TIMURIS
Journal of Computational and Information Research

Vol. 01, No. 1, May, 2026
Pages 27-38

Check for updates

Application of Principal Component Analysis (PCA) to Identify the Main Factors Causing Stunting

Novica Sintasyah Sinaga[1*], Elisabeth Princess[2], Aisyah Novianti[3]

[123]Mathematics Department, Faculty of Mathematic and Natural Sciences, State University of Medan, Jl. Wiliam Iskandar PS. V Kenangan Baru, North Sumatra, 20221, Indonesia

Email correponding Author: **novica.4243260036@mhs.unimed.ac.id*

Article Info

Article history:

Received: April 02, 2026
Revised: May 13, 2026
Accepted: June 20, 2026

Keywords:

Causal Factors,
Dimension Reduction,
North Sumatra,
Principal Component Analysis (PCA),
Stunting.

ABSTRACT

Stunting is a serious public health problem in Indonesia, including in North Sumatra Province where the prevalence is still above the threshold set by WHO. This study aims to identify the main factors that cause stunting in North Sumatra Province using the Principal Component Analysis (PCA) method. PCA is applied to reduce a number of variables that cause stunting into several main components that are able to explain the diversity of data to the maximum and overcome the problem of multicollinearity between variables. The data used is secondary data from stunting vulnerability indicators in districts/cities throughout North Sumatra Province. The results of the analysis showed that PCA succeeded in reducing the data dimension and identifying the main components with the highest eigenvalues that were the dominant factors causing stunting. These findings are expected to provide a comprehensive overview of the factors that have the most influence on stunting incidence in North Sumatra, so that it can be the basis for more targeted and effective intervention policy recommendations for local governments in an effort to accelerate the reduction of stunting rates.

: https://doi.org/10.30598/timurisv1i1pp27-38

1. INTRODUCTION

Nutritional status in toddlers is still a major problem characterized by an increasing incidence of malnutrition. Based on the World Health Organization (WHO) in 2021, the prevalence of deaths in children under five is caused by malnutrition, where the prevalence of malnutrition worldwide reaches 28.5%, in developing countries it is 32.2%, the Asian continent is 30.6%, and Southeast Asia is 29.4% (WHO, 2021). Indonesia is a developing country with an increasing incidence of undernourished toddlers. Based on the Indonesian Health Profile, there was a significant increase from 2019 to 2021, from 16.3% to 17.0% (Sari & Susilowati, 2023).

One of the factors that causes malnutrition problems in toddlers is the lack of food intake caused by the inadequate intake of balanced nutritious food and the wrong diet. Lack of care for toddlers and limited health services can cause various kinds of infectious diseases, as well as limited information about the growth and development of children's nutrition (Ula et al., 2021). Malnutrition is a condition of the body that appears very thin because the food consumed every day cannot meet the nutrients needed, especially energy and protein. The problems faced by the community are the lack of knowledge or information about malnutrition in toddlers, including the lack of availability of specialist doctors which results in difficulties in people seeking treatment and not knowing how to prevent the early symptoms of malnutrition (Alam & Nurcahyo, 2022).

Stunting is a chronic nutritional problem characterized by growth disorders in children due to lack of nutritional intake in the long term. This condition has an impact on children's physical and cognitive development and has the potential to reduce the quality of human resources in the future. The high prevalence of stunting in various regions shows that this problem still requires serious attention and handling based on accurate and comprehensive data analysis (Ramadhani et al., 2026). Stunting not only affects height growth, but is also closely related to poor nutritional conditions, low birth weight, low immunization coverage, and inadequate access to maternal health services, so a multivariate approach is needed to understand the relationship between these factors comprehensively.

Immunization is one of the efforts to cause or increase a person's immunity to a disease, so that if one day exposed to the disease, the person will not get sick or only experience mild illness (Ulfah & Fitra, 2025). With immunization, it is hoped that it can reduce the number of illnesses, deaths, and disabilities due to Diseases Preventable by Immunization (PD3I), such as Hepatitis B, tuberculosis (Tuberculosis), Diphtheria, Pertussis, Tetanus, Polio, Measles, Rubella, and pneumonia (Ministry of Health of the Republic of Indonesia, 2020). Complete basic immunization coverage in Indonesia has fluctuated, where in 2019 it was recorded at 93.7% but decreased in 2020 to 83.3%. Low immunization coverage is caused by various factors, both in terms of information, motivation, and situation, such as mothers' lack of knowledge about immunization schedules, fear of side effects, long distances to health facilities, and sick children at the time of immunization schedules. This data shows that knowledge plays an important role in the success of immunization in infants (Darmin et al., 2023).

Based on the 2010 Population Census, the infant mortality rate (AKB) in North Sumatra was recorded at 25 per 1,000 live births. Although AKB shows a decreasing trend compared to the previous period, various factors still affect this condition, including improving health services, pregnancy maintenance and check-ups, childbirth assistants, breastfeeding and supplementary food, and immunization coverage (Rambe et al., 2022). This condition emphasizes the need for special attention to various maternal and child health indicators in North Sumatra Province as one of the provinces with a considerable nutritional and health burden.

Low Birth Weight (BBLR), which is the condition of babies with a birth weight of less than 2,500 grams, is a serious global health problem with long-term implications. Factors that contribute to BBLR include nutrition during pregnancy, access to healthcare, infections during pregnancy, and genetic factors. BBLR carries the risk of death, impaired child growth and development, as well as the risk of being short (stunting) if not handled properly (Gemilastari et al., 2024). According to WHO data, the prevalence of BBLR is estimated to be around 15–20% of all births in the world, and almost 95% of cases occur in low- to middle-income countries. Based on the results of the 2022 Indonesian Nutrition Status Survey (SSGI), the prevalence of BBLR in Indonesia was recorded at 6.0%. In addition, BBLR has a long-term impact until adulthood, including increasing the risk of coronary heart disease, diabetes, metabolic and immune disorders, and low physical endurance (Hazimah et al., 2024). Disruptions during the brain-forming process have long-term consequences on the structural capacity and function of the brain, contributing to the development of a toddler's cognitive, language, and sensory motor capacities.

K4 coverage is the number of pregnant women who have received antenatal care services in accordance with the standard at least 4 times according to the recommended schedule, compared to the target number of pregnant women in one work area within one year. The indicator shows access to health services for pregnant women and the level of compliance of pregnant women in checking their pregnancy with health workers. Data from the North Sumatra Provincial Health Office in 2013, maternal AKI was 268 per 100,000 live births. According to the results of the 2010 Population Census, the AKI in North Sumatra was 328 per 100,000 live births, this figure is still quite high when compared to the national figure of 259 per 100,000 births (Harahap, 2021).

The complexity of stunting problems involving various public health indicators of the poor, basic immunization coverage, BBLR, malnutrition of toddlers, and maternity K4 coverage requires a statistical analysis approach that is able to summarize all of these dimensions efficiently. The five indicator variables are generally correlated with each other, so analyses that consider only one variable at a time risk ignoring the latent relationships that underlie the overall data pattern. Therefore, a dimension reduction method is needed that is able to extract important information from a set of interrelated variables into a small number of components that are easier to interpret.

Principal Component Analysis (PCA) is a multivariate analysis method that aims to reduce the dimension of data by transforming a set of interrelated origin variables into a small number of new, uncorrelated (orthogonal) variables, called principal components, while retaining most of the data diversity information. PCA has been widely used in public health research to identify latent factors underlying variations in health indicator data and to rank regions based on composite indices formed. The application of PCA to stunting indicator data in 33 districts/cities of North Sumatra Province is expected to produce a single composite index that represents the dimensions of maternal-child health burden and socio-economic related to stunting, as well as identify areas that need priority health interventions (Lesnussa et al., 2025).

Based on the background that has been presented, this study aims to: (1) describe the statistical characteristics of five public health indicator variables related to stunting in 33 districts/cities in North Sumatra Province; (2) test the feasibility of data for analysis using PCA through the Kaiser-Meyer-Olkin Test (KMO) and Bartlett's Test of Sphericity; (3) identify the structure of the correlation relationship between health indicator variables as the basis for justification for reducing the data dimension; (4) determine the key components that optimally summarize the diversity of the data based on the Kaiser Criterion (eigenvalue > 1) and the proportion of cumulative variance, confirmed through the Scree Plot; (5) interpret the first major component (PC1) as a single composite index representing the dimensions of maternal-child health burden and socio-economic burden related to stunting; and (6) identify districts/cities in North Sumatra Province that require priority health interventions based on the main component scores (PC1 scores) obtained from the results of the analysis.

2. METHOD

This study uses a descriptive quantitative design that is strengthened with a multivariate analysis approach through the application of Principal Component Analysis (PCA). The data studied is a cross-section secondary data collected from official publications of the North

Sumatra Provincial Health Office and the Central Statistics Agency (BPS) of North Sumatra Province. The observation units in this study cover all districts and cities in North Sumatra Province, which totals 33 regions. There are five public health indicators that are the focus of the study variables, all of which are closely related to stunting conditions, namely the percentage of poor population (X_1), basic immunization coverage (X_2), prevalence of low birth weight or BBLR (X_3), malnutrition rate in toddlers (X_4), and maternity K4 antenatal service coverage (X_5). The five variables are continuous with a ratio scale and are all expressed in percent units.

As a first step, data exploration was carried out through descriptive statistics that included the minimum, first quartile, median, average, third quartile, and maximum values of each variable. The goal of this stage is to understand the distribution and basic characteristics of the data before entering more complex analysis procedures. After that, the feasibility of the data to be analyzed with PCA was tested using two statistical approaches, namely the Kaiser-Meyer-Olkin test (KMO) and the Bartlett's Test of Sphericity. The KMO test functions to measure the adequacy of the sample by determining that the data is considered feasible if the KMO value reaches 0.5 or more. Meanwhile, Bartlett's Test of Sphericity aims to ensure a significant correlation between indicators; The eligibility criteria are met if the resulting P-value is below 0.05.

Before the extraction process of the main components is carried out, all data is first standardized through Z-score transformation. This procedure guarantees that each variable exerts a balanced influence on the formation of the main component, with no variable dominating simply due to differences in measurement scales. Furthermore, a correlation matrix between variables was built and analyzed to identify the pattern of relationships between indicators as well as detect information redundancy which is an indication that dimension reduction through PCA is the right step. The core stage in PCA is the calculation of the eigenvalue and eigenvector of the correlation matrix. The eigenvalue reflects the amount of data diversity that each main component is able to explain, while the eigenvector contains a loading coefficient that shows the weight of each variable's contribution to the main component in question. From the results of these calculations, the main components are formed as a linear combination of all standardized variables.

The number of key components to be retained in the final interpretation is determined by reference to two criteria at once. Kaiser's criteria stipulate that components with an eigenvalue greater than 1 are considered worthy of retention. This decision is then visually confirmed through the Scree Plot, by paying attention to the breaking point (elbow) on the graph where the decline in eigenvalue begins to slope. The selected main components are then interpreted based on the absolute loading value of each variable; Variables with a load ≥ 0.5 are considered to have a dominant role in forming the main component. In addition, the PCA (principal component scores) score is calculated for each district/city as a composite index that allows comparison between regions and ranking based on the level of vulnerability of health indicators related to stunting.

To make the results of the analysis easier to understand, this study presents various forms of supporting visualization, including correlation matrices, Scree Plot, Individuals PCA Plot, Loading Plot, and PCA Biplot. The entire computation and data visualization process is done using R Studio software, which is supported by packages such as stats, FactoMineR, factoextra, and corrplot.

3. RESULT AND DISCUSSION

3.1 Data Types and Sources

This study is a descriptive quantitative research with a multivariate analysis approach using the Principal Component Analysis (PCA) method. The data used is cross-section secondary data sourced from official publications of the North Sumatra Provincial Health Office and the Central Statistics Agency (BPS) of North Sumatra Province. The data includes five public health indicator variables related to stunting in 33 districts/cities in North Sumatra Province. All variables used in this study, along with their codes and units, are presented in Table 1.

Table 1. Research Variables and Their Codes

Code	Variabel	Units
X_1	Poor Population	%
X_2	Basic Immunization Coverage	%
X_3	Low Birth Weight (BBLR)	%
X_4	Poor nutrition	%
X_5	Maternity K4 Coverage	%

3.2 Description of Research Variable Statistics

This study uses health data from 33 districts/cities in North Sumatra Province with five stunting indicator variables, namely: Poor Population (%), Basic Immunization Coverage (%), BBLR/Low Birth Weight (%), Malnutrition of Toddlers (%), and K4 Maternity Coverage (%). The data is in the form of a tibble with dimensions of 33 rows and 6 columns, where the first column is the name of the district/city and the next five columns are numerical variables. Based on the results of descriptive statistics obtained from the output of R Studio, the characteristics of each variable can be explained as follows:

Table 2. Descriptive Statistics of Stunting Variables in North Sumatra Province

Variabel	Minimum	Q1	Median	Mean	Q3	Maximum
Poor Population (%)	2,80%	5,89%	9,15%	8,89%	11,11%	15,11%
Basic Immunization Coverage (%)	45,16%	49,77%	52,74%	52,30%	55,08%	58,02%
BBLR (%)	7,42%	8,51%	8,78%	8,62%	8,96%	9,41%
Infant Malnutrition (%)	3,36%	3,74%	3,91%	3,89%	4,12%	4,33%
K4 Maternity Coverage (%)	59,65%	65,34%	69,97%	69,04%	72,31%	76,21%

The Poor Population variable has a fairly wide range of values, from 2.80% to 15.11%, with an average of 8.89%. This indicates that there is a significant poverty gap between districts/cities in North Sumatra. The Basic Immunization Coverage variable is in the range of 45.16% to 58.02%, with an average of 52.30%, which shows that basic immunization coverage in most regions is still below the national target of 80%. The BBLR variable has a relatively narrow variation (7.42% - 9.41%), indicating a fairly even distribution. The Malnutrition Variable for Toddlers ranged from 3.36% to 4.33%, while the Maternity K4 Coverage ranged from 59.65% to 76.21% with an average of 69.04%.

3.3 Uji Kelayakan Principal Component Analysis (PCA)

Before conducting the PCA analysis, a feasibility test was first carried out using two methods, namely the Kaiser-Meyer-Olkin Test (KMO) and the Bartlett's Test of Sphericity. Both tests aim to ensure that the data used is eligible for analysis using PCA.

3.4. Kaiser-Meyer-Olkin Test (KMO)

The results of the KMO test showed an MSA (Measure of Sampling Adequacy) value of 0.8712. SME statistics are formulated as:

$$KMO = \Sigma_{i} \neq {}_{j} r^2{}_{ij} / (\Sigma_{i} \neq {}_{j} r^2{}_{ij} + \Sigma_{i} \neq {}_{j} a^2{}_{ij})$$

where *RIM* is the correlation coefficient between the I and Jth variables, and *AIJ* is the partial correlation coefficient. The data is declared eligible for PCA if the SME value is ≥ 0.5. According to Kaiser (1974), the value of SMEs is categorized as: 0.5–0.6 (adequate/marginal), 0.6–0.7 (moderately good), 0.7–0.8 (good), 0.8–0.9 (excellent/meritorious), and ≥ 0.9 (excellent/marvelous).

3.5 Uji Bartlett's Test of Sphericity

Bartlett's Test of Sphericity tests the null hypothesis that the population correlation matrix is an identity matrix (H_0:R = I), which means that there is no correlation between variables. The test statistics used are:

$$\chi^2 = -[(n - 1) - (2p + 5)/6] \cdot \ln|R|$$

where n is the number of observations, p is the sum of variables, and $|R|$ is the determinant of the correlation matrix. This statistic is distributed Chi-Square with the free degree df = $p(p - 1)/2$. If the p-value < 0.05, then H_0 is rejected, which means that there is a significant correlation between variables so that PCA is feasible.

3.6 Correlation Matrix Analysis

The first step in PCA is to understand the relationships between variables through a correlation matrix. After the data was standardized using the Z-score method (z = (x - x̄)/s), a correlation matrix between the five variables was obtained as follows:

Table 3. Correlation Matrix Between Stunting Variables (PM=Poor Population, CID=Basic Immunization Coverage, GBB=Malnutrition of Toddlers, CK4M=K4 Maternity Coverage)

Variabel	PM	CID	BBLR	GBB	CK4M
Poor Population (%)	1,0000	0,4264	0,4822	0,5235	0,4937
Basic Immunization Coverage (%)	0,4264	1,0000	0,8753	0,9048	0,9247
BBLR (%)	0,4822	0,8753	1,0000	0,8783	0,8599
Infant Malnutrition (%)	0,5235	0,9048	0,8783	1,0000	0,9323
K4 Maternity Coverage (%)	0,4937	0,9247	0,8599	0,9323	1,0000

The results of the correlation matrix show some important findings. First, there was a very high correlation between Basic Immunization Coverage and Malnutrition for Toddlers (r = 0.9048) and K4 Maternity Coverage (r = 0.9247). Second, Malnutrition in Toddlers is also very strongly correlated with K4 Maternity Coverage (r = 0.9323), as well as with BBLR (r = 0.8783). Third, BBLR is highly correlated with Basic Immunization Coverage (r = 0.8753) and K4 Maternity Coverage (r = 0.8599). The high correlation pattern between these four variables (Basic Immunization Coverage, BBLR, Toddler Malnutrition, and K4 Maternity Coverage) indicates that the four variables tend to measure the same dimensions, namely the service dimension and the mother-child health status. On the other hand, the Poor variable showed a lower correlation with the other four variables, ranging from r = 0.4264 (with Basic Immunization Coverage) to r = 0.5235 (with Malnutrition in Toddlers). This indicates that the Poor represent a different dimension, namely the socio-economic dimension, which although it has an effect on stunting but has a slightly different mechanism. The high multicollinearity between most of these variables reinforces the reason for using PCA to reduce the data dimension.

3.7. Eigenvalue dan Eigenvector

The third stage in PCA analysis is the calculation of the eigenvalue and eigenvector of the correlation matrix. Eigenvalue represents the magnitude of the variant described by each of the main components, while eigenvector (loading) shows the contribution of each original variable to the main component formed.

Table 4. Eigenvalue and Proportion of Variants of Each Key Component

Components	Eigenvalue (λ)	Proporsi Varian (%)	Cumulative (%)
PC1	3,9976	79,95	79,95
PC2	0,6982	13,96	93,92
PC3	0,1562	3,12	97,04
PC4	0,0884	1,77	98,81
PC5	0,0596	1,19	100,00

Based on Table 3, PC1 has an eigenvalue of 3.9976 and explains 79.95% of the total data variants, far exceeding the Kaiser threshold (eigenvalue > 1) which is generally used as a component selection criterion. PC2 to PC5 have an eigenvalue below 1, with the proportion of variants being only 13.96%, 3.12%, 1.77%, and 1.19%, respectively. Based on the resulting Scree Plot, there was a sharp drop from PC1 to PC2 and then a significant slope, which confirmed that only PC1 needed to be maintained according to the elbow rule criteria. The loading matrix (eigenvector) shows the contribution of each variable to each of the main components. For PC1, the five variables had fairly uniform negative loading: Poor Population (-0.3059), Basic Immunization Coverage (-0.4738), BBLR (-0.4669), Malnutrition of Toddlers (-0.4827), and K4 Maternity Coverage (-0.4806). A uniform negative mark on PC1 indicates that PC1 is a component that reflects the overall health burden, where the higher the PC1 score (negative), the worse the health condition in a district/city.

3.8 Selection of Key Components

Table 4. Selection of the main components

PC	Self-esteem	Propose yourself	Cumulative
PC 1	39.976	79.95	79.95
PC 2	0.6982	13.96	93.92
PC 3	0.1562	3.12	97.04
PC 4	0.0884	1.77	98.81
PC 5	0.0596	1.19	100.00

Based on the two main criteria in the selection of PCA components, namely (1) eigenvalue > 1 and (2) cumulative variance ≥ 70-80%, only PC1 was selected as the main component in this analysis. PC1 meets both criteria simultaneously: PC1's eigenvalue of 3.9976 far exceeds the 1 value, and the cumulative variance covered by PC1 alone has reached 79.95%, which exceeds the required minimum limit of 70%. This means that one main component alone is able to represent almost 80% of the information contained in the five original stunting variables. This condition also confirms the existence of a high redundancy of information between the variables used, which is statistically mediated by the high correlation between variables as shown in the previous correlation matrix.

3.9. First Main Component (PC1)

PC1 is the only component selected and accounts for 79.95% of the total variants. Based on the loading values of all variables on PC1, none of the variables had an absolute loading of ≥ 0.5 individually when referring to the strict loading criteria of > 0.5, however, the four health variables (Basic Immunization Coverage, BBLR, Toddler Malnutrition, and Maternity K4 Coverage) approached the threshold value with loading ranging from -0.4669 to -0.4827. Taking into account the overall magnitude of loading and the consistently negative pattern across variables, PC1 can be interpreted as the Combined Index of Health-Social Burden Related to Stunting. The Malnutrition Variable for Toddlers had the largest contribution (loading -0.4827), followed by Maternity K4 Coverage (-0.4806), Basic Immunization Coverage (-0.4738), BBLR (-0.4669), and Poor Population (-0.3059). The high contribution of the variables of Malnutrition for Toddlers and maternal services (K4 Maternity and Basic Immunization) shows that primary health service factors are the main dimensions that differentiate the level of stunting risk between districts/cities in North Sumatra.

Based on the results of the PC1 score (pc_scores), the distribution of stunting risk burden between districts can be seen. Humbang Hasundutan Regency has the most negative PC1 score (-2.6397), which indicates that this district has the highest stunting risk burden compared to other districts in this analysis. On the other hand, districts/cities with a positive PC1 score such as Asahan Regency (0.4024) showed relatively better conditions. The PCA Biplot visualization confirms this pattern, where the distribution of individuals (districts/cities) that are far in the negative direction of Dim1 describes poorer health conditions.

3.10. Plot Loading Analysis and PCA Biplot

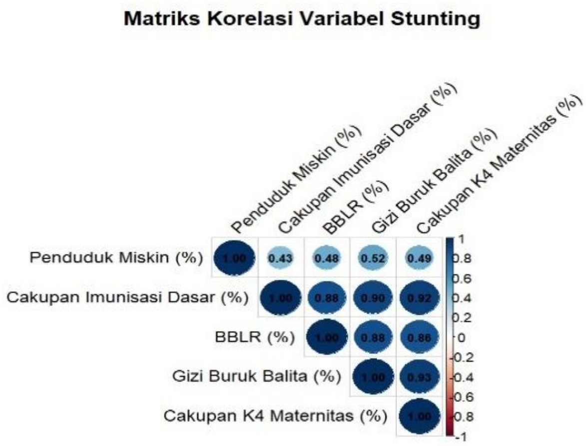

Figure 1. Stunting Variable Correlation Matrix

The results of the visualization show that all correlation coefficients are of positive value (marked by a gradation of a blue circle), which indicates that all variables are moving in unidirection. A very strong and significant correlation relationship (r >= 0.86) was found between the service variable groups and health status, namely between Malnutrition of Toddlers (%) and K4 Maternity Coverage (%) of 0.93; Basic Immunization Coverage (%) and K4 Maternity Coverage (%) by 0.92; Coverage of Basic Immunization (%) and Malnutrition of Toddlers (%) by 0.90; and Basic Immunization Coverage (%) and BBLR (%) by 0.88. This high correlation value confirms the existence of strong multicollinearity between health technical variables, so that the use of PCA as a method of reducing the dimension of data is very valid, precise, and relevant.

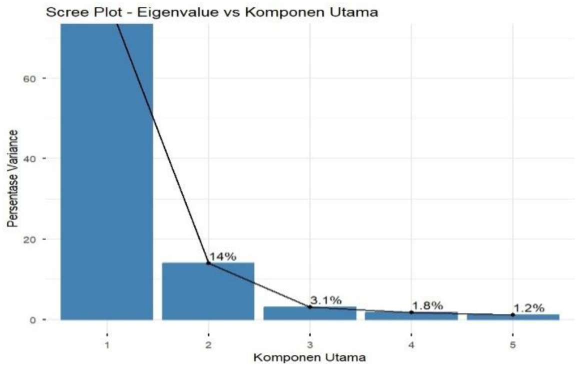

Figure 2. Scree Plot-Eigunvalue Vs Main Components

Based on the elbow plot criteria, there was a drastic decrease in the percentage of variance from the first component to the second component, followed by a pattern that sloped significantly after the second component. This sharp drop (elbow) emphasizes that the absorption of information is saturated after the second component. Therefore, retaining the first two main components (PC1 and PC2) is already very adequate and representative for further analysis, while the remaining three components (PC3 = 3.1%, PC4 = 1.8%, and PC5 = 1.2%) can be eliminated due to their very small and insignificant contribution in explaining the diversity of the data.

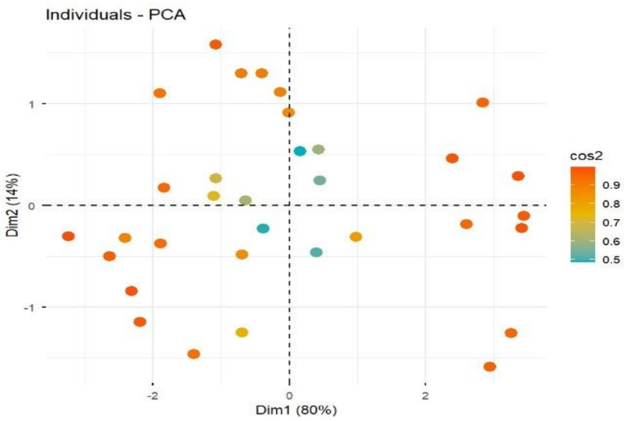

Figure 3. PCA-Individuals

The PCA Individuals graph projects the coordinate position of each district/city in North Sumatra Province into a two-dimensional plane formed by Dim1 (80%) and Dim2 (14%). The quality of representation of each region in this two-dimensional space is measured quantitatively through the cos2 (cosine squared) parameter. Based on the available color scale, the majority of individual districts/cities are dominated by orange to dark red colors with very high cos2 values (ranging from 0.7 to close to 1.0). This proves that the combination of PCA's two-dimensional fields has excellent representation efficiency, so that the geometric position of almost all districts/cities in North Sumatra can be accurately explained without losing a lot of essential information.

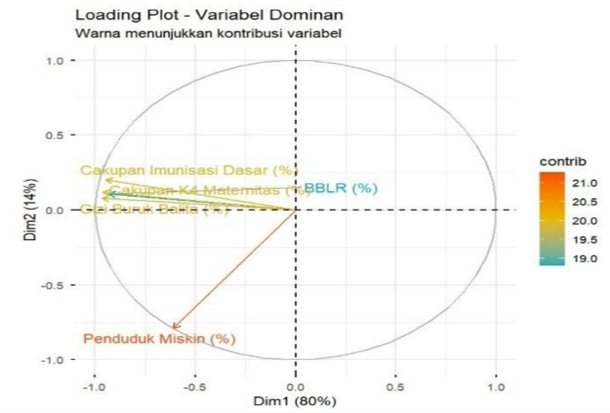

Figure 4. Loading- Dominant Variable

The loading plot on the vector depicts the direction of the vector and the amount of contribution percentage (contrib) of each variable to the formation of Dim1 and Dim2. The contribution parameters are represented by the color gradation of the arrows, where all variables have relatively balanced and significant contribution values, ranging from 19.0% to more than 21.0% (green to orange-red scale). This condition shows that the five variables contribute greatly in explaining the structure of the variability of the data as a whole. A particularly striking characteristic is that the entire variable vector points to the left side of the graph (Dim1 negative), which confirms that all five indicators have a strong negative correlation with the horizontal axis of Dim1.

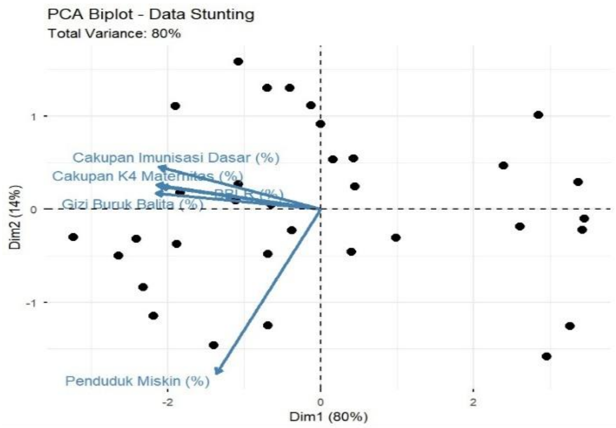

Figure 5. PCA Biplot-Data Stuting

A Biplot PCA image that combines a graph of the individual distribution of districts/cities (black dots) and variable characteristic vectors (blue arrows) in one projection space with a total variance explained of 94%. Through this comprehensive visualization, inter-regional linkages and public health indicators can be interpreted directly. Districts/cities located in the upper left-quadrant (in the direction of the health indicator arrow) show characteristics of areas with high immunization and K4 achievements, but also accompanied by prominent incidence of BBLR and malnutrition. On the other hand, districts/cities in the lower left-left quadrant are areas that are strongly influenced by the Poor Population vector, which reflects pockets with high poverty rates in North Sumatra

4. CONCLUSION

The application of Principal Component Analysis (PCA) to various indicators that are suspected of influencing stunting incidence in all districts/cities of North Sumatra shows that this method is effective in reducing variables that have high correlation to a small number of main components, while sacrificing the substance of the information contained in the data. The main components identified reflect that the determinants of stunting in this region do not stand alone, but are interrelated in three major groups, namely maternal and child health aspects, availability of proper sanitation and access to clean drinking water, and household socio-economic conditions such as income level and parental educational background. Understanding these priority components allows local governments to design and allocate intervention programs in a more targeted and priority-based manner, so that steps to reduce stunting rates can run more effectively and efficiently. Thus, the PCA not only functions as a data reduction tool, but also as a reliable decision-support instrument in the formulation of public policies at the provincial level.

REFERENCES

[1] Alam, S., & Nurcahyo, G. W. (2022). Expert system in diagnosing malnutrition in toddlers using the CBR method. Journal of Information Systems and Technology, 4(4), 143–148

[2] Darmin, Rumaf, F., Ningsih, S. R., Mongilong, R., Goma, M. A. D., & Anggaria, A. D. (2023). The importance of complete basic immunization in infants and toddlers. Journal of Community Service MAPALUS, 1(2), 15–21.

[3] Gemilastari, R., Zeffira, L., Malik, R., & Septiana, V., T. (2024). Characteristics of Babies with Low Birth Weight (BBLR). Scientific Journal, 3(1), 16-26.

[4] Harahap, L. J. (2021). The relationship between knowledge and K4 coverage in pregnant women at the Sangkunur Health Center. Pannmed Scientific Journal, 16(3), 899–903.

[5] Hazimah, M., Akbar, S., Pane, A. H., & Diba, F. (2024). Factors that affect the incidence of low birth weight in Bangka Regency. STM Journal of Medicine (Medical Science and Technology), 7(1), 42–52.

[6] Lesnussa, T. P., Kaseside, M., Utubira, E. E. M. (2025). Application of the Principal Component Analysis Method (Case Study: Poverty Rate in the Maluku Islands). Number: Scientific Journal of Mathematics, Earth and Space, 3(2), 152-160.

[7] Rambe, N. L., Sebayang, W. B. R., & Irsani, N. (2022). Health counseling on the importance of complete basic immunization in the work area of the Terjun Health Center. Scientific Journal of Community Service (Ji-SOMBA), 1(2), 48–52.

[8] Sari, R. K., & Susilowati, E. (2023). Scoping review: Factors that cause malnutrition in toddlers. Journal of Scientific Nutrition, 10(3), 1–9.

[9] Surbakti, E. S. B., Damayanty, Purba, A. M., Azzahra, I. A., Purba, R. R., Santika, & Hasibuan, S. D. (2026). Analysis of factors affecting the incidence of low birth weight babies at Regina Maris Hospital, Medan Maimun District, Medan City, North Sumatra Province in 2025. Scientific Forum and Student Discussion (FORISMA), 7, 799–806.

[10] Suwanti, Ramadhani, S.G., Yuliana, D., Safruddin, N., & Budiman, A. (2026). A systematic review of the application of K-Means clusters and elbow methods for stunting data analysis. Journal of Informatics Engineering Students, 10(2), 3469-3479.

. [11] Ula, M., Ulva, A., F. & Mauliza. (2021). Implementation of Machine Learning with Case Based Reasoning Model in Managing Malnutrition in Children. Journal of Informatics Kaputama(Jik), 5(2),333-339.

[12] Ulfah, S., & Fitrah, N. (2025). Determinants of Complete Basic Immunization Coverage as Part of Stunting Prevention Efforts in Medan City: A Literature Study. Journal of Public Health Clusters, 3(4), 1-10.