International Journal of Engineering and Computer Science Applications (IJECSA)
Vol. 4, No. 2, September 2025, pp. 135∼144
ISSN: 2828-5611
DOI: doi.org/10.30812/ijecsa.v4i2.5395

A Gaussian Mixture Model Approach to Profiling Stunting Risk Across Indonesian Provinces

Masithoh Yessi Rochayani, Iut Tri Utami
Universitas Diponegoro, Semarang, Indonesia

Article Info

Article history:

Received July 23, 2025
Revised July 31, 2025
Accepted August 9, 2025

Keywords:

Clustering
Gaussian Mixture Model
Indonesian Health Survey
Stunting

ABSTRACT

Stunting is still a major health problem in Indonesia, with notable differences between provinces. Although the national rate has decreased over time, regional gaps continue, emphasizing the role of data in helping to explain what contributes to the issue. This study aims to segment 38 provinces in Indonesia based on maternal and child health indicators associated with stunting prevalence. The variables used include the percentage of low birth weight (LBW) infants, the percentage of infants born short, the percentage of pregnant women with chronic energy deficiency (CED), exclusive breastfeeding (EBF) coverage, prevalence of diarrhea in toddlers, and prevalence of acute respiratory infections (ARI) in toddlers. The clustering analysis was performed using the Gaussian Mixture Model (GMM) with the number of clusters varied from 2 to 7. Model selection was based on the Bayesian Information Criterion (BIC), where the lowest value indicated the optimal model. The results show that the model with two clusters was selected, with a BIC value of 1358.24, which indicates the best balance between model fit and complexity. This clustering reveals that provinces are grouped based on similarities in maternal and child health profiles, not on geographic proximity, meaning that the GMM method does not rely on spatial location to form clusters.

Corresponding Author:

NMasithoh Yessi Rochayani,
Universitas Diponegoro, Semarang, Indonesia,
Email: yessi.rochayani@live.undip.ac.id

How to Cite: M. Y. Rochayani and I. T. Utami, “A Gaussian Mixture Model Approach to Profiling Stunting Risk Across Indonesian Provinces”, International Journal of Engineering and Computer Science Applications (IJECSA), vol. 4, no. 1, pp. 135–144, Sep. 2025. doi: 10.30812/ijecsa.v4i2.5395.

1. INTRODUCTION

Stunting is a condition of growth failure in toddlers due to chronic malnutrition, particularly during the first 1000 days of life [1]. A child is considered stunted, or shorter than expected for their age, when their height-for-age falls more than two standard deviations below the median of the WHO Child Growth Standards [2]. Stunting is an indicator of human resource quality. This is because stunted children are at risk of delayed brain development, reduced learning ability, and are at risk of chronic diseases. In Indonesia, the prevalence of stunting was 21.5% in 2023, which declined to 19.8% in 2024. The government aims to reduce this figure to 18.8% by 2025. Although conditions have improved nationally, there are still large disparities between provinces, with some regions remaining far above the national average. These regional variations suggest that stunting is a multidimensional issue influenced by complex factors, making it important to understand how determinants of stunting vary across provinces.

Several factors have been widely recognized as contributors to stunting, particularly those related to maternal and child health. These include the incidence of low birth weight (LBW) among infants [3], birth length less than 48 cm [4], and the level of chronic energy deficiency (CED) in pregnant women, which reflects nutritional and health status during pregnancy. Research [5] shows that pregnant women who experience CED have a 13.6 times greater risk of giving birth to short or very short babies (stunting) compared to mothers who do not experience CED. In addition, exclusive breastfeeding during the first six months of life plays an important role in supporting optimal growth and development [6]. Poor breastfeeding practices, particularly early weaning, can lead to nutritional deficits. On the other hand, the child's health condition, especially when affected by frequent illnesses such as diarrhea [7] and acute respiratory infections (ARI) [8], can impair nutrient absorption and further worsen nutritional status. These multiple, interrelated factors show that stunting arises from both direct and indirect causes.

To explore how different risk factors are distributed across provinces, one approach that can be used is clustering. Clustering is a statistical analysis that aims to combine objects based on variables into groups that have different characteristics between one group and another. The objects will be organized into groups, with each group consisting of items that have similar features or characteristics. Several clustering methods have been used to group regions based on factors influencing stunting, including K-Means [9], Hierarchical Clustering [10], and Possibilistic Fuzzy C-Means (PFCM) [11]. However, these methods have limitations. For example, K-Means assigns each data point to only one cluster and assumes that all clusters have round shapes and the same size [12], which may not reflect the true structure of real-world health data. Hierarchical clustering can be sensitive to linkage methods and distance metrics. Meanwhile, fuzzy clustering methods such as PFCM allow for soft membership. Still, they are often sensitive to noise and initialization parameters, and lack the probabilistic rigor required for likelihood-based model evaluation. These limitations indicate the need for a more flexible and statistically grounded clustering method better to capture the complexity of stunting determinants across different regions.

Although various clustering approaches have been employed in previous studies, relatively few have utilized Gaussian Mixture Models (GMM) to group regions based on stunting-related determinants. Filling this gap is essential because GMM is more flexible in handling complex data. Compared to traditional clustering methods such as K-Means, which assume spherical clusters and perform hard partitioning, GMM adopts a probabilistic framework that better captures data uncertainty by allowing soft assignments and more adaptable cluster shapes [13]. Although fuzzy clustering methods like Fuzzy C-Means and PFCM allow soft membership, they are not based on a probabilistic model. In contrast, GMM allows likelihood-based estimation, supports model selection using criteria such as BIC, and offers better interpretability when modeling real-world distributions. Given these advantages, GMM is a suitable method for finding groups of regions with similar health conditions related to stunting. This study aims to use GMM to identify clusters of 38 Indonesian provinces based on maternal and child health indicators related to stunting, including LBW, short birth, CED in pregnant women, exclusive breastfeeding, history of diarrhea, and ARI. This study contributes to stunting reduction efforts by using GMM to group provinces based on health-related risk factors.

2. RESEARCH METHOD

2.1. Gaussian Mixture Model

The Gaussian Mixture Model (GMM) is a model-based clustering method that assumes data comes from a mix of several normal distributions. Instead of putting each data point into just one group, GMM gives a probability that the data belongs to each group. This classifies the method as belonging to the soft clustering approach (see Equation (1)). Let K be the number of components (clusters), and let the GMM define the probability density function as a weighted sum of multivariate Gaussian distributions (see Equation (2)).

$$p(x) = \sum_{k=1}^{K} \pi_k N(x|\mu_k, \Sigma_k) \tag{1}$$

where π_k is the mixing weight for the k-th component, satisfying

$$\sum_{k=1}^{K} \pi_k = 1, \quad 0 \leq \pi_k \leq 1 \tag{2}$$

$N(x|\mu_k, \Sigma_k)$ is the multivariate normal distribution defined in Equation (3).

$$N(x|\mu_k, \Sigma_k) = \frac{1}{(2\pi)^{d/2} |\Sigma_k|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1} (x-\mu_k)\right) \tag{3}$$

with $x \in R^d$ is a d-dimensional observation vector, $\mu_k \in R^d$ is the mean vector of the k-th component, $\Sigma_k \in R^{d\times d}$ is the covariance matrix of the k-th component, $|\Sigma_k|$ is the determinant of the covariance matrix, and Σ_k^{-1} is the inverse of the covariance matrix [14].

2.2. Parameter Estimation of GMM with EM Algorithm

To estimate the parameter $\theta = \{\pi_k, \mu_k, \Sigma_k\}_{k=1}^{K}$, the Expectation-Maximization (EM) algorithm is used. The steps of the EM algorithm are described as follows.

1. Initialization
   Determine initial values for the parameters $\pi_k^{(0)}, \mu_k^{(0)}, \Sigma_k^{(0)}$ for all k.
2. Expectation step
   Calculate the posterior probability that the data x_i, comes from the k-th component using Equation (4).

$$\gamma_{ik}^{(t)} = \frac{\pi_k^{(t)} N\left(x_i \middle| \mu_k^{(t)}, \Sigma_k^{(t)}\right)}{\sum_{j=1}^{K} \pi_k^{(t)} N\left(x_i \middle| \mu_k^{(t)}, \Sigma_k^{(t)}\right)} \tag{4}$$

   For all $x_i, i = 1, \ldots, n$ and each component $k = 1, \ldots, K$.
3. Maximization step
   Update GMM parameters based on values $\gamma_{ik}^{(t)}$ including:
   Mixing coefficient update using Equation (5):

$$\pi_k^{(t+1)} = \frac{1}{n} \sum_{i=1}^{n} \gamma_{ik}^{(t)} \tag{5}$$

   Update component mean using Equation (6).

$$\mu_k^{(t+1)} = \frac{\sum_{i=1}^{n} \gamma_{ik}^{(t)} x_i}{\sum_{i=1}^{n} \gamma_{ik}^{(t)}} \tag{6}$$

   Covariance matrix update using Equation (7).

$$\Sigma_k^{(t+1)} = \frac{\sum_{i=1}^{n} \gamma_{ik}^{(t)} \left(x_i - \mu_k^{(t+1)}\right)\left(x_i - \mu_k^{(t+1)}\right)^T}{\sum_{i=1}^{n} \gamma_{ik}^{(t)}} \tag{7}$$

4. Convergence evaluation
   Calculate log-likelihood using Equation (8).

$$\log L(\theta) = \sum_{i=1}^{n} \log \left(\sum_{k=1}^{K} \pi_k N(x_i | \mu_k, \Sigma_k) \right) \tag{8}$$

   If the difference in log-likelihood between iterations is smaller than a specified threshold or the maximum number of iterations is reached, then the algorithm stops.

2.3. Bayesian Information Criterion (BIC)

The implementation of the Gaussian Mixture Model (GMM) involves determining the optimal number of clusters. One evaluation method commonly used is the Bayesian Information Criterion (BIC), which incorporates information from the log-likelihood value and model complexity. The BIC formula is expressed in Equation (9).

$$BIC = -2\, L(\boldsymbol{\theta}; \mathbf{x}) + df\, ln(n) \tag{9}$$

where $L(\boldsymbol{\theta}; \mathbf{x})$ is the log-likelihood and n is the number of observations. The number of parameters or degrees of freedom (df) is calculated by Equation (10).

$$df = (k-1) + kd + h \tag{10}$$

where s is the number of clusters and d is the number of variables. The number of parameters for the mixture proportions is $(k-1)$. The number of mean parameters is (kd). Additionally, h denotes the number of covariance parameters. In this study, it is assumed that all clusters share the same covariance matrix (see Equation (11)) [15].

$$h = \frac{d(d+1)}{2} \tag{11}$$

2.4. Research Data

The data used in this study comes from the 2023 Indonesian Health Survey (SKI 2023). This study utilizes provincial-level data representing various maternal and child health indicators related to stunting prevalence. The variables used in the analysis include: the percentage of infants with low birth weight (LBW), the percentage of infants born with a body length below the standard (short birth), the percentage of experiencing chronic energy deficiency (CED) among pregnant women, the percentage of exclusive breastfeeding (EBF), the prevalence of diarrhea in toddlers, and the prevalence of acute respiratory infections (ARI) in toddlers.

2.5. Stages of Data Analysis and Modeling

This study involves three main analytical procedures. First, an Exploratory Data Analysis (EDA) was conducted, which included calculating descriptive statistics such as the minimum, maximum, mean, and standard deviation for each variable. The distributions of the variables were visualized using boxplots to assess their spread and to detect potential outliers, while multivariate outlier detection was performed using Mahalanobis squared distances. In addition, Pearson correlation coefficients were computed to evaluate the strength and direction of linear relationships among the variables.

The second stage involved Gaussian Mixture Model (GMM) modeling, where clustering was performed with the number of clusters ranging from two to seven. Model parameters were estimated using the Expectation-Maximization (EM) algorithm until convergence was achieved. For each fitted model, the Bayesian Information Criterion (BIC) was calculated, and the BIC values were compared across models. The model with the lowest BIC value was then selected as the optimal clustering solution. Finally, the interpretation phase focused on examining the characteristics of each cluster by evaluating the average values of all variables, allowing meaningful profiling of the resulting groups.

3. RESULT AND ANALYSIS

3.1. Exploratory Data Analysis (EDA)

Table 1 summarizes the descriptive statistics of six health indicators related to maternal and child health across Indonesian provinces. The percentage of low birth weight (LBW) infants shows relatively low variability, with a mean of 5.90% and a standard deviation of 1.35%. In contrast, the percentage of infants born with short birth length (Short Birth) and the prevalence of chronic energy deficiency (CED) among pregnant women exhibit higher variability, with standard deviations of 6.85% and 7.57%, respectively. Exclusive breastfeeding (EBF) coverage has the highest variation (s = 9.74%), indicating significant differences in infant feeding practices across provinces. Meanwhile, the average prevalence of diarrhea and acute respiratory infections (ARI) in toddlers remains low, although some provinces show much higher values. These variations across indicators suggest notable disparities in maternal and child health conditions, supporting the need for clustering analysis to group provinces with similar characteristics.

Table 1. Descriptive Statistics of Risk Factors for Stunting Across Provinces

Statistic	LBW	Short Birth	CED	EBF	Diarrhea	ARI
Minimum	2.7	7	5.2	33.4	1.4	0.67
Maximum	8	32.4	44.7	71.4	17.5	11.8
Mean	5.9	19.88	17.15	53.18	4.62	3.93
Median	5.95	19.95	17.55	54.45	4.3	3.26
Standard deviation	1.35	6.85	7.57	9.74	2.66	2.89

The boxplot in Figure 1 shows how the six variables used for clustering are spread out. Among them, exclusive breastfeeding (EBF) has the widest spread, with a median of around 55% and a broad range, which means there are big differences between provinces. On the other hand, low birth weight (LBW) and diarrhea rates in toddlers have lower medians and tighter spreads, suggesting the values are close to the middle and similar across regions. Some variables, such as CED, diarrhea, and ARI, also have outliers, indicating that a few provinces have significantly higher or lower values than the rest.

Identifying outliers in multivariate data requires more than just a boxplot approach. Instead, the Mahalanobis squared distance is used and compared against a critical chi-square value at the 0.05 significance level, with degrees of freedom equal to the number of variables. In this study, which includes six variables, the threshold for detecting outliers is 12.592. The calculation results show that two provinces have Mahalanobis distance values exceeding this limit, Papua Tengah and Papua Pegunungan. This suggests that these two provinces exhibit significantly different characteristics in multivariate aspects compared to other provinces and can be considered outliers. However, in this study, outliers were not removed during GMM modeling. According to [16], GMM can still work well without removing outliers. In addition, removing outliers means that grouping is only carried out on provinces that are not detected as outliers.

Figure 2 displays the Pearson correlation matrix between the six variables. A moderate positive correlation was observed between LBW and Short Birth (r = 0.45), indicating that provinces with a high proportion of LBW infants tend also to have a high proportion of Short Birth. Another significant positive correlation was between maternal CED and diarrhea prevalence (r = 0.53), reflecting a possible link between maternal nutritional status and children's susceptibility to infectious diseases. Another moderate correlation was observed between diarrhea prevalence and ARI prevalence in toddlers (r = 0.55).

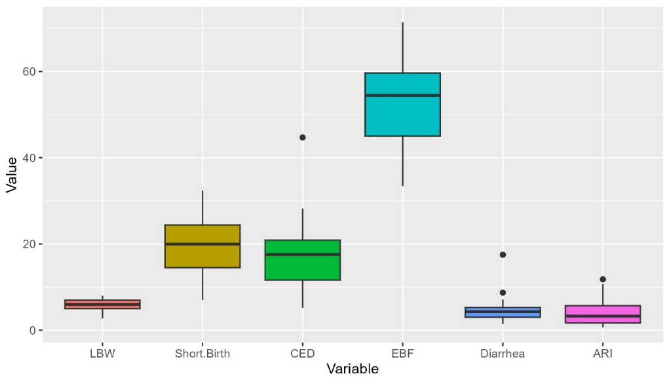

Figure 1. Boxplots of Variables Related to Stunting Risk

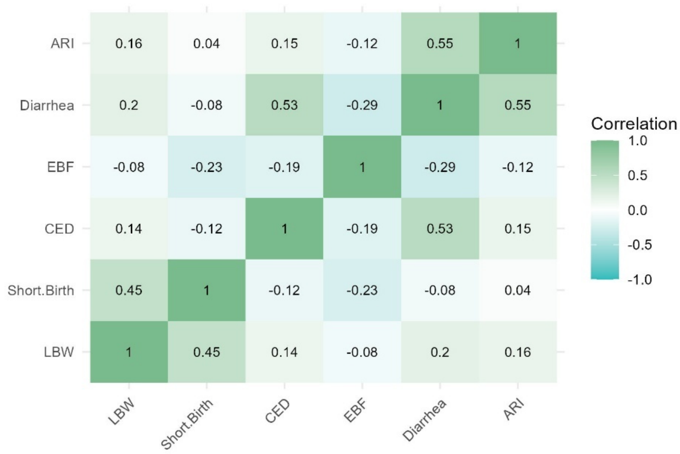

Figure 2. Correlation Matrix of Variables Related to Stunting Determinants

Meanwhile, a significant negative correlation was found between exclusive breastfeeding coverage (EBF) and the prevalence of diarrhea in toddlers (r=-0.29), indicating that areas with higher exclusive breastfeeding practices tend to have lower diarrhea cases. This is consistent with medical knowledge that exclusive breastfeeding protects infants from exposure to pathogens that cause gastrointestinal infections [17]. Furthermore, EBF was also negatively correlated with the proportion of stunted infants (-0.23) and chronic energy deficiency (CED) in pregnant women (r=-0.19), which may reflect better maternal and infant health conditions in areas with high breastfeeding coverage. Another negative correlation, although weak, was observed between stunted infants and diarrhea (r=-0.08), suggesting a slight tendency that areas with high stunting rates may have a slightly lower diarrhea prevalence.

3.2. GMM Clustering

In this study, Gaussian Mixture Model (GMM) modeling was performed with 2 to 7 clusters. The best model was selected based on the Bayesian Information Criterion (BIC), with the lowest BIC value indicating the best model. The GMM modeling results are summarized in Table 2.

Table 2. Descriptive Statistics of Risk Factors for Stunting Across Provinces

Number of Clusters	Number of Iterations	Log-likelihood	Degrees of freedom	BIC
2	17	-617.28	34	1358.24
3	24	-606.01	41	1361.17
4	8	-593.16	48	1360.93
5	9	-592.71	55	1385.49
6	25	-590.62	62	1406.77
7	18	-567.83	69	1386.66

Table 2 shows the relationship between the number of clusters, the log-likelihood value, degrees of freedom (the number of model parameters), and the BIC. In general, the log-likelihood value increases with the number of clusters. This is because the model has more parameters, allowing it to fit the data better. The more clusters formed, the greater the number of degrees of freedom, as stated in Equation (10). The increase in parameters causes a penalty in the BIC calculation (the second term in Equation (10)), because BIC not only considers the model's fit to the data (through the log-likelihood value), but also takes into account the model's complexity. As a result, although the log-likelihood value tends to increase with increasing number of clusters, the BIC value can actually increase if a significant increase in model fit does not accompany the additional parameters. In other words, even though the likelihood continues to increase, the BIC value does not always decrease. In this study, the BIC reached its lowest value in the two-cluster model (BIC = 1358.242) and increased as the number of clusters increased, for example, to 1406.770 in six clusters and 1386.660 in seven clusters. This finding suggests that increasing the number of clusters does not always result in a better model. This result is consistent with [15], which explains that although the log-likelihood increases with increasing the number of clusters, the BIC may increase if the increase is not proportional to the increase in the number of parameters (see in Figure 3).

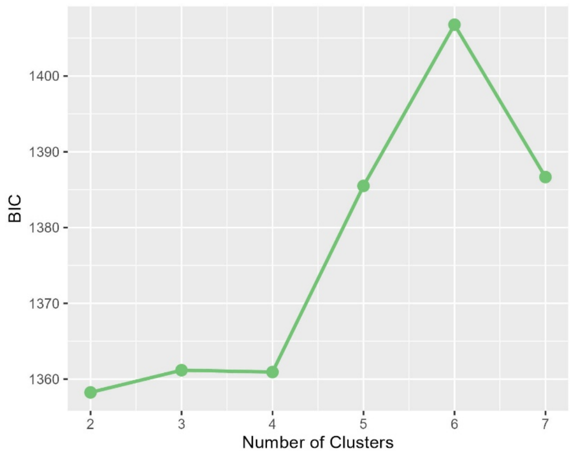

Figure 3. BIC Values for GMM Models with Varying Number of Clusters

3.3. Interpretation of Clustering Results

The model with two cluster was selected in this study because it yielded the lowest BIC. The BIC for the two-cluster model was 1358.24, indicating the best balance between model fit (goodness-of-fit) and model complexity compared to other tested models. The clustering result with two main groups illustrates distinct characteristics in factors influencing stunting prevalence across provinces. Cluster 1, consisting of 16 provinces, generally reflects relatively better conditions in maternal and child health. The average percentage of low birth weight (LBW) in this cluster ranges from 2.7% to 7.8%, with most provinces reporting values below 6%. The percentage of short birth is also relatively low to moderate, with the majority below 16%, except for some provinces like Banten (20.9%) and West Java (19.9%). The proportion of chronic energy deficiency (CED) among pregnant women in this cluster varies but remains within manageable levels, approximately 5.2% to 20%. The coverage of exclusive breastfeeding (EBF) is mostly high, ranging between 50% and 68%, indicating a strong adherence to breastfeeding practices that support child development. Additionally, the prevalence of diarrhea and acute respiratory infections (ARI) among toddlers in this cluster is relatively low, typically not exceeding 7

The results of this study obtained 2 clusters, as shown in Table 3, where Cluster 2 includes 22 provinces characterized by more complex conditions. The average percentage of infants born with short birth length is higher compared to Cluster 1, with several provinces reporting figures exceeding 25%, such as Sulawesi Utara (32.4%), Gorontalo (29.8%), and D.I. Yogyakarta (28.7%). Although the coverage of exclusive breastfeeding is relatively high in many provinces within this cluster (above 60% in Jawa Tengah, Nusa Tenggara Timur, and Sumatera Barat), some provinces show significantly low coverage, such as Papua Barat Daya (33.4%) and Gorontalo (37.6%). The percentage of pregnant women experiencing chronic energy deficiency (CED) is also high in several provinces, including Papua Selatan (28.2%) and Papua Pegunungan (44.7%). Moreover, disparities are evident in the prevalence of diarrhea and acute respiratory infections (ARI) among toddlers, with several provinces recording considerably high rates, such as Papua Tengah for ARI (11.8%) and Papua Pegunungan for diarrhea (17.5%).

The findings of this research are that certain provinces with distinct geographic locations, such as DKI Jakarta, Banten, and Papua, were grouped into the same cluster (Cluster 2). This implies that the clustering by GMM relies on similar characteristics in stunting indicators rather than spatial location. This result is in line with [18] which groups provinces in Indonesia based on poverty indicators. Even though it is not stated directly in [18], the clustering results demonstrate that the groupings are not based on geographic proximity. This confirms that the clusters were formed based on similar health-related indicators, rather than their physical or regional locations.

Table 3. Cluster Membership Based on the Optimal GMM Solution

Cluster	Number of cluster members	Provinces included
Cluster 1	16	Aceh, Sumatera Utara, Riau, Jambi, Bengkulu, Kepulauan Riau, Jawa Barat, Jawa Timur, Banten, Bali, Nusa Tenggara Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Utara, Sulawesi Tenggara, Papua Pegunungan

Cluster	Number of cluster members	Provinces included
Cluster 2	22	Sumatera Barat, Sumatera Selatan, Lampung, Kep. Bangka Belitung, D.K.I. Jakarta, Jawa Tengah, D.I. Yogyakarta, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Timur, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, Papua Barat Daya, Papua, Papua Selatan, Papua Tengah

4. CONCLUSION

This study shows how the Gaussian Mixture Model (GMM) can help group provinces in Indonesia based on maternal and child health factors linked to stunting. Provinces were grouped into two clusters based on the lowest BIC value, which was 1358.24. The first cluster generally includes provinces with better maternal and child health conditions, while the second cluster comprises provinces with more complex and challenging health profiles. Interestingly, GMM was able to group provinces not by geographic proximity, but based on similar patterns in health indicators. These findings suggest that GMM can find hidden health patterns and help improve stunting prevention.

ACKNOWLEDGEMENTS

This research was funded by the Faculty of Science and Mathematics, Universitas Diponegoro, through grant number: 25.F/UN7.F8/PP/II/2025.

REFERENCES

[1] Kementerian Kesehatan Republik Indonesia, *Survei Kesehatan Indonesia (SKI) 2023 Dalam Angka*. 2023.

[2] O. Karlsson *et al.*, "Maternal Height-standardized Prevalence of Stunting in 67 Low- and Middle-income Countries," *Journal of Epidemiology*, vol. 32, no. 7, pp. 337–344, Jul. 5, 2022. DOI: 10.2188/jea.JE20200537. PMID: 33612705.

[3] S. C. Sholihah, "Hubungan Berat Badan Lahir Rendah (BBLR) terhadap Kejadian Stunting di Wilayah Kerja Puskesmas Dradah," *PREPORTIF: Jurnal Kesehatan Masyarakat*, vol. 7, no. 1, pp. 135–140, Dec. 20, 2023. DOI: 10.31004/prepotif.v7i1.10859.

[4] J. Judiono *et al.*, "Faktor Determinan Panjang Badan Bayi Lahir Pendek sebagai Faktor Risiko Stunting di Jawa Barat: Determinant Factors of Short Birth Length Baby as a Risk Factor of Stunting in West Java," *Amerta Nutrition*, vol. 7, no. 2, pp. 240–247, Jun. 9, 2023. DOI: 10.20473/amnt.v7i2.2023.240-247.

[5] W. Rohmawati, P. D. Wintoro, and T. W. Sari, "Hubungan Kekurangan Energi Kronik pada Ibu Hamil dengan Kejadian Stunting di Klaten," *MOTORIK Jurnal Ilmu Kesehatan*, vol. 16, no. 1, pp. 40–44, Jul. 15, 2021. DOI: 10.61902/motorik.v16i1.233.

[6] S. O. M. Ahmed *et al.*, "Impact of exclusive breastfeeding on physical growth," *Clinical Nutrition Open Science*, vol. 49, pp. 101–106, Jun. 2023. DOI: 10.1016/j.nutos.2023.04.008.

[7] C. Desyanti and T. S. Nindya, "Hubungan Riwayat Penyakit Diare dan Praktik Higiene dengan Kejadian Stunting pada Balita Usia 24-59 Bulan di Wilayah Kerja Puskesmas Simolawang, Surabaya," *Amerta Nutrition*, vol. 1, no. 3, p. 243, Oct. 23, 2017. DOI: 10.20473/amnt.v1i3.6251.

[8] A. Muthiyah, "Literature review: Korelasi kejadian ispa terhadap stunting," *Jurnal Ilmiah Kesehatan Sandi Husada*, vol. 10, no. 2, pp. 729–733, 2021. DOI: 10.35816/jiskh.v10i1.1146.

[9] Z. I. Alfianti, "Pengelompokan Wilayah Kasus Balita Stunting di Indonesia Menggunakan Algoritma K-Means," *Jurnal Ilmiah Informatika Komputer*, vol. 28, no. 3, pp. 173–185, 2023. DOI: 10.35760/ik.2023.v28i3.8876.

[10] D. Satriawan and D. A. Styawan, "Pengelompokkan Provinsi di Indonesia Berdasarkan Faktor Penyebab Balita Stunting Menggunakan Analisis Cluster Hierarki," *Jurnal Statistika dan Aplikasinya*, vol. 5, no. 1, pp. 61–70, Jun. 30, 2021. DOI: 10.21009/JSA.05106.

[11] A. R. Maulana, M. Y. Rochayani, and M. A. Mukid, "Regional Clustering for Stunting Prevalence Analysis in Central Java Using Possibilistic Fuzzy C-Means (PFCM) Algorithm," in *2025 International Conference on Computer Sciences, Engineering, and Technology Innovation (ICoCSETI)*, Jan. 2025, pp. 635–640. DOI: 10.1109/ICoCSETI63724.2025.11019156.

[12] J. Heidari, N. Daneshpour, and A. Zangeneh, "A novel K-means and K-medoids algorithms for clustering non-spherical-shape clusters non-sensitive to outliers," *Pattern Recognition*, vol. 155, p. 110 639, Nov. 2024. DOI: 10.1016/j.patcog.2024.110639.

[13] M. Omari *et al.*, "Advancing Image Compression Through Clustering Techniques: A Comprehensive Analysis," *Technologies*, vol. 13, no. 3, p. 123, Mar. 19, 2025. DOI: 10.3390/technologies13030123.

[14] M. P. Deisenroth, A. A. Faisal, and C. S. Ong, *Mathematics for Machine Learning*, 1st ed. Cambridge University Press, Feb. 29, 2020. DOI: 10.1017/9781108679930.

[15] C. Fraley and A. E. Raftery, "Model-Based Clustering, Discriminant Analysis, and Density Estimation," *Journal of the American Statistical Association*, vol. 97, no. 458, pp. 611–631, Jun. 1, 2002. DOI: 10.1198/016214502760047131.

[16] C. Guyeux *et al.*, "Introducing and Comparing Recent Clustering Methods for Massive Data Management in the Internet of Things," *Journal of Sensor and Actuator Networks*, vol. 8, no. 4, p. 56, Dec. 9, 2019. DOI: 10.3390/jsan8040056.

[17] B. Branger *et al.*, "Breastfeeding and respiratory, ear and gastro-intestinal infections, in children, under the age of one year, admitted through the paediatric emergency departments of five hospitals," *Frontiers in Pediatrics*, vol. 10, p. 1 053 473, Feb. 15, 2023. DOI: 10.3389/fped.2022.1053473.

[18] N. N. Alyarahma, G. Kholijah, and C. Sormin, "Pengelompokan Provinsi di Indonesia Menggunakan Gaussian Mixture Model Berdasarkan Indikator Kemiskinan," *Journal of Mathematics: Theory and Applications*, vol. 6, no. 2, pp. 158–167, Oct. 14, 2024. DOI: 10.31605/jomta.v6i2.4032.

[This page intentionally left blank.]