STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **x**, Month 2025, pp 0–9.
Published online in International Academic Press (www.IAPress.org)

Utilizing Multi-Arm Bandit and Partitioning Around Medoids for Clustering Food Security Conditions in Sumatra Island

Muhammad Subianto [1,2,*], Naila Anastasya Anshor [1], Evi Ramadhani [1], Nany Salwa [1]

[1] *Department of Statistics, Universitas Syiah Kuala, Aceh, Indonesia*
[2] *Department of Informatics, Universitas Syiah Kuala, Aceh, Indonesia*

Abstract This study applies the Multi Arm-Bandit (MAB) and Partition Around Medoid (PAM) methods to cluster food security status in Sumatra Island in 2022. Food issues in Indonesia are becoming increasingly complex with challenges covering food availability, accessibility and food security. Data were obtained from the Food Security and Vulnerability Atlas (FSVA). The MAB method was used to identify the most influential variables in decision-making, while the PAM method was used for clustering based on medoids. The results showed that the PAM method was effective in categorizing provinces of Sumatra Island based on food security variables. Additionally, significant variables identified included poverty, food expenditure, disease rate, and stunting. This study provides important contributions to the government and the National Food Agency in designing food security improvement programs in various regions of Sumatra Island and serves as a reference for other researchers applying similar methods in complex data analysis.

Keywords MAB, PAM, Cluster, Food Security Status.

AMS 2010 subject classifications 62-07, 62P10, 62P12, 68T01, 68U01

DOI: 10.19139/soic-2310-5070-2456

1. Introduction

Technological advancements in area computing and data storage are progressing rapidly. A significant development is the occurrence of processors specifically designed for machine learning, which enables simultaneous data processing and facilitate the formation of more sophisticated algorithms [1]. These advancements have made the process increasingly possible in applying machine learning to large datasets, uncovering patterns that are difficult to detect through conventional methods, a process often known as data mining.

Data mining is an interdisciplinary field that combines machine learning, pattern recognition, statistics, databases, and visualization methods to extract valuable information from large databases. A major component of data mining is the use of neural networks, which operates under two main types of learning, namely supervised and unsupervised. Supervised learning comprises classification or labeling of data/variables, while unsupervised handles data that lack predefined classes or labels. Common unsupervised learning methods include estimation models, association analysis such as linear regression, and cluster analysis [2].

Cluster analysis is a data mining method used to identify groups of objects with similar characteristics, enabling the differentiation of distinct classes in a dataset. Various clustering methods have been developed, including K-Means, Partition Around Medoids (PAM), single-link, and complete-link [3]. Different from K-Means which relies on mean value of each cluster, PAM uses medoids, data points with the smallest total distance to all other points in

*Correspondence to: Muhammad Subianto (Email: subianto@usk.ac.id). Department of Statistics, Universitas Syiah Kuala, Aceh, Indonesia.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

the cluster. This distinction makes PAM a more flexible clustering method, enabling comprehensive and in-depth analysis [4]. The method has proven highly useful diverse applications across various fields, including food security patterns on Sumatra Island in 2022.

During this study, the PAM clustering method is combined with the Multi-Arm Bandit (MAB) algorithm. MAB algorithm is a decision-making method used to identify the best choice among several options. The fundamental aspect of this method is decision-making for a variable that does not affect the properties of the variable or others. Consequently, MAB method aims to identify the most influential factors based on the accuracy of exploitation and exploration trade-offs. This method is applied to data containing numerical variables during the analysis.

Food issues in Indonesia are becoming increasingly complex, with challenges incorporating food availability, accessibility, and security. Poor food quality remains prevalent in some areas, including Sumatra Island. Moreover, food security data for Sumatra Island in 2022 were obtained from Food Security and Vulnerability Atlas (FSVA) [5]. The National Food Agency (NFA) Indonesia has conducted surveys and developed digital maps to identify clusters at the district, sub-district, and even village levels [6]. Following the discussion, food security standards include numerous factors and variables. Food should always be sufficient and accessible for purchase by the population to meet daily needs, reflecting the prosperity and welfare of Indonesian citizens. Food sufficiency plays a crucial role in determining the quality of of human resources. Studies have shown that the primary causes of food vulnerability include poverty, high prevalence of stunting among children under five, limited access to clean water, and inadequate healthcare services. In line with this discussion, stunting remains a critical global health issue, affecting 22% of children under five, with the burden disproportionately higher in Low and Middle Income Countries (LMICs).

Countries including Nepal, Bangladesh, and Vietnam have reduced stunting by over 30% through comprehensive, multisectoral strategies that combine both nutrition-specific and as well as nutrition-sensitive interventions. Major drivers include policy commitment, community health programs, and poverty reduction. Concerning the ongoing discussion, challenges such as wasting and inequities. Lessons learned from these countries show scalable methods to success offer scalable methods that can help accelerate progress toward achieving global stunting targets [7]. Another study related to food security [8], in this study examines the role of food security (availability, access, usage) in stunting prevention among 113 mothers of stunted children in Bondowoso, Indonesia. Using PLS-SEM, results show food availability as the strongest predictor (β=0.58-0.86) of prevention behaviors, while access moderately improves feeding practices. The 32.1% local stunting rate signifies the need for incorporated food security and nutrition interventions in rural agrarian communities to address chronic under nutrition effectively.

Managing the issues presents significant challenges due to the complexity and dynamic nature of the data included, comprising a large number of numerical variables that continuously change over time. MAB algorithm aids in intelligent adaptive decision-making by allowing resource allocation to adjust based on new information gathered over time. Similarly, clustering algorithms such as PAM have the potential to group complex numerical data into more structured clusters, enabling the identification of relevant patterns for understanding food security levels.

2. Method

2.1. Data and Variable

The data used in this study were obtained from the National Food Agency Indonesia in the form of FSVA (https://fsva.badanpangan.go.id/) for the year 2022. This study used eight variables during the process, as shown in Table 1.

2.2. Procedure Analysis

The data analysis in this study was conducted using R 4.3.1, QGIS 3.38.0, and Python 3.11.5. The steps in the data analysis procedure were as follows.

Table 1. Research Variables.

Notation	Variable	Scale
X_1	Poverty	Percent
X_2	Food Expenditure	Percent
X_3	Without Electricity	Percent
X_4	Without Clean Water	Percent
X_5	Women's Education Length	Index
X_6	Health Worker	Ratio
X_7	Disease Rate	Percent
X_8	Stunting	Percent

1. Descriptive statistical analysis
   Descriptive statistical analysis was performed to acquire an initial understanding of the data characteristics, such as mean, median, minimum, and maximum values.
2. Data preparation
   Data preparation during the process included normalization using the z-score method. Z-score normalization helped standardize different measurement scales, making variables more comparable in the analysis. It was particularly effective in reducing the influence of original measurement units and ensuring variables were more equitable in the analysis [9]. The transformation ensured the data met necessary assumptions, and the z-score was calculated using the following formula.

$$Z_{\text{score}} = \frac{X_i - \mu}{\sigma} \tag{1}$$

   where:

   - X = raw data value
   - μ = mean of the dataset
   - σ = standard deviation of the dataset

3. Implementation of Multi-Arm Bandit (MAB)
   MAB method was applied based on the balance of exploitation and exploration values, representing various factors affecting food security. During the analysis, the epsilon-greedy algorithm was used, which was a popular method in many studies due to its effective balance between exploration and exploitation. Major actions in epsilon-greedy were [10]:

   - Exploration: Selecting a random option to gather information about choices that might not have been comprehensively explored.
   - Exploitation: Selecting the best-known option based on previously gathered information.

   The probability of selecting arm i at time $t+1$, signified as $P_i(t+1)$, was calculated using [11]:

$$P_i(t+1) = \begin{cases} 1 - \epsilon + \frac{\epsilon}{\kappa} \\ \frac{\epsilon}{\kappa} \end{cases} \tag{2}$$

where:

- ϵ: A parameter controlling the balance between exploration and exploitation.
- κ: Total number of arms available.

In the application of MAB:
Each option, or "arm," had its probability of producing a particular result. The model continuously updated the estimation of each value of arm based on the outcomes obtained. In addition, the distribution of exploration and exploitation actions ensured the method was both easy to implement as well as stable in performance, determined by the epsilon parameter.
The method adaptively balanced trying new options to acquire further knowledge and exploiting proven options to achieve the best outcomes [12]. In this study, the MAB framework was used to iteratively select variables that improve clustering performance related to food security metrics. The aim was to balance exploration (trying new variables) and exploitation (reusing variables that had shown strong performance in the past). During the analysis, the cumulative reward function in MAB setting was defined as follows.

$$R_T = \max_{x \in \mathcal{X}} \sum_{t=1}^{T} X_i(t) \tag{3}$$

where:

- R_T: Total reward over
- T is the total number of iterations,
- $X_i(t)$ is the reward received from selecting variable (or "arm") i at time t,
- $\mathcal{X}$ is the set of all possible variables.

During the analysis, each $X_i(t)$ quantified the contribution of variable i to improving the quality of clustering. This contribution was measured using the PAM algorithm with the Silhouette Coefficient as the primary validation metric.

Silhouette Coefficient as a Reward Signal: To compute the reward for variable i, the study performed clustering with and without the inclusion of the mentioned variable, and compared the resulting average silhouette score $\bar{s}$:

$$\text{Reward}_i = \bar{s}_{\text{with } i} - \bar{s}_{\text{without } i} \tag{4}$$

A positive reward showed that the inclusion of variable i improved the cluster compactness and separation, which we interpret. The outcome was interpreted as stronger relevance to distinguishing food security conditions across observations (e.g., households, regions).

Silhouette Coefficient Definition: Given a data point x_j, the silhouette coefficient $s(j)$ was defined as follows.

$$s(j) = \frac{b(j) - a(j)}{\max\{a(j), b(j)\}} \tag{5}$$

where:

- $a(j)$ was the average distance between x_j and all other points in the same cluster,
- $b(j)$ was the minimum average distance between x_j and all points in the nearest different cluster.

The coefficient $s(j)$ ranged from -1 to 1, with higher values showing better-defined clusters. By using the silhouette improvement as a reward, the MAB algorithm was guided to prefer variables that improved the structural clarity of clustering based on food security characteristics such as dietary diversity, food expenditure, water access, and anthropometric measures.

4. Application of Partitioning Around Medoids (PAM)
   The determination of the optimal number of clusters was conducted using the silhouette coefficient method. This method evaluated the quality of clustering by calculating the maximum global silhouette index value for cluster numbers ranging from 2 to a defined maximum. The silhouette coefficient used during the analysis was defined as follows [4].

$$SC = \max_{x} SI(k) \tag{6}$$

   where:
   - SC: Silhouette coefficient
   - $SI(k)$: Global silhouette index for K cluster

   Categories of the silhouette coefficient:
   - $0.7 < SC \leq 1$: Strong cohesion among objects in the cluster
   - $0.5 < SC \leq 0.7$: Moderate cohesion
   - $0.25 < SC \leq 0.5$: Weak cohesion.
   - $SC \leq 0.25$: No cohesion among objects

   The PAM algorithm grouped data based on the distance between data points and the nearest medoids, which served as the representative center for each cluster.
5. Evaluation of results
   Measuring the performance of MAB algorithm in identifying the main factors influencing food security. Measurement was based on secondary data obtained through FSVA website (https://fsva.badanpangan.go.id/). On FSVA, it was stated that 8 independent variables influenced food security in Indonesia. Therefore, a trial was conducted using MAB method in this study. Performance was measured based on probability values and epsilon greedy variables that influenced food security with MAB. Following the process, the clustering results during the analysis from PAM were determined.

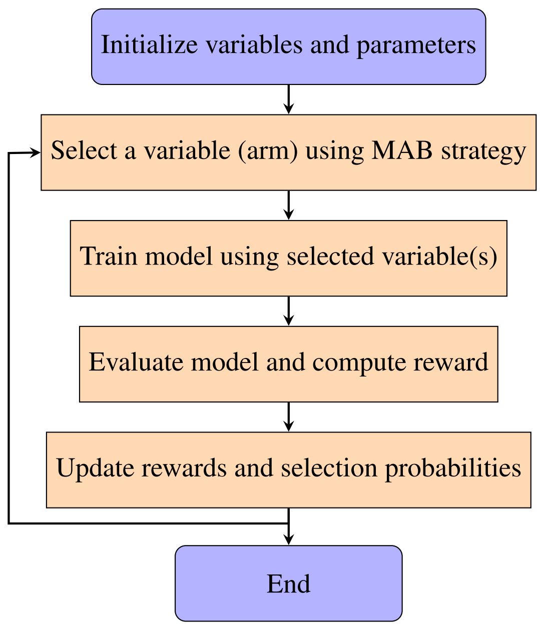


Figure 1. Flowchart to illustrate MAB iteratively selects variables.

3. Results and Discussion

3.1. Data Preprocessing

All variables had different data measurement scales, needed to be transformed before clustering. The transformation process during the analysis was data normalization using z-score. The process was selected because it normalized data, enabling each variable to have the same scale, namely with a mean of zero and a standard deviation of one. By using z-score transformation, the process reduced the effect of outliers that might be present in the raw data and avoided bias in the analysis results.

3.2. Multi Arm-Bandit (MAB)

A flowchart to show how MAB iteratively selects variables is shown in the Figure 1. The application of the MAB method in this study is to examined the variables influencing food security status on Sumatra Island based on the highest accuracy values. Food security status variables were obtained through FSVA by NFA with a total of 8 variables, namely Poverty (P), Food Expenditure (FE), Without Electricity (WE), Without Clean Water (WCW), Women's Education Length (WEL), Health Worker (HWR), Disease Rate (DR), and Stunting (S). During the process, the selection of influential variables was conducted using transformation. The selection of influential variables in the analysis was performed with several possibilities. In this study, the variables were divided into 5 possibilities, namely 0, 0.25, 0.5, 0.75, 1. The use of five possibilities in MAB aimed to cover the range from not influential to very influential in the evaluation of food security variables, allowing for systematic and efficient analysis of variable influences.

A total of 100 trials were conducted because MAB method produced varying outputs, allowing it to be repeated to obtain the highest accuracy value. After running the data 100 times, the accuracy for each variable was obtained.

Table 2. Multi Arm-Bandit accuracy values (%).

Variable	P	FE	WE	WCW	WEL	HWR	DR	S
	93.95	93.43	49.30	49.87	49.19	69.54	71.10	72.20
	93.84	93.32	27.54	29.94	28.85	50.23	73.29	93.74
	93.01	94.05	29.84	29.53	29.11	50.76	71.57	93.48
	93.64	94.16	49.82	27.91	49.50	53.47	71.88	71.26
Accuracy	93.32	92.80	27.91	27.80	49.77	50.76	72.46	73.92
	72.77	94.11	28.43	26.13	28.53	50.03	73.45	94.26
	...	...	...	...	...	...	...	...
	93.43	49.30	49.87	49.19	69.54	71.10	72.20	93.43
	93.32	27.54	29.94	28.85	50.23	73.29	93.74	93.32

The analysis results in Table 2 showed the accuracy results of various variables analyzed using MAB for food security status on Sumatra Island. The Table 2 showed that the variables P, FE, DR, and S had high accuracy, ranging from 70% to 100%, signifying high relevance and influence on food security status of Sumatra Island. Other variables such as WE, WCW, WEL, and HWR signified lower accuracy, ranging from 26% to 70%. Therefore, a total of 4 variables influenced food security, including poverty, food expenditure, disease rate, and stunting. The accuracy for each variable was calculated 100 times to ensure the consistency of the results.

Table 3. Silhouette coefficient values.

No	Number of clusters (k)	Silhouette coefficient
1	2	0.35*
2	3	0.26
3	4	0.26
4	5	0.25
5	6	0.25
6	7	0.25

3.3. Partition Around Medoid (PAM)

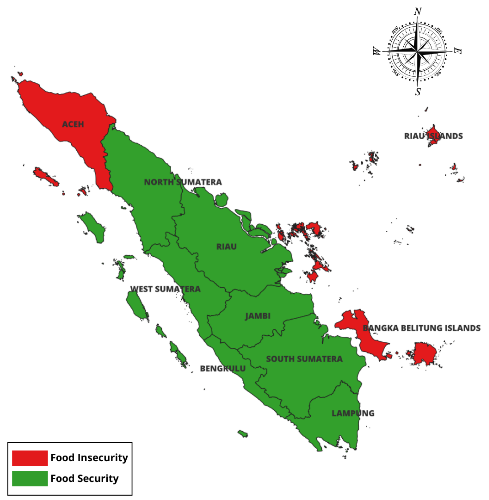
Before applying PAM algorithm, an important step was taken to determine the optimal number of clusters. Determining the number of clusters was important because it influenced the clustering results obtained. In this process, methods such as the silhouette score and gap statistic were used to determine the most suitable number of clusters for the data.

The silhouette score was calculated for different cluster numbers (k) ranging from 2 to 7. Table 3 showed that the highest silhouette coefficient (0.35) was obtained at $k = 2$. This signified that two clusters were the most appropriate for grouping provinces based on food security levels. Therefore, the optimal number of clusters in this method was 2 clusters. The silhouette method was used as a tool to evaluate the quality of clustering by measuring how well the data in each cluster was grouped. For smaller datasets, the silhouette value generally implied that using a fewer number of clusters, such as two clusters, provided more consistent and interpretable clustering results. This was because dividing into many clusters could lead to overfitting with limited data, where the resulting clusters were less representative of the actual data.

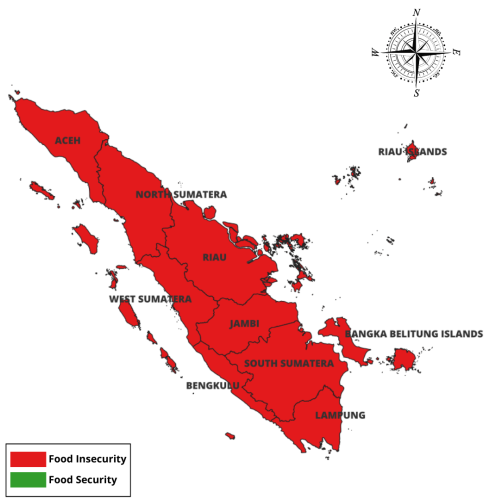
Table 4. Clusters based on provinces.

No	Region	Poverty	Food Expenditure	Disease Rate	Stunting
1	Aceh	Vulnerable	Vulnerable	Vulnerable	Vulnerable
2	North Sumatra	Vulnerable	Secure	Vulnerable	Secure
3	West Sumatra	Vulnerable	Secure	Vulnerable	Vulnerable
4	Riau	Vulnerable	Secure	Vulnerable	Vulnerable
5	Riau Islands	Vulnerable	Vulnerable	Vulnerable	Vulnerable
6	Jambi	Vulnerable	Secure	Vulnerable	Vulnerable
7	South Sumatra	Secure	Secure	Vulnerable	Vulnerable
8	Bengkulu	Vulnerable	Secure	Vulnerable	Vulnerable
9	Lampung	Vulnerable	Secure	Vulnerable	Vulnerable
10	Bangka Belitung Islands	Vulnerable	Vulnerable	Vulnerable	Vulnerable

(a) Poverty



(b) Food Expenditure



(c) Disease Rate

(d) Stunting

Figure 2. Map of food security cluster distribution based on variables.

3.4. Cluster Result Interpretation

Based on the cluster results from PAM method, the characteristics of each formed cluster were observed. The data used in this study included 10 provinces on Sumatra Island. Table 4 showed the values of each variable and the categories assigned.

Table 4 showed the condition of several provinces on Sumatra Island based on four main variables, namely poverty, food expenditure, disease rate, and stunting. These provinces were grouped based on the vulnerability level for each variable. Aceh was categorized as vulnerable for all variables, signifying that this province had high levels of poverty, food expenditure, disease rate, and stunting. North Sumatra was also vulnerable for most variables except for food expenditure and stunting, which were more controlled with a secure category. West Sumatra showed vulnerability to poverty and disease rate, and was vulnerable to stunting, but was more secure in terms of food expenditure.

The results shows that Aceh, Riau Islands, and Bangka Belitung Islands are the most vulnerable in all categories, while North Sumatra and South Sumatra signify greater resilience. This clustering analysis provides crucial perceptions for policy makers to implement targeted food security interventions in high-risk provinces.

Riau Province has better conditions in terms of food expenditure but remains vulnerable to other variables. Jambi and Lampung show similar patterns of vulnerability, with some variables being better than others province. For example, the Riau Islands are vulnerable to all variables, while Jambi as well as Lampung show security in food expenditure except poverty, disease rate, and stunting. South Sumatra shows security in poverty and food expenditure but is vulnerable to disease rates as well as stunting. Following the discussion, Bengkulu shows security in food expenditure except for poverty, disease rate, and stunting. During the process, the Bangka Belitung Islands also show vulnerability to all variables. This analysis helps identify provinces that require more attention in efforts to address health and economic issues.

Figure 2 was a thematic map showing the cluster results for each province in Sumatra Island, aiming to explain the distribution of cluster characteristics in the context of food security in the region. Based on the Figure 2, the map showed the distribution of food security clusters based on four different variables, namely poverty, food expenditure, disease rate, and stunting. These maps provided a visual representation of the level of vulnerability (vulnerable) and security (secure) of food in various provinces on Sumatra Island. In Figure 2a, South Sumatra Province was in the secure category, while other provinces such as Aceh, North Sumatra, West Sumatra, Riau, Jambi, Bengkulu, Lampung, Riau Islands, and Bangka Belitung Islands were in the vulnerable group. This implies that high poverty levels significantly influence food security vulnerability in a province.

In Figure 2b, the majority of provinces, such as Aceh, Riau Islands, and Bangka Belitung Islands, are in the vulnerable category. Provinces such as North Sumatra, West Sumatra, Riau, Jambi, South Sumatra, Bengkulu, and Lampung are in the secure category. Moreover, high food expenditure in some provinces signifies problems in distribution or adequate food accessibility. Figure 2c shows that all provinces are in the vulnerable category as high disease rates affect all provinces in Sumatra in terms of food security.

Figure 2d shows that provinces such as Aceh, West Sumatra, Riau, Jambi, South Sumatra, Bengkulu, Lampung, Riau Islands, and Bangka Belitung Islands are in the vulnerable category, while North Sumatra is in the secure category. High stunting rates in some provinces imply serious nutritional problems that can affect food security.

Figure 2 shows that provinces more vulnerable to food security have varying problems depending on the analyzed variable. Provinces such as Aceh, Riau Islands, and Bangka Belitung Islands are consistently vulnerable across several variables, while provinces such as North Sumatra and South Sumatra tend to be more secure. This signifies that policy interventions need to be modified to the specific needs of each province to improve total food security.

4. Concluding Remarks

1. In conclusion, the results of clustering food security status in Sumatra Island in 2022 using MAB and PAM methods show the following.

- The Multi-Arm Bandit (MAB) method identified four key variables affecting food security: namely Poverty, Food Expenditure, Disease Rate, and Stunting. These variables showed accuracy levels ranging from 70% to 100% during the analysis.
- The clustering results from Partition Around Medoid (PAM) divided the provinces into two groups, namely vulnerable and secure regions, based on food security status.

2. The results show that provinces such as Aceh, Riau Islands, and Bangka Belitung Islands are the most vulnerable in terms of food security, requiring targeted intervention from policy makers.

Declarations

- Author Contributions: All authors contributed equally as primary contributors to this paper. All authors read and approved the final manuscript.
- Funding Statement: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.
- Conflict of Interest: The authors declare no conflicts of interest.
- Data availability: The dataset used in this study was retrieved from the National Food Agency Indonesia and is accessible at the following link: https://fsva.badanpangan.go.id/. The code utilized for this study can be accessed on GitLab at the following repository: https://gitlab.com/subianto/food-sumatra.

REFERENCES

1. Stuart J. R. and Peter N., *Artificial Intelligence: A Modern Approach*, Pearson Education, 4th edition, ISBN 978-0-1346-1099-3, 2020.
2. Aurelien G., *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media, Inc., Sebastopol, CA, 2nd edition, ISBN:978-1-4920-3264-9, 2019.
3. Charu C. A. and Chandan K. R. *Data Clustering: Algorithms and Applications*, CRC Press, 1st edition, ISBN:978-1-4665-5821-2, 2013.
4. Leonard K. and Peter J. R., *Finding Groups in Data: An Introduction to Cluster Analysis*, Wiley Series in Probability and Statistics. Wiley, https://doi.org/10.1002/9780470316801, 2005.
5. NFA, *Food Security and Vulnerability Atlas – FSVA*, The National Food Agency (NFA), Indonesia, https://fsva.badanpangan.go.id/, 2022.
6. Emalia, Z., Budiarty, I., and Asnawa, M. J., Environmental quality in sumatra island in economic and spatial perspective, In *Proceedings of the International Conference of Economics, Business, and Entrepreneur (ICEBE 2022)*, pages 369–377. Atlantis Press, https://doi.org/10.2991/978-2-38476-064-0_39, 2023.
7. Jalaludin MY, Fauzi MD, Sidiartha IGL, John C, Aviella S, Novery E, Permatasari A, and Muhardi L, Addressing Stunting in Children Under Five: Insights and Opportunities from Nepal, Bangladesh, and Vietnam-A Review of Literature, *Children*, 12(5):641. https://doi.org/10.3390/children12050641, 2025.
8. Prayitno, G., Auliah, A., Zuhriyah, L., Efendi, A., Arifin, S., Rahmawati, R., Nugraha, A. T., and Siankwilimba, E., Exploring the Role of Food Security in Stunting Prevention Efforts in the Bondowoso Community, Indonesia, *Societies*, 15(5):135. https://doi.org/10.3390/soc15050135, 2025.
9. Volkan C. and Oktay Y. A comprehensive review on data preprocessing techniques in data analysis, *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, 28(2):299–312, https://dergipark.org.tr/tr/pub/pajes/issue/69632/1110427, 2022.
10. Joannes V. and Mehryar M., Multi-armed bandit algorithms and empirical evaluation, In *Proceedings of the 16th European Conference on Machine Learning (ECML 2005)*, pages 437–448. Springer, https://doi.org/10.1007/11564096_42, 2005.
11. Alexandre M. and Ricado L.A.R., An adaptive implementation of ϵ-greedy in reinforcement learning. *Procedia Computer Science*, 109:1146–1151, https://doi.org/10.1016/j.procs.2017.05.431, 2017.
12. Tor L. and Csaba S., *Bandit Algorithms*, Cambridge University Press, https://doi.org/10.1017/9781108571401, 2020