming programs, and increased access to healthcare services for mothers and children. Cross-sector collaboration is crucial to holistically address the various risk factors contributing to stunting. Research indicates that improved sanitation effectively reduces stunting [3] while mass deworming programs shows promising results when combined with other interventions [8].

The implementation of this program relies on a community-based approach by engaging community leaders and local health cadres, integrating various health services, and continuous monitoring through a health information system [3]. Training healthcare workers is vital to enhance their capacity in addressing stunting [25]. Periodic evaluations conducted every six months will measure the program's effectiveness and allow for necessary strategic adjustments [4]. This cluster-based approach is designed to optimize resource allocation and improve the effectiveness of stunting prevention programs in the District X. By implementing targeted and locally responsive interventions, this strategy has the potential to significantly reduce the prevalence of stunting in the region.

Table 8 groups subdistricts according to risk levels, the number of sub-districts per cluster, and recommended intervention strategies. Each cluster has different intervention priorities, ranging from preventive education to sanitation improvements, intensification of immunization programs, control of worm infections, and increased access to health services for mothers and children.

**Table 8.** Intervention strategy

Cluster	Region	Number of Districts	Intervention Strategy
Cluster 1	Low-risk subdistrict	16	Preventive education about worm infections, immunization, the dangers of smoking; routine monitoring
Cluster 2	Districts with medium risk	5	Intensification of worm control programs, improvement of immunization, anti-smoking campaigns
Cluster 3	High-risk subdistrict	4	Sanitation improvements, mass deworming, increased access to healthcare services for mothers and children

Cluster 1: Sub-districts in the low-risk category have a relatively low prevalence of stunting (around 15%) with high immunization coverage (90%). However, the prevalence of smoking remains significant (25%), necessitating further education on the dangers of smoking within households. Cluster 2: The medium-risk cluster demonstrates a stunting prevalence of 25% and a worm infection prevalence of 35%, highlighting the need for intensified worm control programs. Immunization coverage in this cluster remains less optimal at 70%, indicating the need for improvement. Cluster 3: High-risk sub-districts show a stunting prevalence of 40%, and inadequate sanitation coverage in 50% of the area. Additionally, the coverage of maternal and child health services is critically low, around 30%, making comprehensive intervention very necessary.

While K-Means provides a useful segmentation of sub-districts based on stunting risk, it has several limitations that should be considered. The method assumes that clusters are spherical and evenly distributed, which may not accurately reflect the complex and overlapping nature of stunting risk factors. The algorithm is also sensitive to outliers, meaning

that extreme values—such as a sub-district with exceptionally high or low stunting prevalence—could disproportionately influence cluster assignments. Moreover, K-Means treats all features equally in its distance calculations, potentially overlooking the varying importance of factors like sanitation, immunization, and healthcare access. Addressing these limitations may require complementary methods, such as hierarchical clustering or dimensionality reduction techniques, to refine the analysis and ensure more meaningful groupings for targeted interventions.

Following the categorization of regions into risk-based clusters, the next step involves preparing competent speakers and scheduling the delivery of educational materials effectively. To optimize the impact of stunting prevention programs, several key factors need to be considered. Effective speaker selection criteria include a minimum background in public health or nutrition education, fluency in local languages, and completion of effective communication training. Speakers should receive specialized training in technical content and communication strategies, and demonstrate the ability to answer technical questions in simple language while having



experience working with similar communities. Educational sessions should ideally be conducted weekly for 60-90 minutes over a three-month period, followed by monthly reinforcement sessions, with material divided into small modules (20-30 minutes) accompanied by interactive breaks for optimal information absorption. Delivery methods must be tailored to the characteristics of target groups, such as direct cooking demonstrations and role-playing for pregnant women and toddlers, social media and interactive mobile applications for adolescents, and colored visual aids for groups with low literacy. Effective follow-up strategies include monthly home visits by health workers, formation of WhatsApp support groups, provision of nutritional monitoring diaries, incentive programs such as healthy family certificates, and periodic evaluations through child growth measurements and parental knowledge tests to measure program success.

This process is crucial to ensure that the information is well received by the public. With proper scheduling, regular counseling sessions can be conducted, thus allowing the community to understand and apply the knowledge provided to improve their overall health and well-being.

The results of this clustering align with previous findings [26], which also identified three main groups of stunting risk factors. However, this research provides a localized perspective to the District X. The use of K-Means clustering proves effective in identifying public health patterns, even with a relatively smaller dataset. As explained in the findings [26] this method is effective in classifying provinces in Indonesia based on stunting prevalence, breastfeeding practices, and the adequacy of energy and protein intake. Despite certain limitations, such as the small dataset size, this study makes a valuable contribution to evidence-based stunting prevention strategies. By combining a data-driven approach with targeted public health strategies, this research provides a solid foundation for the development of more relevant and effective health interventions at the community level [27].

#### 4. CONCLUSION

This study successfully applied the K-Means method, validated using the Elbow method and Cluster Profiling to classify the health conditions of toddlers as part of a stunting prevention effort. The Elbow Method identified three optimal clusters that reflect variations in health conditions with differing risk levels. The first cluster represents low-risk areas, characterized by high immunization coverage and low worm infection rates. Recommended strategies for this cluster include preventive education and routine monitoring of family health. The second cluster consists of medium-risk areas, where interventions focus more on intensifying worm infection control programs, increasing immunization coverage, and smoking hazard campaigns in families. The third cluster, identified as high risk, requires a comprehensive approach, including improved sanitation, mass deworming programs, and enhanced access to maternal and child health services. These clusters have been validated through the Cluster Profiling.

The clustering approach significantly enhances the formulation of targeted stunting prevention strategies. By aligning interventions to the specific needs of each region, this framework establishes a strong foundation for effective public health policies. Furthermore, the data-driven methodology

facilitates the creation of precise and impactful health counseling materials, optimizing community health outcomes and resource allocation.

The results of this classification also provide a basis for designing more specific and targeted health interventions. By understanding the distribution patterns of health conditions, policymakers can develop more effective stunting prevention programs, such as increasing access to clean water in areas with a high incidence of worm infections or focusing immunization on areas with low coverage. This approach helps allocate resources more efficiently and improve the effectiveness of health interventions.

In addition, the distribution of health conditions supports the identification of regional groups and their management strategies. Interventions are better aligned with local conditions, enabling more effective stunting prevention through targeted counseling. The Elbow method has proven effective in determining the optimal number of clusters in the K-Means analysis, demonstrating the importance of data-driven strategies and consideration of local contexts in more effective stunting efforts.

This research is expected to inspire further studies to integrate additional health parameters and expand the dataset to improve model accuracy. With periodic data updates, the classification of intervention groups can be updated dynamically, allowing for more precise adjustments to intervention strategies and serving as a valuable reference for more effective interventions.

#### REFERENCES

- [1] Silva, J.M., Vieira, L.L., Abreu, A.M., de Souza Fernandes, E., Moreira, T.R., da Costa, G.D., Cotta, R.M. (2023). Water, sanitation, and hygiene vulnerability in child stunting in developing countries: A systematic review with meta-analysis. *Public Health*, 219: 117-123. <https://doi.org/10.1016/j.puhe.2023.03.024>
- [2] De Onis, M., Branca, F. (2016). Childhood stunting: A global perspective. *Maternal & Child Nutrition*, 12: 12-26. <https://doi.org/10.1111/mcn.12231>
- [3] Danaei, G., Andrews, K.G., Sudfeld, C.R., Fink, G., et al. (2016). Risk factors for childhood stunting in 137 developing countries: A comparative risk assessment analysis at global, regional, and country levels. *PLoS Medicine*, 13(11): e1002164. <https://doi.org/10.1371/journal.pmed.1002164>
- [4] Black, R.E., Victora, C.G., Walker, S.P., Bhutta, Z.A., et al. (2013). Maternal and child undernutrition and overweight in low-income and middle-income countries. *The Lancet*, 382(9890): 427-451. [https://doi.org/10.1016/S0140-6736\(13\)60937-X](https://doi.org/10.1016/S0140-6736(13)60937-X)
- [5] Ruel, M.T., Alderman, H. (2013). Nutrition-sensitive interventions and programmes: How can they help to accelerate progress in improving maternal and child nutrition? *The Lancet*, 382(9891): 536-551. [https://doi.org/10.1016/S0140-6736\(13\)60843-0](https://doi.org/10.1016/S0140-6736(13)60843-0)
- [6] Hoddinott, J., Alderman, H., Behrman, J.R., Haddad, L., et al. (2013). The economic rationale for investing in stunting reduction. *Maternal & Child Nutrition*, 9: 69-82. <https://doi.org/10.1111/mcn.12080>
- [7] Adedokun, S.T., Yaya, S. (2021). Factors associated with adverse nutritional status of children in sub-Saharan Africa: Evidence from the demographic and health

- surveys from 31 countries. *Maternal & Child Nutrition*, 17(3): e13198. <https://doi.org/10.1111/mcn.13198>
- [8] Prendergast, A.J., Humphrey, J.H. (2014). The stunting syndrome in developing countries. *Paediatrics and International Child Health*, 34(4): 250-265. <https://doi.org/10.1179/2046905514Y.0000000158>
- [9] Kusumawardani, L.H., Rachmawati, U., Jauhar, M., Rohana, I.G.A.P.D. (2020). Community-based stunting intervention strategies: Literature review. *Dunia Keperawatan: Jurnal Keperawatan dan Kesehatan*, 8(2): 259-268. <https://doi.org/10.20527/dk.v8i2.8555>
- [10] Martony, O. (2023). Stunting di Indonesia: Tantangan dan solusi di era modern. *Journal of Telenursing*, 5(2): 1734-1745. <https://doi.org/10.31539/joting.v5i2.6930>
- [11] Haque, M.A., Choudhury, N., Wahid, B.Z., Ahmed, S.T., et al. (2023). A predictive modelling approach to illustrate factors correlating with stunting among children aged 12–23 months: A cluster randomised pre-post study. *BMJ Open*, 13(4): e067961. <https://doi.org/10.1136/bmjopen-2022-067961>
- [12] Abdullah, A., Sucipto, S. (2023). Liver disease classification using the elbow method to determine optimal K in the k-nearest neighbor (K-NN) algorithm. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, 12(2): 218-228. <https://doi.org/10.32736/sisfokom.v12i2.1643>
- [13] Chowdhury, J., Ravi, R.P. (2022). Healthcare accessibility in developing countries: A global healthcare challenge. *Journal of Clinical & Biomedical Research*, pp. 1-5. [https://doi.org/10.47363/JCBR/2022\(4\)152](https://doi.org/10.47363/JCBR/2022(4)152)
- [14] Aguayo, V.M., Menon, P. (2016). Stop stunting: Improving child feeding, women's nutrition and household sanitation in South Asia. *Maternal & Child Nutrition*, 12: 3-11. <https://doi.org/10.1111/mcn.12283>
- [15] Waller, A., Lakhanpaul, M., Godfrey, S., Parikh, P. (2020). Multiple and complex links between babyWASH and stunting: An evidence synthesis. *Journal of Water, Sanitation and Hygiene for Development*, 10(4): 786-805. <https://doi.org/10.2166/washdev.2020.265>
- [16] Rykov, A., De Amorim, R.C., Makarenkov, V., Mirkin, B. (2024). Inertia-based indices to determine the number of clusters in K-Means: An experimental evaluation. *IEEE Access*, 12: 11761-11773. <https://doi.org/10.1109/ACCESS.2024.3350791>
- [17] Sutomo, F., Maaafii, D.A., Al Rasyid, D.N., Kurniawan, Y.I., et al. (2023). Optimization of the k-nearest neighbors algorithm using the elbow method on stroke prediction. *Jurnal Teknik Informatika*, 4(1): 125-130. <https://doi.org/10.52436/1.jutif.2023.4.1.839>
- [18] Pérez-Rodrigo, C., Gil, Á., González-Gross, M., Ortega, R.M., et al. (2015). Clustering of dietary patterns, lifestyles, and overweight among Spanish children and adolescents in the ANIBES study. *Nutrients*, 8(1): 11. <https://doi.org/10.3390/nu8010011>
- [19] Nagari, S.S., Inayati, L. (2020). Implementation of clustering using k-means method to determine nutritional status. *Jurnal Biometrika dan Kependudukan*, 9(1): 62-68. <https://doi.org/10.20473/jbk.v9i1.2020.62-68>
- [20] Umargono, E., Suseno, J.E., Gunawan, S. (2019). K-means clustering optimization using the elbow method and early centroid determination based-on mean and median. In *Proceedings of the International Conferences on Information System and Technology*, Yogyakarta, Indonesia, pp. 234-240. <https://doi.org/10.5220/0009908402340240>
- [21] Maori, N.A., Evanita, E. (2023). Metode elbow dalam optimasi jumlah cluster pada k-means clustering. *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, 14(2): 277-288. <https://doi.org/10.24176/simet.v14i2.9630>
- [22] Ashabi, A., Sahibuddin, S.B., Salkhordeh Haghghi, M. (2020). The systematic review of K-means clustering algorithm. In *Proceedings of the 2020 9th International Conference on Networks, Communication and Computing*, Tokyo, Japan, pp. 13-18. <https://doi.org/10.1145/3447654.3447657>
- [23] Onumanyi, A.J., Molokomme, D.N., Isaac, S.J., Abu-Mahfouz, A.M. (2022). AutoElbow: An automatic elbow detection method for estimating the number of clusters in a dataset. *Applied Sciences*, 12(15): 7515. <https://doi.org/10.3390/app12157515>
- [24] Bhutta, Z.A., Das, J.K., Rizvi, A., Gaffey, M.F., et al. (2013). Evidence-based interventions for improvement of maternal and child nutrition: What can be done and at what cost? *The Lancet*, 382(9890): 452-477. [https://doi.org/10.1016/S0140-6736\(13\)60996-4](https://doi.org/10.1016/S0140-6736(13)60996-4)
- [25] Simons, K., Bradfield, O., Spittal, M.J., King, T. (2023). Age and gender patterns in health service utilisation: Age-Period-Cohort modelling of linked health service usage records. *BMC Health Services Research*, 23(1): 480. <https://doi.org/10.1186/s12913-023-09456-x>
- [26] Yusuf, A. (2022). K-means clustering based on distance measures: Stunting prevalence clustering in south kalimantan. In *2022 5th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, Yogyakarta, Indonesia, pp. 706-710. <https://doi.org/10.1109/ISRITI56927.2022.10052925>
- [27] KM, N.I.A.M.S., Purnomo, W., Mahmudah, I. (2019). Implementation of the k-means clustering method on stunting case in Indonesia. *International Journal of Advances in Scientific Research and Engineering*, 5(6): 103-107. <https://doi.org/10.31695/IJASRE.2019.33258>