Indonesian Journal of Business Analytics (IJBA)
Vol. 5 , No. 6 (December) 2025: 3896-3916

Transformation of Food Security Analysis: A Data Science Approach to Agricultural Commodities in Ntt

Rolland E. Fanggidae[1], Rikhard T. Ch. Bolang[2], Aldarine Molidya[3], Edy Purwanto[4], Ginanjar Aji Nugroho[5], Bonifasius Irmgardus Nggajo[6]
Nusa Cendana University, Directorate General of Treasury
Corresponding Author: Aldarine Molidya: aldarine.molidya@staf.undana.ac.id

A R T I C L E I N F O

Keywords: Food Security, Forecasting, Clustering.

Received : 20, October
Revised : 22, November
Accepted: 24, December

©2025 Fanggidae, Bolang, Molidya, Purwanto, Nugroho, Nggajo (s): This is an open-access article distributed under the terms of the Creative Commons Atribusi 4.0 Internasional.

A B S T R A C T

This study aims to assess food commodity data forecasting in East Nusa Tenggara (NTT) Province using the ARIMA model and to conduct a clustering analysis of food security status using the K-Means Clustering and Hierarchical Agglomerative Clustering (HAC) approaches. The analytical method used was the Autoregressive Integrated Moving Average (ARIMA) model to calculate the forecast of time series food commodity data in East Nusa Tenggara (NTT) Province, and the food security analysis used K-Mean Clustering and Hierarchical Agglomerative Clustering (HAC). The ARIMA test results produced forecasts of food commodity data or variables for 2025 to 2027. However, these results did not yield satisfactory results when evaluated using three different coefficients, one of which was the Mean Absolute Percentage Error (MAPE), which produced a relatively large value. This concludes that the test results for forecasting data or food commodity variables in NTT are inaccurate, requiring additional factors capable of producing good predictions. The importance of food security policies that adapt to regional capacity. Cluster 1 focuses on increasing production and stabilizing prices through simple irrigation, agricultural assistance, and market operations. Cluster 3 focuses on strengthening logistics through road infrastructure, distribution hubs, and digitalization. Across clusters, a simple food information system, interregional partnerships, and local climate adaptation are needed to make policies more realistic and measurable.

DOI: https://doi.org/10.55927/ijba.v5i6.15921
ISSN: 2808-0718
https://journal.formosapublisher.org/index.php/ijba

INTRODUCTION

Specifically, food vulnerability hotspots worldwide are found in several regions in sub-Saharan Africa, South Asia, and Indonesia (Riptanti et al., 2020); (Lal, 2013); (Bruinsma, 2017), and one region in particular that remains vulnerable to food insecurity is East Nusa Tenggara Province (Riptanti et al., 2018). Based on data from the National Development Planning Agency (Bapanas, 2025), the average food security index for districts/cities in East Nusa Tenggara Province ranges from "somewhat food vulnerable" to "food secure." This means that overall, districts/cities in NTT Province are not yet fully secure, necessitating priority programs to boost food security in the province. This is influenced by climate factors and drought, primarily caused by the short rainfall intensity in East Nusa Tenggara, which causes this region to be categorized as an agricultural dryland region (Nursyamsi et al., 2014). This category influences the implementation system of the new national program initiated by the Indonesian government, the Free Nutritious Meals (MBG).

The Free Nutritious Meals (MBG) program implemented by the government is being implemented throughout the Republic of Indonesia, including the East Nusa Tenggara Province. The MBG program's primary targets are children, students, and pregnant women. This is based on data from the Ministry of Health and the Coordinating Ministry for Human Development and Culture, which shows that 41% of students experience hunger, which impacts the quality of education (Kiftiyah et al., 2025); (Merlinda & Yusuf, 2025). However, the government is potentially faced with a situation where national and local food supplies cannot meet demand due to the implementation of the MBG program (Dwijayanti, 2024), which impacts the dynamics of the agricultural sector (Moffitt et al., 2023); (Orsini & Smith, 2007). Therefore, studies are needed to develop preventive measures for the implementation of the MBG program that could impact national and local food security.

One study that needs to be conducted is forecasting local food availability at the regional level, specifically in East Nusa Tenggara (NTT) Province. Research by Da Veiga et al., 2014, shows that food availability analysis can be conducted using the Autoregressive Integrated Moving Average (ARIMA) model and the Holt-Winters (HW) method. Similar analyses have been conducted in studies by Thapa et al., 2022; Ilić et al., 2016; and Nath et al., 2019. This study is necessary to predict regional food availability in terms of meeting demand resulting from the implementation of the MBG program in East Nusa Tenggara (NTT) Province.

National and local food availability are theoretically correlated with existing food security status (Béné, 2020). Implementationally, food security status in Indonesia is classified by the National Food Agency (Bapanas, 2025). However, this food security analysis does not include clustering factors based on regional similarities at a specific regional level. Therefore, studies that incorporate regional or spatial factors into the existing food security status are needed. Research conducted by (Sinaga et al., 2025) classifies food security status using two different approaches: K-Means Clustering and Hierarchical Agglomerative Clustering (HAC) methods. K-Means is used to group data into k clusters based on the distance to the nearest centroid, then update these centroids until the maximum number of iterations is reached (Ay et al., 2023); (Oti et al., 2021). HAC, on the other hand, operates with a bottom-up approach, meaning each piece of data is considered a separate cluster and is gradually combined based on similarities to form the final cluster (Chhabra & Mohapatra, 2022); (Yu & Hou, 2022).

Research related to food security has been widely discussed in previous studies (Riptanti et al., 2018); (Nursyamsi et al., 2014); (Asmara et al., 2012). However, analytically, there is less discussion of food clustering analyzed using a data science or machine learning-based approach. The integration of regional food commodity data and the use of machine learning are utilized in this study to produce comprehensive analysis results. Therefore, the problem formulation in this study is: first, how is food data forecasting on food commodities in East Nusa Tenggara Province? Second, how is the clustering of food security status using food commodity data in East Nusa Tenggara Province? Therefore, the purpose of this study is intended to answer this research question, namely to determine the forecasting of food commodity data in East Nusa Tenggara Province (NTT) using the ARIMA model and conducting.

THEORETICAL REVIEW

Food

Food is a fundamental human need, and therefore, its fulfillment is a fundamental right of every citizen, as stipulated in Law No. 18 of 2012. Food availability refers to the availability of food from domestic production and national food reserves, as well as imports if these two primary sources are unable to meet demand (Government Regulation No. 17 of 2015). Food availability refers to the quantity of food available in a region, encompassing production, imports/exports, seeds/seedlings, raw materials for the food industry, raw materials for the food and non-food industries, depreciation, waste, and availability for consumption. Food distribution refers to any activity or series of activities involved in distributing food or food products to the public, whether for trade or not (Yuwono et al., 2019).

Food availability can essentially be met through domestic production or imports. However, for Indonesia, there is arguably no better option than producing food independently through the utilization of existing resources nationally. Paddy field resources have an important role in producing food, around 90% of national rice is produced from paddy fields and the rest from dry land (Asmara et al., 2012).

Food Security and Local Food

Food security, according to Republic of Indonesia Regulation Number 17 of 2015, is defined as the state's ability to meet nutritional needs for its citizens, as reflected in the availability of adequate food, both in quantity and quality, that is protected, diverse, nutritious, fair, and reasonable, and does not conflict with religion, beliefs, and local culture. Food security is defined as the condition in which food is available to the community, down to the individual level (Nugroho & Mutisari, 2015). Food security is assessed by the availability of sufficient food, both in quantity and quality, that is safe, diverse, equitable, nutritious, and affordable (Asmara et al., 2012). Food security is a condition in which the population can meet its food needs (Asmara et al., 2012). According to the Ministry of Agriculture's Food Security Agency (2020), there are three aspects that influence the level of food security:

1. Food availability refers to the availability of food from domestic production, food reserves, and food imports if these two primary sources are insufficient to meet demand. Food availability can be calculated at the national, regional, district/city, and community levels.
2. Food accessibility refers to the ability of households to obtain sufficient nutritious food through one or a combination of various sources, such as: own production and supply, barter, purchasing, loans, gifts, and food aid. Food availability in an area exists, but it cannot be accessed by certain households if they lack the physical, economic, or social capabilities to access sufficient food diversity and quantity.
3. Food utilization refers to the use of food by households and the individual's ability to metabolize and absorb nutrients. Food utilization includes food processing, storage, and preparation methods, water safety for cooking and drinking, feeding habits, hygiene conditions, and food distribution within the household. The mother's significant role in improving the nutritional health of the family, especially infants and children, is also often used as a variable to measure household food utilization.

According to (Nugroho & Mutisari, 2015), there are 200 definitions and 450 indicators concerning Food is a strategically important commodity, considering that it is a necessity for humans to have at all times, in sufficient quantities, of adequate quality, and safe for consumption at affordable prices (Utami & Budiningsih, 2015).

Local food encompasses all aspects of food production and development, tailored to the potential and resources of the local region and culture. Therefore, the type, quantity, and quality of local food products depend on the environmental conditions of the region. These conditions are not only focused on land suitability, climate, soil, or other cultural factors that may influence it, but also relate to the social, economic, and cultural conditions of the community within the region. A diverse range of local food crops are found throughout Indonesia, including corn, cassava, ganyong (canna), gembili (peanut), and others. Local foods also possess advantages in various aspects, including quality and quantity, and also contribute to biodiversity and ecosystem sustainability (Utami & Budiningsih, 2015).

Indonesia is the third-largest country in the world in terms of diverse natural resources (megadiversity), so it is natural that Indonesian food resources are rich (Muhammad Farhan & Hani Nurul Fadhilah, 2023). Local foods must be resistant to pests and diseases, have minimal dependence on pesticides, herbicides, and fungicides, and minimal reliance on chemical fertilizers. This means they have greater health value, high calorie and other nutritional content, and minimal chemical contamination (Utami & Budiningsih, 2015).

METHODOLOGY

This research was conducted in East Nusa Tenggara Province regarding local food availability and security. This research analysis focused on several food commodities, including rice, spring onions, shallots, garlic, spinach, green beans, chilies, long beans, water spinach, cauliflower, potatoes, cucumbers, cabbage, chayote, Chinese cabbage, eggplant, tomatoes, carrots, melons, watermelons, fisheries, corn, seaweed, green beans, cassava, peanuts, sweet potatoes, and soybeans. The data used in this study is secondary data. In this study, annual data was used, namely from 2013-2023.

ARIMA (Autoregressive Integrated Moving Average)

The Autoregressive Integrated Moving Average (ARIMA) model completely ignores the independent variable in forecasting. ARIMA uses the past and present values of the dependent variable to generate accurate short-term forecasts.

Box and Jenkins introduced the ARIMA model in 1970. This model, also known as the Box-Jenkins methodology, consists of a series of steps to identify, estimate, and diagnose ARIMA models using time series data. This model is the most prominent method in financial forecasting (Pai & Lin, 2005; Nochai & Nochai, 2006). In the ARIMA model, the future value of a variable is a linear combination of its past values and past errors, which is expressed as follows:

Y_t=∅_0+ ∅_1 Y_(t-1)+ ∅_2 Y_(t-2)+⋯+ ∅_p Y_(t-p)+ ε_t- θ_1 ε_(t-1)-…- θ_q ε_(t-q)……….. (1)

Yt is the actual value, εt is the random error in period t, $\emptyset 1$ and $\theta 1$ are the variable coefficients, and p and q are autoregressive and moving averages. The basic nature of ARIMA is that the data is stationary.

$\Delta Y_t = Y_t - Y_{t-1}$..................(2)

$\Delta^2 Y_t = \Delta Y_t - \Delta Y_{t-1}$....................(3)

With a series, it is possible to return to the sum in total d times. This can be written as $Yt = \Sigma dwt$, where Σ is the total value of:

$\Sigma^2 w_t = \Sigma_{i=-\infty}^{t} w_i$............(4)

$\Sigma^2 w_t = \Sigma_{i=-\infty}^{j} w_i$---............(5)

It should be noted that the sum total Σ is the inverse of the difference operator D. Since $\Delta Yt = Yt - Yt{-}1$, it can be written as $\Delta = 1 - B$ and $\Sigma = \Delta{-}1 = (1 - B){-}1$.

K-Means Clustering

The K-Means algorithm is an iterative clustering algorithm that divides a dataset into k predetermined clusters. The K-Means Clustering algorithm is presented below (Sinaga et al., 2025).

1. Determine the desired number of clusters (k) in the dataset.
2. Determine the initial cluster centers (centroids) by taking the smallest, average, and largest values.
3. Calculate the shortest distance between each data point and the centroid. Calculate the shortest distance to the centroid using the Euclidean distance formula. The formula can be seen below:

   $d\ (xi, \mu j) = \sqrt{(Xi - \mu j)^2}$.............(6)

Description:
xi: Criteria data
μj: Cluster center point j

4. Recalculate the cluster centers using the current cluster members. The formula can be seen below:
   $\mu j(t + i) = \frac{1}{N_{sj}} \sum j \in sjxj$.............(7)
   Note:
   $\mu j(t + 1)$: New centroid at the 1st iteration
   Nsj: Number of data in cluster sj

Hierarchical Agglomerative Clustering (HAC)
Hierarchical Agglomerative Clustering is a clustering method that builds a data hierarchy using a bottom-up approach, namely by combining data points one by one to form a single large cluster. The HAC algorithm is presented as follows (Chhabra & Mohapatra, 2022); (Sinaga et al., 2025).

1. Calculate the Euclidean distance matrix (as in Formula 1).
2. Merge the two closest clusters. If the distance between objects a and b is the smallest compared to the distances between other objects in the Euclidean distance matrix, then the two clusters merged in the first stage are d_ab.
3. Update the distance matrix according to the Agglomerative clustering technique. If d(ab) is the closest distance from the Euclidean distance matrix, then the formula for the agglomerative method is:
   a. Single bond formula
      $d_{(ab)c} = min\,\{d_{a,c};\ d_{b,c}\}$................(8)
   b. Average relationship formula
      $d_{(ab)c} = average\,\{d_{a,c};\ d_{b,c}\}$...........(9)
   c. Complete relationship formula
      $d_{(ab)c} = max\,\{d_{a,c};\ d_{b,c}\}$.............(10)
4. Repeat steps 2 and 3 until only one cluster remains.
5. Draw a Dendrogram

RESULTS
The analysis was conducted using two different calculation approaches. For food commodity data forecasting, the Autoregressive Integrated Moving Average (ARIMA) Model was used in the first calculation stage. Meanwhile, for local food security analysis, two different clustering methods were used: K-Mean Clustering and Hierarchical Agglomerative Clustering (HAC). The following are the analysis results for each calculation stage using the two calculation approaches used.

1. Data Forecasting Analysis Using the ARIMA Model

Next, food production forecasting will be conducted to determine the amount of food production that can be produced in the forecast year. Food commodity forecasting is tested using the ARIMA (Autoregressive Integrated

Moving Average) method. The ARIMA method has several stages that must be completed before forecasting. The following are the ARIMA testing stages, with the first stage being the Unit Root Test using the Augmented Dickey Fuller (ADF).

Table 4.1 Augmented Dickey Fuller Unit Root Tests

Variabel	**Stationary Data**			**Variabel**	**Stationary Data**	
	Level	**Difference**			**Level**	**Difference**
Bawang Daun	-0.848	-3.669		Terung	-1.116	-3.546
	(0.8048)	(0.0046) ***			(0.7088.)	(0.0069) ***
Bawang Merah	-1.065	-4.676		Tomat	-0.802	-5.350
	(0.7288)	(0.0001) ***			(0.8186)	(0.0000) ***
Bawang Putih	-1.476	-2.623		Wortel	-1.860	-4.090
	(0.5454)	(0.0883) *			(0.3510)	(0.0010) ***
Bayam	0.044	-3.915		Melon	-2.731	
	(0.9620)	(0.0019) ***			(0.0689) *	
Buncis	-1.539	-5.396		Semangka	-1.692	-3.217
	(0.5143)	(0.0000) ***			(0.4351)	(0.0190) **
Cabai	-1.598	-6.736		Perikanan	-2.516	-4.062
	(0.4846)	(0.0000) ***			(0.1116)	(0.0011)
Kacang Panjang	-0.751	-3.058		Jagung	-1.183	-2.609
	(0.8331)	(0.0298) **			(0.6807)	(0.0912) *
Kangkung	-0.604	-3.812		Rumput Laut	-2.226	-4.875
	(0.8700)	(0.0028) ***			(0.1970)	(0.0000) ***
Kembang Kol	-1.341	-2.989		Padi	-2.357	-2.579
	(0.6104)	(0.0360) **			(0.1543)	(0.0975) *
Kentang	-3.439	-5.895		Kacang Hijau	-2.811	
	(0.0097)	(0.0000) ***			(0.0567) *	
Ketimun	-1.828	-2.826		Ubi Kayu	-2.211	-5.084
	(0.3668)	(0.0547) *			(0.2024)	(0.0000) ***
Kubis	-1.774	-3.427		Kacang Tanah	-2.740	
	(0.3936)	(0.0101) **			(0.0674) *	
Labu Siam	-1.934	-3.063		Ubi Jalar	-1.448	-2.806
	(0.3161)	(0.0294) **			(0.5592)	(0.0574) *
Petsai	-0.999	-3.256		Kedelai	-2.268	-3.888
	(0.7538)	(0.0170) **			(0.1826)	(0.0021) ***
***, **, * indicates 1%, 5% and 10% significance level. The critical values of ADF tests are with intercept at 1%, 5% and 10% are -3.750, -3.000 and -2.630. Probability values are based on MacKinnon approximate p-value for Z(t)						

The results of the data stationarity test indicate that there are several food commodity variables that are directly stationary at the level degree, namely Melon, Green Beans, and Peanuts. Meanwhile, for other food commodity variables, they are stationary at the first and second differences. If all commodities show stationary at the level degree, the method used is ARMA (Autoregressive Moving Average). However, because overall, many food commodity variables are stationary at the first and second differences, testing of data forecasting is carried out using ARIMA (Autoregressive Integrated Moving

Average). The next step is to select the order (p, d, q) using the ACF (Autocorrelation Function) and PACF (Partial Autocorrelation Function) tests at the specified lag.

Table 4.2 Estimasi Model ARIMA

Variabel	ARIMA (p, d, q) Model	Prob > chi2	MAPE	AIC Value	BIC Value	Forecast Value		
						2025	2026	2027
Bawang Daun	1,1,1	0.9449	91.43	198.62	199.83	17944	19690	21437
Bawang Merah	1,1,1	0.1443	97.76	217.50	218.10	162055	171963	181842
Bawang Putih	0,1,1	0.0264	104.56	168.92	169.31	3168	2897	2625
Bayam	1,1,1	0.8900	98.54	185.12	185.91	58838	56679	57904
Buncis	1,1,2	0.0000	101.52	185.13	186.11	43097	53024	42371
Cabai	1,1,2	0.0010	112.78	224.27	225.25	201162	-100711	177782
Kacang Panjang	1,1,1	0.9742	97.18	203.36	204.15	67121	66372	67161
Kangkung	1,1,2	0.3071	98.53	209.89	210.68	147503	144341	146450
Kembang Kol	1,1,3	0.0014	156.44	168.43	169.61	5696	5104	5009
Kentang	1,1,2	0.0002	156.57	182.33	183.32	5124	3620	4862
Ketimun	1,1,1	0.8843	106.88	199.28	200.07	49027	46151	45018
Kubis	1,1,2	0.0000	104.22	198.90	199.68	26568	24812	23150
Labu Siam	1,1,2	0.6963	103.06	232.43	233.42	107015	108123	99530
Petsai	1,1,2	0.3670	101.60	210.52	211.50	128769	136101	129677
Terung	0,1,1	0.1772	103.44	212.40	212.79	104412	103306	102200
Tomat	1,1,1	0.0015	100.72	206.17	206.76	124702	135968	129629
Wortel	1,1,2	0.0935	108.95	188.85	189.64	26671	25572	25252
Semangka	1,1,2	0.2455	333.06	208.16	208.95	3274	2743	3406
Melon	1,1,2	0.0069	362.52	189.84	190.75	15169	14839	14743
Perikanan	1,1,1	0.6564	100.57	223.77	224.56	175021	176542	172714

Jagung	1,1,2	0.0000	99.79	236.16	237.15	521042	603482	516539
Rumput Laut	1,1,1	0.2791	103.23	233.46	233.78	1973806	1817497	1853787
Padi	0,1,0	0,0000	100.54	208.55	208.71	729690	723299	716909
Kacang Hijau	1,1,1	0.0153	102.53	172.88	173.67	7914	7462	7749
Ubi Kayu	1,1,2	0.0000	109.15	252.48	253.46	634540	893946	630120
Kacang Tanah	1,1,1	0.9765	110.14	191.16	192.37	5623	5290	4853
Ubi Jalar	1,1,1	0.9910	97.90	204.26	205.05	22791	24273	25099
Kedelai	0,1,1	0.9994	210.89	200.97	201.56	2426	2326	2226

Source : Local Expert and Undana Research Team Analysis (2025)

The results of the ARIMA analysis used to forecast data produce deviations based on the MAPE coefficient value. A large MAPE value indicates that the forecasting method's forecasting accuracy is small. This means that the smaller the percentage error value in the MAPE, the more accurate the forecasting results (Ilić et al., 2016); (Nath et al., 2019). Therefore, alternative forecasting calculations are needed to minimize the deviations in the values that can be produced. The recommended analysis method is the Exponential Smoothing analysis method. Exponential Smoothing analysis is a fixed-model time series prediction method. Initially referred to as an "exponentially weighted moving average," Exponential Smoothing is a prediction scheme that uses weighted values from previous series observations to calculate future values (Shastri et al., 2018); (Shastri et al., 2015). However, this method requires long time series data, so the use of data in subsequent analyses that can be recommended is relatively long time series data, namely over 15 years, to produce small data variability and not cause large value deviations.

2. **Analysis of the Food Security Index (FSI) using K-Mean Clustering and Hierarchical Agglomerative Clustering**

Clustering Using K-Means Clustering

The stages begin with determining the number of clusters for the K-Means Clustering method using the Elbow Method and using the Silhouette Score to assess the optimal cluster score based on the Elbow Method. The results from both graphs presented show that the optimal clusters used are 3 clusters for further use in the K-Means Clustering method.

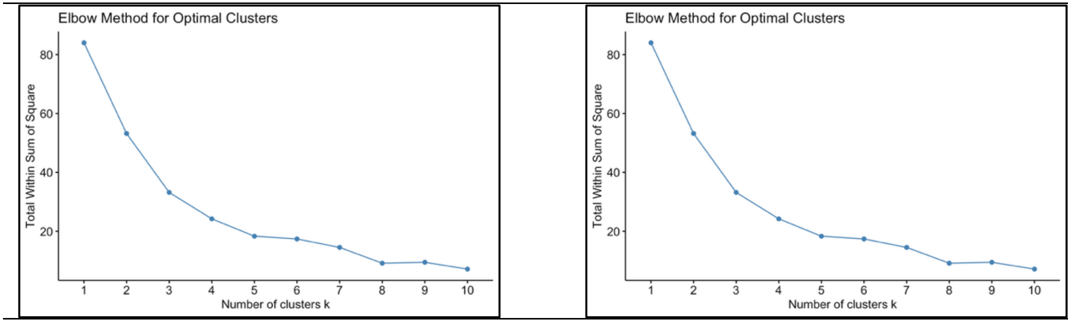

Figure 4.3 Metode Elbow dan Silhouette Score

The Elbow Method graph above shows the relationship between the number of clusters (k) and the Total Within Sum of Squares (WSS) value, or within-cluster variance. When the number of clusters is small, the WSS value decreases sharply because dividing the data into several initial groups significantly reduces internal variation. However, after reaching a certain point, the decline in the WSS value begins to level off even as the number of clusters continues to increase (Sinaga et al., 2025); (de Oliveira et al., 2024). In this graph, the elbow point is seen at k = 3 or k = 4. This means that adding clusters after this point does not significantly reduce variance. Therefore, the optimal number of clusters for this data is 3 or 4. With this number, clustering is considered quite efficient.

The graph below shows the results of dimensionality reduction into two different dimensions: dimension 1 (Dim 1) and dimension 2 (Dim 2) based on the K-Means method using Multidimensional Scaling (MDS). The results in the graph above show the division of regional clusters based on the analysis attributes used. These clusters follow the clustering pattern from the K-Means calculation in the clustering results table using K-Means Clustering. The following is an interpretation of the cluster visualization in the graph above.

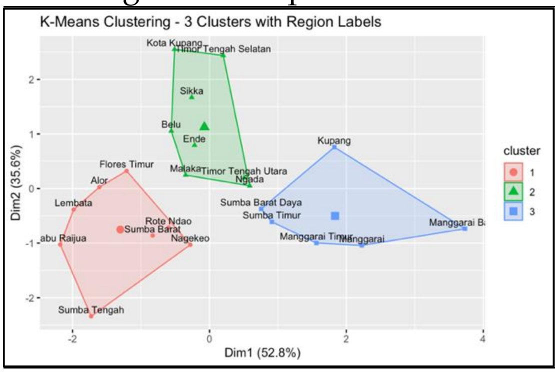

Figure 4.4 Cluster Plot Visualization Using K-Mean Clustering

This stage begins with cluster determination for the K-Means method using a plot visualization based on three clusters, as shown in the image above. The image above also shows the distribution of clusters, consisting of Dimension 1 (52.8%) and Dimension 2 (35.6%). The clustering visualization results in the image above use the recommended number of clusters from the elbow analysis and silhouette score method.

Table 4.4 Clustering Results with K-Means Clustering

No.	Produksi Beras	Luas Panen Padi	Jumlah Penduduk	Harga Beras	Daerah	Cluster
1.	-0.3068996	-0.2098478	-0.94930779	0.6788107	Sumba Barat	1
2.	0.6019048	0.9294336	0.02045521	0.6408775	Sumba Timur	3
3.	1.0780389	0.7701124	1.14456208	0.5377771	Kupang	3
4.	-0.5950484	-0.6779432	2.08457204	0.5204316	Timor Tengah Selatan	2
5.	0.1740042	0.3527897	0.17015017	0.5862473	Timor Tengah Utara	2
6.	-0.7110710	-0.6227140	-0.19897342	0.5167031	Belu	2
7.	-1.0133500	-1.05579372	-0.29846014	-1.5906952	Alor	1
8.	-1.0972965	-1.17994006	-1.05809232	-1.5906952	Lembata	1
9.	-0.9434144	-0.81752824	0.32914295	-1.5906952	Flores Timur	1
10.	-0.6796507	-0.67507828	0.77357892	0.5167031	Sikka	2
11.	-0.3426723	-0.41723588	0.22593711	0.5642006	Ende	2
12.	0.4879093	0.27066217	-0.77171936	0.5167031	Ngada	2
13.	1.7930151	1.65473470	0.71965154	-1.5906952	Manggarai	3
14.	-0.1836628	-0.11928466	-0.96790343	0.6484966	Rote Ndao	1
15.	2.6262199	2.59840604	0.17851821	0.5167031	Manggarai Barat	3
16.	-0.3614529	-0.40736784	-1.53135154	0.8409182	Sumba Tengah	1
17.	0.2746446	0.88868812	0.66665395	-1.5906952	Sumba Barat Daya	3
18.	0.2617068	0.02412027	-0.82471695	0.6788107	Nagekeo	1
19.	1.3515208	1.23152301	0.36354490	0.6788107	Manggarai Timur	3

20.	-1.0261089	-1.10545225	-1.50903676	-1.5906952	Sabu Raijua	1
21.	-0.3136964	-0.25568648	-0.59134158	0.5609585	Malaka	2
22.	-1.0746405	-1.17659766	2.02413619	0.5410192	Kota Kupang	2

Source: Local Expert and Undana Research Team Analysis (2025)

Cluster 1 includes the regions of Central Sumba, Sabu Raijua, Lembata, Alor, East Flores, West Sumba, Rote Ndao, and Nagekeo. These regions tend to be concentrated on the negative side of Dimension 1 (52.8%) and Dimension 2 (35.6%), which together explain most of the variation in the data. This pattern demonstrates common characteristics of low rice production, small harvested rice areas, and a relatively small population. Furthermore, rice prices in this region tend to be higher than in other clusters, indicating potential supply pressure on local demand. In general, this cluster represents regions with a weak food production base and limited harvest scale.

Cluster 2 includes Kupang City, South Central Timor, North Central Timor, Sikka, Belu, Ende, Malaka, and Ngada. These regions are located in the upper right quadrant of the cluster map, indicating larger harvested areas, higher food production, and relatively dense populations compared to the first cluster. These characteristics indicate a balance between food availability and demand, although production capacity is still not as optimal as in the next cluster. Regions within this cluster have the potential to strengthen food security through increased productivity and distribution efficiency.

Cluster 3 consists of Kupang, East Manggarai, West Manggarai, Manggarai, Southwest Sumba, and East Sumba. These regions are located in the upper right corner of the analysis graph, indicating the highest levels of rice production and harvested area among all clusters. Furthermore, this region has a large population and more stable rice prices. These characteristics indicate that the regions within the blue cluster serve as major food production centers in NTT Province, contributing significantly to regional food security.

Clustering Using Hierarchical Agglomerative Clustering

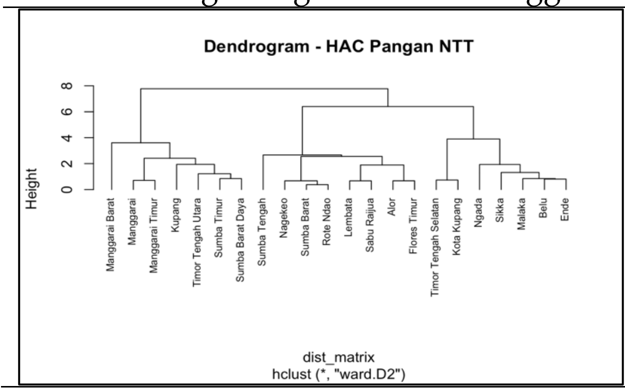

Figure 4.5 Visualization of Hierarchical Agglomerative Clustering Dendogram

The steps begin with determining the number of clusters for the HAC method using the linkage (single) method, which produces a dendrogram as shown in the figure above. The dendrogram is evaluated using the Cophenetic

Correlation Coefficient (CCC) to assess the optimal cluster score. The dendrogram results presented show that the optimal clusters used are 3 clusters and the CCC score obtained is 0.6672, indicating that by using 3 clusters based on the dendrogram, HAC is classified as moderate to strong for determining optimal clusters.

Table 4.5 Hasil Clustering dengan Hierarchical Agglomerative Clustering

No.	Rice Production	Rice Harvest Area	Total population	Price of Rice	Area	Cluster
1.	-0.3068996	-0.2098478	-0.94930779	0.6788107	Sumba Barat	1
2.	0.6019048	0.9294336	0.02045521	0.6408775	Sumba Timur	2
3.	1.0780389	0.7701124	1.14456208	0.5377771	Kupang	2
4.	-0.5950484	-0.6779432	2.08457204	0.5204316	Timor Tengah Selatan	3
5.	0.1740042	0.3527897	0.17015017	0.5862473	Timor Tengah Utara	2
6.	-0.7110710	-0.6227140	-0.19897342	0.5167031	Belu	3
7.	-1.0133500	-1.05579372	-0.29846014	-1.5906952	Alor	1
8.	-1.0972965	-1.17994006	-1.05809232	-1.5906952	Lembata	1
9.	-0.9434144	-0.81752824	0.32914295	-1.5906952	Flores Timur	1
10.	-0.6796507	-0.67507828	0.77357892	0.5167031	Sikka	3
11.	-0.3426723	-0.41723588	0.22593711	0.5642006	Ende	3
12.	0.4879093	0.27066217	-0.77171936	0.5167031	Ngada	3
13.	1.7930151	1.65473470	0.71965154	-1.5906952	Manggarai	2
14.	-0.1836628	-0.11928466	-0.96790343	0.6484966	Rote Ndao	1
15.	2.6262199	2.59840604	0.17851821	0.5167031	Manggarai Barat	2
16.	-0.3614529	-0.40736784	-1.53135154	0.8409182	Sumba Tengah	1
17.	0.2746446	0.88868812	0.66665395	-1.5906952	Sumba Barat Daya	2

18.	0.2617068	0.02412027	-0.82471695	0.6788107	Nagekeo	1
19.	1.3515208	1.23152301	0.36354490	0.6788107	Manggarai Timur	2
20.	-1.0261089	-1.10545225	-1.50903676	-1.5906952	Sabu Raijua	1
21.	-0.3136964	-0.25568648	-0.59134158	0.5609585	Malaka	3
22.	-1.0746405	-1.17659766	2.02413619	0.5410192	Kota Kupang	3

Source: Local Expert and Undana Research Team Analysis (2025)

The graph above represents the distribution of clusters, which were also mapped using a dendrogram in the previous figure. Cluster 1 includes the regions of West Sumba, Alor, Lembata, East Flores, Rote Ndao, Central Sumba, Nagekeo, and Sabu Raijua. These regions have sufficient food availability with relatively even logistics distribution. Community access to food is moderate to good, and they do not experience price pressures or extreme poverty. This condition reflects a region with sufficient food stability, but still needs to strengthen production to become more self-sufficient.

Cluster 2 consists of East Sumba, Kupang, North Central Timor, Manggarai, West Manggarai, Southwest Sumba, and East Manggarai. These regions are centers of food production with high levels of availability. However, uneven distribution makes some areas potentially food insecure. The imbalance between supply and demand also tends to fluctuate food prices. Therefore, the main challenge in this cluster lies in improving distribution efficiency and ensuring equitable access.

Meanwhile, Cluster 3 includes South Central Timor, Belu, Sikka, Ende, Ngada, Malaka, and Kupang City. These regions boast strong logistical access, high food diversity, and the most stable food security compared to other clusters. This stability makes the third cluster a key pillar of provincial food security.

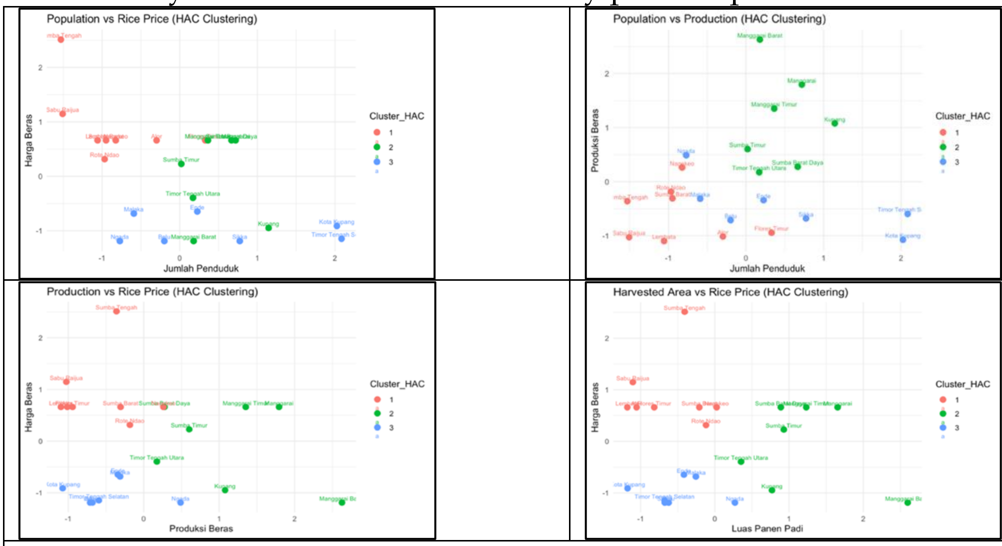

Figure 4.6 Visualization of Relationships between Attributes with Hierarchical Agglomerative Clustering

The analysis of food security in East Nusa Tenggara Province reveals three regional clusters with very different characteristics, vulnerabilities, and potential. These differences are primarily influenced by a combination of population, rice production, harvested area, rice prices, and the effectiveness of the food distribution system in each region.

Cluster 1, which includes West Sumba, Alor, Lembata, East Flores, Rote Ndao, Central Sumba, Nagekeo, and Sabu Raijua, is food insecure and vulnerable to high prices. These regions are characterized by low to moderate populations, limited production and harvested area, and high rice prices. These conditions indicate an insufficient local rice supply, leading to high price pressures. These regions are highly dependent on external supplies and are at risk of a food crisis if the distribution chain is disrupted. To address this vulnerability, government policy interventions are needed through subsidies, production support, irrigation development, and improved food distribution.

Cluster 2 includes East Sumba, Kupang, North Central Timor, Manggarai, West Manggarai, Southwest Sumba, and East Manggarai. This region is a potential food barn with high production, a large harvest area, and relatively stable rice prices. Despite the population diversity, a strong food supply keeps prices moderate. This region plays a strategic role as a food distribution hub to other areas experiencing food deficits. To optimize its function, policies are needed to protect agricultural land, increase sustainable productivity, and ensure efficient post-harvest distribution.

Meanwhile, Cluster 3, consisting of South Central Timor, Belu, Sikka, Ende, Ngada, Malaka, and Kupang City, is a consumer or distribution receiving area. While the population in this region is high, rice production and harvest area are low. Interestingly, rice prices remain low thanks to good logistical access and efficient food distribution, particularly through Kupang City, the center of consumption and supply from production areas (Cluster 2). Therefore, relevant policies include maintaining smooth rice distribution and improving logistics infrastructure such as roads and storage warehouses.

Overall, this pattern indicates that food security in NTT depends not only on production volume but also on distribution efficiency and regional connectivity. Cluster 1 requires strengthening production and distribution to reduce food vulnerability. Cluster 2 serves as a production center whose sustainability must be maintained. Meanwhile, Cluster 3 serves as a consumption center that relies heavily on smooth logistics flows.

DISCUSSION

The results of forecasting model testing using Autoregressive Integrated Moving Average (ARIMA) on various food commodity variables in East Nusa Tenggara Province for the period 2025–2027 showed an inadequate level of accuracy. Model evaluation using three main performance measures, namely Mean Absolute Percentage Error (MAPE), Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC), indicated the model's weakness in accurately explaining the dynamics of food production data. For example, for the

Scallion commodity, the MAPE value reached 91.34%, meaning there was a 91.34% prediction error rate against the actual data. This value indicates that the ARIMA model is unable to provide reliable predictions for this commodity. Similar findings also emerged for other food commodity variables, where relatively high AIC and BIC values strengthen the conclusion that the model is inefficient in describing medium-term data patterns. High AIC and BIC values indicate suboptimal model complexity and weak generalization ability to future data (Mahaluca et al., 2025). (Ezziyyani et al., 2024).

These results confirm that the historical data structure used in the modeling may not be long enough or stable enough, resulting in the model's inability to adequately capture seasonal patterns and long-term trends. Therefore, to improve forecasting accuracy, increasing the time series length or using a more adaptive model, such as the Exponential Smoothing method, may be a more suitable option. Unlike ARIMA, this method is more flexible with short time series data and unstable seasonal patterns. Exponential Smoothing, particularly the Holt-Winters variant, is able to capture seasonal trends and fluctuations more responsively, resulting in more stable and realistic predictions for food commodities. This method is also relatively simple and efficient, allowing it to be applied simultaneously to various types of commodities (Ishaque & Ziblim, 2013); (Starychenko et al., 2020); (Septiana, 2024).

The results of the food security clustering test for districts/cities in East Nusa Tenggara (NTT) Province show differences in spatial and economic characteristics between regions, which can be grouped using two main approaches: K-Means Clustering and Hierarchical Agglomerative Clustering (HAC). Theoretically, both have the same goal of grouping regions based on certain similar characteristics, but they have different mechanisms in the analysis process. The K-Means Clustering method works by direct partitioning, where each region is placed into one of several clusters based on the distance from a predetermined center point (centroid). This approach is efficient and widely used for large data sets. However, this method has limitations because it is unable to describe natural or hierarchical relationships between regions and is highly dependent on a predetermined number of clusters (Oti et al., 2021); (Ay et al., 2023). Unlike K-Means, Hierarchical Agglomerative Clustering (HAC) works by the principle of gradual grouping based on the level of similarity between regions. The process begins by considering each region as a separate cluster, then gradually combining them based on proximity or similar attributes to form a hierarchical structure (dendrogram). This approach provides a more organic, natural, and interpretive picture of interregional relationship patterns. In the context of food security, this is important because it can demonstrate the degree of similarity in food conditions, production, and distribution between districts/cities (Sinaga et al., 2025); (Hii et al., 2024).

In the NTT food security study, the use of HAC is considered more appropriate because the province's regions have highly diverse geographic conditions, infrastructure, and production capacity. The hierarchical approach allows for the identification of tiered vulnerability patterns and natural relationships between regions based on variables such as population, rice production, harvested area, rice prices, and logistical access. Thus, the clustering results not only divide regions into groups but also explain the food security continuum from the most vulnerable to the most stable. The HAC analysis results indicate three main clusters representing different food security conditions. Cluster 1 represents areas vulnerable to food insecurity and prone to high prices, where government intervention is needed in the form of subsidies, production assistance, irrigation infrastructure development, and increased food distribution to stabilize supplies. Cluster 2 consists of areas with high productivity and significant potential as regional food barns, so policies are directed at maintaining production sustainability, optimizing post-harvest efficiency, and protecting agricultural land from non-agricultural conversion. Cluster 3 includes areas with efficient market access and distribution, such as Kupang City and its surrounding areas, which act as centers for rice consumption and distribution. For this cluster, the appropriate policy is to maintain smooth interregional logistics flows and ensure stable food stocks. This hierarchical approach not only provides a map of food security conditions but also opens up opportunities for more contextual and spatial intervention planning. Several international studies have also shown the effectiveness of the HAC method in analyzing food vulnerability and poverty in heterogeneous regions, such as those conducted by (Van Dijk & Meijerink, 2014) in their study of global food security and (Paloviita et al., 2016). This method helps understand the complex interactions between food production, distribution, and consumption factors across regions.

Therefore, the use of Hierarchical Agglomerative Clustering in the context of food security in NTT is not merely a statistical technique but also an analytical approach that enables evidence-based policy formulation. Through this cluster mapping, local governments can design differentiated food policy strategies that strengthen production in vulnerable areas (Cluster 1), increase efficiency in surplus areas (Cluster 2), and strengthen distribution in consumption areas (Cluster 3), thereby achieving sustainable and equitable regional food security.

CONCLUSIONS AND RECOMMENDATIONS

Based on the previous results and discussion, the conclusions and recommendations of this study are as follows:

1. Based on the overall analysis, it can be concluded that the ARIMA forecasting model does not provide optimal results in predicting food commodity production in East Nusa Tenggara Province for the period 2025–2027. High MAPE values, along with large AIC and BIC values, indicate a significant level of prediction error, as seen for Scallions, which achieved an MAPE of 91.34%. This indicates that the ARIMA model is not sufficiently sensitive to food production dynamics, which are heavily influenced by external factors such as climate, planting season, and distribution limitations. Therefore, alternative, more adaptive forecasting methods are needed, such as Exponential Smoothing, which is more responsive to short data sets and unstable seasonal patterns.

2. Furthermore, the results of the spatial analysis using K-Means and HAC provide a deeper understanding of food security patterns between regions in NTT. K-Means helps group regions based on shared characteristics in a partitioned manner, while HAC provides a gradual depiction of the natural relationships between regions. Based on the HAC, the NTT region is divided into three clusters: Cluster 1, which is a vulnerable region with low production and high prices, requiring production and distribution interventions. Cluster 2, which is a surplus region serving as a regional food barn, needs to be optimized to support deficit areas. Cluster 3, which is a consumption region with efficient distribution, needs to maintain stable supply and logistics. The combination of forecasting and clustering results provides a strong foundation for formulating more targeted and evidence-based food security policies. Improved forecasting methods will strengthen the accuracy of food production predictions, while clustering results enable the design of different policies for each type of region, including interventions in vulnerable areas, optimization in surplus areas, and strengthening distribution in consumption areas. This strategy will support the achievement of more equitable, adaptive, and sustainable food security in East Nusa Tenggara Province.

This research makes a significant contribution to the application of scientific calculations for forecasting and the development of new food security status models relevant to regional conditions. Key findings from the forecasting and clustering analysis indicate that food security policies in East Nusa Tenggara Province need to be formulated adaptively and realistically, taking into account implementation capacity at the regional level.

For Cluster 1, areas with low rice production and high rice prices, interventions are directed at increasing production capacity and improving food distribution. Priority measures that are still realistic at the regional level include: rehabilitation and construction of simple irrigation in priority areas, distribution of agricultural production input assistance through technical service programs, and implementation of market operations coordinated with the Logistics Agency (Bulog) and district/city governments. The focus of this policy is not on large-scale development, but rather on strengthening local production systems, which have a direct impact on increasing farmer productivity and stabilizing prices in local markets.

For Cluster 2, areas with potential food barns, policies are directed at optimizing production and reducing yield losses through a phased approach and in accordance with regional fiscal capacity. Regional governments can utilize the central government's agricultural machinery (alsintan) assistance program, strengthen post-harvest storage systems at the village or farmer group level, and establish simple food distribution centers that can serve as buffer stocks between regions. This strategy can be implemented through partnerships between regional governments, farmer groups, village-owned enterprises (BUMDes), and the local private sector, without relying on large external investments.

Meanwhile, Cluster 3, as a consumption area with better distribution access, requires pragmatic policies to strengthen logistics and regional connectivity. The main focus will be improving connecting roads between production and consumption areas using regional and central funding sources (DAK Infrastructure), establishing small- and medium-scale logistics hubs at strategic locations such as Kupang City, and utilizing simple digital platforms to accelerate the flow of information on food distribution. This approach can be implemented through collaboration with BUMDs, market players, and the local private sector.

FURTHER STUDY

Across clusters, regional food security policies need to be strengthened by developing a simple yet functional integrated food information system based on real-time data on prices, stocks, and distribution flows. Local governments can also encourage interregional partnerships to ensure smooth supply from surplus areas (Cluster 2) to consumption areas (Cluster 3). Furthermore, adaptation to climate change needs to be gradually integrated through weather monitoring and planting calendars based on local data.

ACKNOWLEDGMENT

This research is a collaborative research between Local Expert and the Directorate General of Treasury (DJPb) of the East Nusa Tenggara Representative Office, so the author would like to express his gratitude for the cooperation during this research.

REFERENCES

Asmara, R., Hanani, N., & Mutisari, R. (2012). Analisis Ketahanan Pangan di Kota Batu. *Agricultural Socio-Economics Journal*, *12*(3), 232.

Bapanas. (2025). *Indeks Ketahanan Pangan (IKP) Kabupaten/Kota - DATASET | Portal Data Badan Pangan Nasional*. Satudata.Badanpangan.Go.Id. https://satudata.badanpangan.go.id/datasetpublications/xi/indeks-ketahanan-pangan-kab-kota

Béné, C. (2020). Resilience of local food systems and links to food security–A review of some important concepts in the context of COVID-19 and other shocks. *Food Security*, *12*(4), 805–822.

de Oliveira, B. R., dos Santos Silva, F. C., Mezzomo, R., Barrozo, L. M., da Costa Zanatta, T. S., dos Santos, J. C., Sousa, C. H. C., Coelho, Y. P., Caldas, A. do N. S., & Zuffo, A. M. (2024). Temporal variability in soybean sowing and harvesting according to K-means and silhouette scores. *Trends in Agricultural and Environmental Sciences*, e240010–e240010.

Dwijayanti, A. (2024). Policy Spillover: Analisis Jaringan Dampak Kebijakan Makan Siang Bergizi Gratis terhadap Sektor Pertanian. *Jurnal Administrasi Publik*, *20*(2), 281–308.

Ezziyyani, M., Cherrat, L., Chrit, S., & Mamoune, S. El. (2024). Forecasting Food Security in Morocco. *International Conference on Advanced Intelligent Systems for Sustainable Development*, 734–747.

Facendola, R., Ottomano Palmisano, G., De Boni, A., Acciani, C., & Roma, R. (2023). Profiling citizens on perception of key factors of food security: An application of k-means cluster analysis. *Sustainability*, *15*(13), 9915.

Hidayah, N. (2011). Kesiapan psikologis masyarakat pedesaan dan perkotaan menghadapi diversifikasi pangan pokok. *Humanitas*, *8*(1), 88.

Hii, E. C. S., Lim, S. M., Loo, S. X., Yeap, S. K., & Tan, C. Y. (2024). Clustering Analysis of Food Security, Waste and Loss: Malaysia Agricultural Insights. *JOIV: International Journal on Informatics Visualization*, *8*(3–2), 1892–1904.

Mahaluca, F., Carsane, F., & Vilanculos, A. (2025). Forecasting Rice Production in Mozambique Using Time Series Models and Artificial Neural Networks: Implications for Food and Nutritional Security in the Context of SDG 2. *J Eco Res & Rev*, *5*(1), 1–15.

Merlinda, A. A., & Yusuf, Y. (2025). Analisis Program Makan Gratis Prabowo Subianto Terhadap Strategi Peningkatan Motivasi Belajar Siswa di Sekolah Tinjauan dari Perspektif Sosiologi Pendidikan. *Ranah Research: Journal of Multidisciplinary Research and Development*, *7*(2), 1364–1373.

Moffitt, S. L., O'Neill, M. K., & Cohen, D. K. (2023). 4 Problems of Policy Spillover. In *Reforming the Reform* (pp. 65–91). University of Chicago Press.

Muhammad Farhan, F., & Hani Nurul Fadhilah, H. (2023). *INOVASI PANGAN LOKAL PRODUK GULA AREN JAHE SARI NABATI UNTUK MENINGKATKAN NILAI NUTRISI MINUMAN DALAM SEDIAAN SERBUK*. Fakultas Teknik Universitas Sultan Ageng Tirtayasa.

Nath, B., Dhakre, D. S., & Bhattacharya, D. (2019). Forecasting wheat production in India: An ARIMA modelling approach. *Journal of Pharmacognosy and Phytochemistry*, *8*(1), 2158–2165.

Nochai, R., & Nochai, T. (2006). ARIMA model for forecasting oil palm price. *Proceedings of the 2nd IMT-GT Regional Conference on Mathematics, Statistics and Applications*, 13–15.

Septiana, D. (2024). Forecasting Rice Prices with Holt-Winter Exponential Smoothing Model. *Hanif Journal of Information Systems*, *1*(2), 62–67.

Shastri, S., Sharma, A., Mansotra, V., Sharma, A., Bhadwal, A., & Kumari, M. (2018). A study on exponential smoothing method for forecasting. *International Journal of Computer Sciences and Engineering*, *6*(4), 482–485.

Starychenko, Y., Skrypnyk, A., Babenko, V., Klymenko, N., & Tuzhyk, K. (2020). Food security indices in Ukraine: forecast methods and trends. *Studies of Applied Economics*, *38*(4).

Thapa, R., Devkota, S., Subedi, S., & Jamshidi, B. (2022). Forecasting area, production and productivity of vegetable crops in Nepal using the Box-Jenkins ARIMA model. *Turkish Journal of Agriculture-Food Science and Technology*, *10*(2), 174–181.

Yu, H., & Hou, X. (2022). Hierarchical clustering in astronomy. *Astronomy and Computing*, *41*, 100662.

Yuwono, T., Widodo, S., Darwanto, D. H., Indradewa, D., Somowiyarjo, S., & Hariadi, S. S. (2019). *Pembangunan Pertanian: Membangun Kedaulatan Pangan*. UGM PRESS.