CLUSTER MODELING FOOD SECURITY BASED ON MACHINE LEARNING APPROACH FOR REGIONAL PLANNING IN LAMPUNG PROVINCE

Feri Utomo, *Fitriani*, *Septafiansyah* Dwi Putra, *Irmayani* Noer
Master's Program in Food Security,
Politeknik Negeri Lampung, Bandar Lampung 35141

Abstract. Food security is crucial in ensuring a region's welfare and socio-economic stability. However, economic inequality and poverty often hinder achieving sustainable food security, particularly in regions such as Lampung Province. This research aims to analyze the relationship between Gross Regional Domestic Product at Constant Prices (GRDP ADHK), the Human Development Index (HDI), and the percentage of household expenditure on food as indicators of food security. Using the Random Forest Classifier machine learning algorithm, a quantitative approach is applied to evaluate each indicator's relative influence, and the K-Means Clustering method is used to group regions based on their socio-economic characteristics. The model evaluation results indicate high performance, with an R^2 value of 0.994 and very low prediction error. The feature importance analysis reveals that food expenditure is the most dominant indicator (45.6%), followed by GRDP ADHK (35.1%) and HDI (19.3%). Decision tree visualization and cluster analysis identify three regional typologies: Cluster 0 (food-vulnerable, middle-income, high consumption), Cluster 1 (high food security, high economic status, and HDI), and Cluster 2 (low economic status, high HDI, efficient consumption). These findings indicate that food security is influenced not only by income but also by social policies and resource distribution. This study provides a strong data-driven foundation for formulating food and social development policies in Lampung Province and recommends targeted interventions based on the identified regional typologies.

Keywords: GDP at Constant Prices, Household Expenditure Percentage, Human Development Index (HDI), K-Means Clustering, Random Forest Classifier.

1. Introduction

Food security is one of the fundamental pillars in maintaining a country's economic, social, and political stability. This concept refers to a condition in which all people, at all times, have physical, social, and economic access to sufficient, safe, and nutritious food that meets their dietary needs and food preferences for an active and healthy life[1]. However, economic inequality—marked by income disparity, unequal wealth distribution, and limited access to resources—has been shown to hinder the achievement of inclusive and sustainable food security. [2]; [3].

Economic inequality and food insecurity are closely intertwined, particularly in developing countries. In these regions, income disparities exacerbate poverty, limit access to productive resources, and increase vulnerability to economic shocks. In Indonesia, specifically in Lampung Province, the relationship between economic inequality and food security is highly relevant to analyze as a basis for formulating pro-poor development policies and promoting more equitable growth.

Lampung Province, located at the southern tip of Sumatra Island, is known for its abundant agricultural resources such as coffee, rubber, and palm oil. [4]. Beyond its contribution to the regional economy, the province's agrarian and social structure has also been shaped by migration patterns from Java. [5]. Despite its natural wealth and agrarian potential, Lampung continues to face economic inequality and poverty, which directly impact food vulnerability, particularly among poor households and marginalized communities.

These challenges are further exacerbated by external factors such as climate change, market price fluctuations, and global economic dynamics, which often disproportionately affect vulnerable groups. On the other hand, household food consumption patterns are strongly influenced by commodity prices, income levels, household size, and demographic shifts. [6]. Therefore, a comprehensive understanding of household income and expenditure patterns is crucial to identify potential food insecurity risks and to design well-targeted policy interventions. [7].

Poverty has long been recognized as a significant cause of food insecurity. Limited purchasing power among people experiencing poverty prevents them from accessing adequate and nutritious food, thereby increasing the risk of malnutrition, health problems, and low productivity. This creates a vicious cycle between poverty and food security, reinforcing each other. In addition, income inequality can indirectly impact food security by restricting access to essential services such as education and healthcare. [8]. The government's role becomes crucial in policy contexts. Poverty alleviation and food security programs must be strategically designed and sustainably implemented. [9]; [10].

Economic growth and social welfare disparity are not locally confined phenomena. Studies in various countries show that increases in GDP do not always correspond to significant reductions in poverty. For instance, India has recorded rapid GDP growth but still ranks 74th in the Multidimensional Poverty Index. [11]. A similar pattern is

1 Corresponding author: utomoferi8@gmail.com

observed in the Maminasata region of Indonesia, where there is a clear imbalance between economic growth and the achievement of social well-being. [12].
The dynamics between poverty, income inequality, and food security are highly complex. Increasing food availability is insufficient to eradicate poverty and hunger as envisioned in the Sustainable Development Goals. [13]. Addressing food security issues requires a holistic approach considering multiple factors such as economic growth, income distribution, resource accessibility, and social inclusion. Globally, more than 800 million people experience food insecurity and malnutrition due to monotonous diets, often linked to low income, climate change, conflict, and population growth. [14]. Food production and rural development, food and nutrition security, the protection and sustainability of natural resources and the environment, and water resource pollution are all interrelated issues. [15].
In light of these challenges, a more in-depth data-driven analysis is necessary. Previous research has shown that conventional analysis often fails to capture the complexity of the relationship between economic growth and food security. Therefore, this study proposes a novel approach using AI (machine learning) technology. This technology enables the discovery of hidden patterns and nonlinear relationships between GDP and welfare indicators such as malnutrition, food consumption patterns, and food production. The platform ensures that all innovative agriculture approaches aim to optimize outcomes in pursuing national food security [6]. This research aims to analyze the relationship between Gross Regional Domestic Product based on Constant Prices (GRDP ADHK), the Human Development Index (HDI), and the percentage of household per capita expenditure on food as indicators of food security. Furthermore, it seeks to cluster districts/cities in Lampung Province to identify patterns of economic disparity and food-vulnerable regions. The study is expected to contribute to the development of food policy in response to socio-economic dynamics and to provide guidance for implementing models in regions with varying characteristics.

2. Research method

This study was conducted in the province of Lampung. The research process took place from March to April 2025 and encompassed the stages of data collection, data processing, and implementing and evaluating machine learning models. The variables used in this study include the Gross Regional Domestic Product at Constant Prices (GRDP ADHK), accessibility (measured by the percentage of per capita household expenditure on food), and economic inequality (measured by the Human Development Index or HDI). The data span the years 2014 to 2024.
Data analysis was performed using the Random Forest Classifier to determine the most influential factors affecting food expenditure or food insecurity, and the K-Means algorithm was used to cluster districts based on the combination of the three indicators. Model evaluation employed RMSE and R^2 metrics, which were then used to interpret the modelling results. The mathematical formulation applied in the analysis follows the approach by [16] As described below:

1. Data normal

$$X' = \frac{X - \mu}{\sigma} \qquad (1)$$

Explanation:
X = original data value
M = mean of the data
σ = standard deviation of the data
X′ = value of the data after standardization

2. *Random Forest* Modelling

$$\hat{y}_{RF} = \text{mode}(\hat{y}_1, \hat{y}_2, ..., \hat{y}_N) \qquad (2)$$

Explanation:
Ŷi = the prediction from the *i*-th decision tree
N = the total number of trees in the Random Forest
ŷRF = the average of all tree predictions (for regression)

3. Feature Importance based on *Decrease in Impurity* (*Gini/Entropy*):

$$FI_j = \sum_{t \in T_j} \frac{N_t}{N} \cdot \Delta i(t) \qquad (3)$$

4. K-Means Algorithm Classification

$$J = \sum_{k=1}^{K} \sum_{X_i \in \text{Cluster}_k} \|X_i - C_k\|^2 \qquad (4)$$

1 Corresponding author: utomoferi8@gmail.com

$||X_i - C_k||^2$ = The squared Euclidean distance between the data point and the cluster centroid.
J = An objective function that reflects the total clustering error.

5. Model Evaluation Using RMSE and R^2 Metrics

RMSE (*Root Mean Squared Error*): Used to measure the prediction error.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{5}$$

Explanation:
Yi : The actual value of GRDP (Gross Regional Domestic Product).
ŷi: The predicted value from the decision tree model.
R-squared (R^2) indicates how well the model explains the variance in the data.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{6}$$

Where $\bar{y}$ represents the mean of the actual values.
A schematic overview of the workflow of the Random Forest and K-Means algorithms in modelling food security clusters in Lampung Province is presented in Figure 1. This figure illustrates the research methodology.

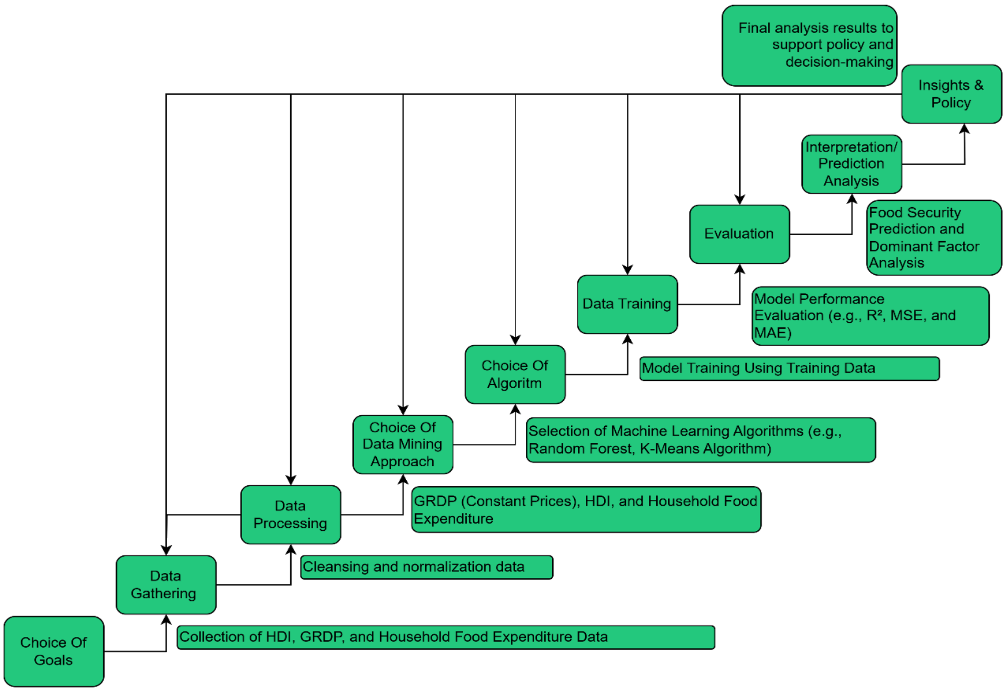

Figure 1. Research Implementation Method

3. Results and discussion

Correlation Between GRDP (Constant Prices), HDI, and the Percentage of Household Expenditure on Food

The three indicators used in this study are interrelated in shaping the overall picture of food security in the Lampung region. Gross Regional Domestic Product (GRDP) serves as an indicator of economic development, which is essentially accompanied by an increase in per capita income and can be used to measure the impact of income distribution inequality. [17]. Per capita GRDP reflects the region's economic capacity and the purchasing power of its population, which directly affects households' ability to access sufficient and nutritious food. A high regional income value can increase the Human Development Index (HDI) through improved income levels and better access to healthcare and education services. A high HDI can influence household consumption patterns toward healthier and more diverse diets. Meanwhile, the percentage of household consumption allocated to food can be used to identify households vulnerable to food insecurity. A high proportion of income spent on food

1 Corresponding author: utomoferi8@gmail.com

indicates a significant allocation of household income solely to meet basic nutritional needs, reflecting limited purchasing power and potential food insecurity.

This phenomenon aligns with Engel's Law, which describes the relationship between household consumption demand and income levels. According to Engel's Law, as income increases, the proportion of income spent on food decreases, even if the nominal amount continues to rise (Wan, 1996; Puspita & Agustina, 2020). This reflects that more prosperous communities have better access to quality food, a key pillar of food security. Equitable income growth is crucial for strengthening food security, enabling all segments of society to access adequate, safe, and nutritious food. This is supported by findings published in the journal *Food Policy*, which state that inclusive economic growth has a direct impact on improving household food status in developing countries [19]. The relationship among these three indicators in the study is mutually reinforcing: GRDP at Constant Prices affects the structure of regional income, which influences the quality of human development (HDI), and is ultimately reflected in household spending patterns on food.

Evaluation of the *Random Forest* Model

The Random Forest analysis in this study aims to identify the most dominant indicator in determining the regional clustering in Lampung Province. After conducting data cleansing, imputation, data normalization, and model testing on the three selected indicators, the model evaluation results are presented in Table 1.

Table 1. Results of the *Random Forest* Model Evaluation

Evaluation Methods	Score
R^2 Score	0,994
Mean Squared Error (MSE)	0,002
Mean Absolute Error (MAE)	0,007

Based on the evaluation results of the classification model using the Decision Tree algorithm on the food security clustering in Lampung Province, the coefficient of determination (R^2) was found to be 0.994. This value indicates that the model can explain 99.4% of the variation in the cluster data, suggesting an exceptionally high level of goodness of fit. The coefficient of determination (R-squared) has the advantage of providing more accurate and transparent information regarding the proportion of data variation explained by the model. Empirical studies have shown that R-squared does not have interpretability limitations like other metrics. [20]. Furthermore, the very low Mean Squared Error (MSE) of 0.0023 and the Mean Absolute Error (MAE) of 0.0071 indicate that the model has a very low prediction error. Therefore, this model can be considered highly accurate in classifying regions into their respective food security clusters based on the indicators of GRDP at constant prices, Human Development Index (HDI), and Percentage of Household Expenditure on Food.

***Feature importance* analysis**

The feature importance analysis measures how much each variable contributes to dividing the data into clusters. The importance values, expressed as percentages, indicate the relative influence of each feature on the model's classification decisions. The results of this analysis are presented in Table 2.

Table 2. Results of *Feature Importance* Analysis

Factor	Importance
Percentage of Household Expenditure on Food	45.6%
GRDP AT CONSTANT PRICES	35.1%
IPM	19.3%

The feature importance analysis from the decision tree classification model indicates that the Percentage of Household Expenditure on Food indicator has the most significant influence in distinguishing food security clusters, with a contribution weight of 45.6%. This highlights that the proportion of household spending on food consumption is a decisive indicator in categorizing regions based on their level of food security. Meanwhile, the Gross Regional Domestic Product at Constant Prices accounts for 35.1% of the model's decision-making, reflecting the significant role of a region's real economic capacity in shaping food security. On the other hand, although the Human Development Index (HDI) contributes only 19.3%, it remains a relevant indicator as it reflects the human development dimension that also affects the socioeconomic conditions of a region.

***Decision tree* visualization**

1 Corresponding author: utomoferi8@gmail.com

The classification pattern of food security in Lampung Province, modelled using a decision tree algorithm, is visualized in the modelling output presented in Figure 2. This visualization illustrates how each indicator contributes to determining the cluster of regions based on their level of food security.

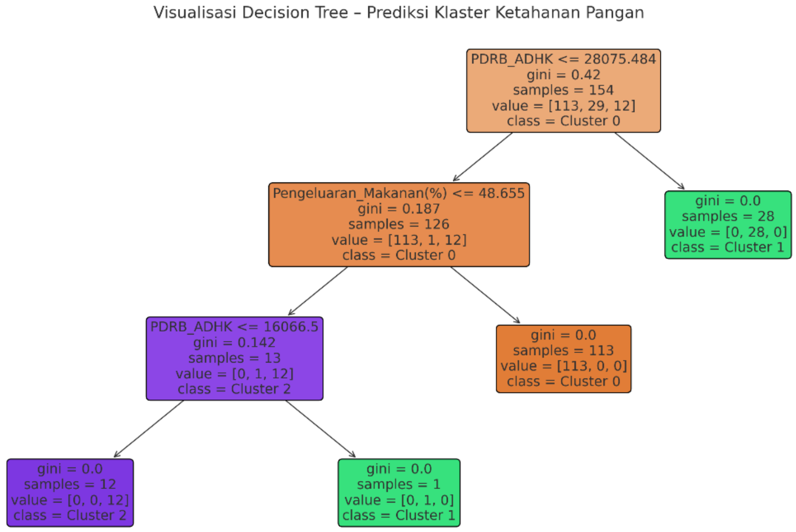

Figure 2. Decision Tree Visualization

Based on the Decision Tree visualization in Figure 2, it can be analyzed that the classification model successfully maps regions into three clusters based on the indicators of GRDP at Constant Prices (GRDP at constant prices) and food expenditure. The model demonstrates that a combination of economic and social indicators categorizes regions into three main clusters: Cluster 0, Cluster 1, and Cluster 2. The initial node splits the data based on the threshold of GRDP at constant prices ≤ 28,075.484, indicating that regions with higher GRDP levels tend to fall into Cluster 1. In contrast, regions with lower GRDP are divided based on the percentage of food expenditure. If the proportion of food expenditure exceeds 48.655%, the region is classified under Cluster 0, which is dominated by households with limited non-food or high food consumption. Regions with very low GRDP at constant prices values (≤ 16,066.5) and high food expenditure are classified into Cluster 2. Meanwhile, the Human Development Index (HDI) indicator does not appear in the visualization, as this variable does not significantly improve cluster separation. However, HDI values can still serve as supporting information in policy formulation.

Clustering Using the *K-Means Algorithm*

The following chart presents the results of the *elbow method* analysis to determine the optimal number of clusters for grouping regions based on three (3) observed indicators.

1 Corresponding author: utomoferi8@gmail.com

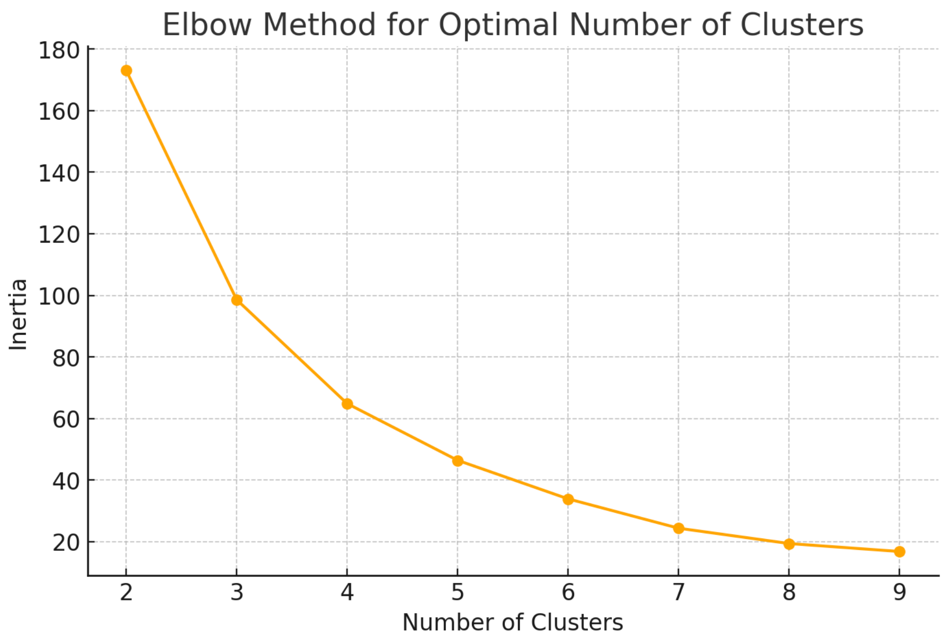

Figure 3. *Elbow Method* Chart

The graph above illustrates the result of the Elbow Method. The "elbow" point is most clearly visible at 3 clusters, where the decrease in inertia begins to slow down. Selecting 3 clusters is considered the optimal choice for segmenting regions based on the indicators used: Gross Regional Domestic Product at Constant Prices (GRDP at Constant Prices), Food Expenditure (%), and Human Development Index (HDI). The regional segmentation clustering is presented in Table 3.

Table 3. Regional Segmentation Clusters

Cluster	PDRB_ADHK	Percentage of Household Expenditure on Food	IPM	Total Region-Year Observations
0	9731,57	55,41	66,21	113
1	39291,55	50,95	72,2	29
2	3972,25	40,8	75,33	12

Based on the K-Means algorithm clustering results, three main regional profiles were identified in Lampung Province, reflecting the diversity of socio-economic conditions and food security levels. Cluster 0, comprising 113 region-years, has an average GRDP at Constant Prices of 9,731.57, food expenditure of 55.41%, and a relatively low Human Development Index (HDI) of 66.21. This cluster represents areas with moderate income levels, high food expenditure, and a medium HDI. These characteristics typically describe suburban or semi-urban regions in a transitional development phase, where access to food remains challenging due to heavy reliance on household spending.

Cluster 1, which includes 29 region-years, displays regions with a relatively high GRDP (39,291.55), still high but lower food expenditure (50.95%), and a medium HDI of 72.20. This cluster reflects areas with high income, high HDI, and moderately lower food expenditure. The profile indicates economically and socially developed urban areas with the highest food security among the three clusters. Regions in this cluster tend to enjoy better access to infrastructure, public services, and food markets.

Cluster 2 is the smallest, comprising only 12 region-years but has distinct characteristics. Areas in this cluster have a very low GRDP (3,972.25), yet the lowest food expenditure (40.8%) and the highest HDI (75.33) among all clusters. Although this cluster represents a small portion of the regions, it exhibits a unique pattern: extremely low-income levels, high HDI, and low food expenditure proportions. This pattern suggests the possible presence of significant social support, food subsidies, or a high level of non-commercial food consumption, such as subsistence farming. Therefore, although these areas are economically disadvantaged, food security is maintained through non-market mechanisms. These findings highlight that food security is not solely determined by income, but also by access, distribution, and policies that support community well-being.

Regional distribution by Human Development Index (HDI) and Gross Regional Domestic Product (GRDP) at Constant Prices.

1 Corresponding author: utomoferi8@gmail.com

The distribution graph of regions based on the Human Development Index (HDI) and Gross Regional Domestic Product at Constant Prices (GRDP ADHK) is presented in Figure 4. Each point represents a combination of a district/city and a specific year, with the color indicating a similar socio-economic cluster.

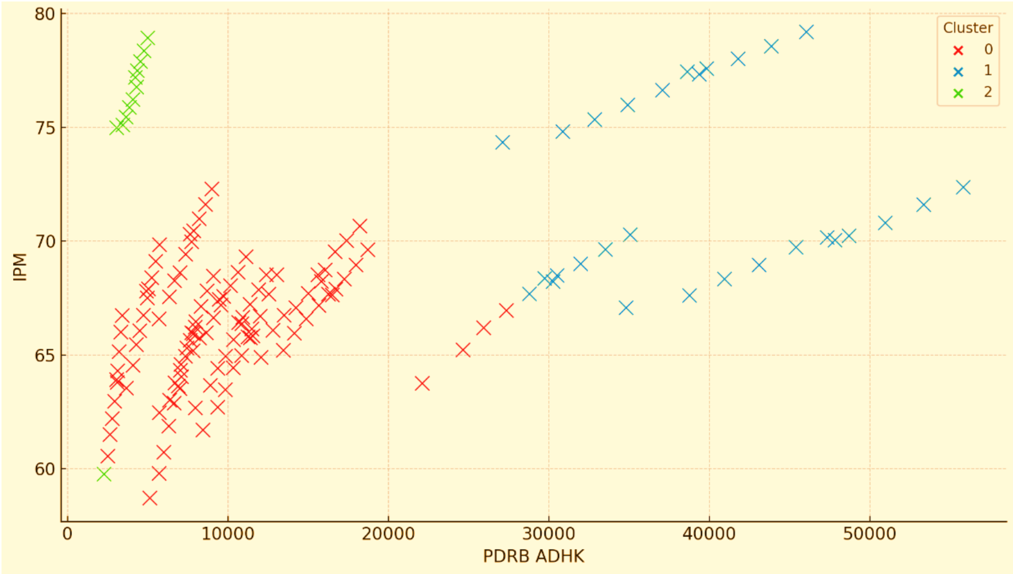

Figure 4. Cluster Distribution Based on HDI and GRDP at Constant Prices (2014–2024)

The clustering results of regions in Lampung Province based on GRDP at Constant Prices (GRDP at constant prices) and the Human Development Index (HDI) reveal three main patterns that reflect varying economic progress and human development levels.

Cluster 0 comprises regencies/cities characterized by moderate economic activity and a medium-level HDI. Regions such as South Lampung, North Lampung, West Lampung, Tanggamus, Way Kanan, Tulang Bawang, East Lampung, West Tulang Bawang, Pesawaran, Pringsewu, and Mesuji fall under this group. These areas show signs of emerging economic activity but have not yet achieved significant improvements in residents' quality of life. This situation highlights the need for balanced economic and social development interventions to prevent these regions from falling further behind.

Cluster 1 includes areas with both high GRDP and high HDI, such as Bandar Lampung and, in certain years, South Lampung and Central Lampung. These regions represent developed areas with strong economic systems and well-established human development, supported by relatively good public services and integrated regional connectivity. The development model of these regions can serve as a reference or role model for the development of other areas within the same province.

Cluster 2 presents a unique and paradoxical profile—regions such as Metro City and West Coast (Pesisir Barat) with a high HDI but low GRDP. Although these areas do not possess substantial economic capacity, they maintain a high level of human development. This phenomenon suggests the presence of efficient social service management or the possibility of effective government interventions or external support, such as subsidies. Therefore, this cluster warrants further study to explore the successful strategies or policies that could be replicated in other economically disadvantaged but socially promising areas.

Strategically, these three clusters provide valuable insights for formulating regional development policies. Cluster 1 can serve as a best-practice model for integrated human and economic development. Cluster 0 requires integrated interventions to bridge development gaps. Meanwhile, Cluster 2 deserves focused research to identify effective social development models despite economic limitations.

Regional distribution is based on GDP at Constant Prices (GRDP AT CONSTANT PRICES) and the percentage of food expenditure.

The distribution chart of regions based on GRDP at Constant Prices (GRDP at constant prices) and food expenditure percentage is grouped into clusters. This chart helps identify the relationship patterns between regional income and the proportion of food expenditure in each area and year. Presented in Figure 5.

1 Corresponding author: utomoferi8@gmail.com

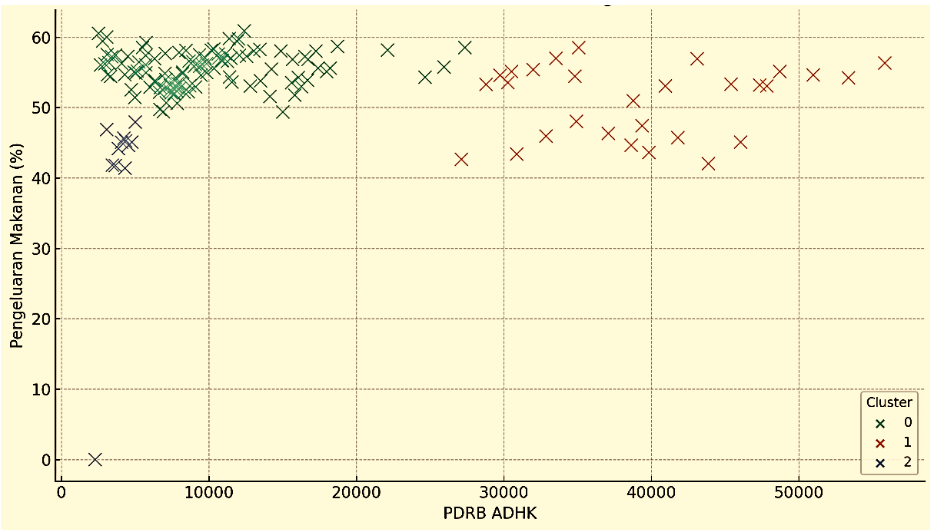

Figure 5. Cluster Distribution Based on GDP at Constant Prices and Food Expenditure.

Cluster analysis based on GRDP (at Constant Prices) and the proportion of food expenditure in Lampung Province reveals three regions with distinct economic and food consumption characteristics. Cluster 0 reflects regions with lower-middle economies and high proportions of household food expenditure. This indicates that a significant portion of household income is spent on basic needs, especially food, signalling vulnerability to price fluctuations. Regions such as South Lampung, East Lampung, West Lampung, North Lampung, Way Kanan, Tanggamus, Tulang Bawang, West Tulang Bawang, Pesawaran, Pringsewu, Mesuji, and West Coast (in certain years) fall into this category. These conditions suggest limited purchasing power and highlight a crucial need for prioritizing food security interventions in these areas.

Cluster 1 represents urban areas or regions with strong and established economies. A high average GRDP and a relatively low proportion of food expenditure indicate that people in these areas have moved beyond fulfilling basic needs and are likely to have more diverse spending allocations, including education, healthcare, and recreation. In certain years, regions such as Bandar Lampung, Metro, Central Lampung, and South Lampung fall into this cluster. This consumption pattern reflects higher purchasing power and a more stable economic structure, making these regions ideal models for development planning focused on sustainable welfare improvement.

Cluster 2 shows an unusual phenomenon, where regions have low GRDP and a low proportion of food expenditure. This may suggest access to self-sourced food, government subsidy programs, or minimalist consumption patterns. Regions such as the West Coast, Metro, and Tanggamus in specific years are categorized in this cluster. These conditions require deeper investigation to determine whether the low food expenditure reflects efficiency or a hidden issue of insufficient food intake in terms of quality and quantity.

These findings point to differentiated policy directions for each cluster. Cluster 1 can serve as a model for ideal development, successfully combining economic strength and healthy food consumption. Cluster 0 requires targeted government attention through programs to strengthen local economies, provide employment opportunities, and improve access to and stabilize food prices. Meanwhile, Cluster 2 warrants further study to ascertain whether its low consumption pattern results from efficiency or indicates underlying problems in food access and nutritional security.

4. Conclusion

This study successfully identified and mapped food security characteristics in Lampung Province through a combined approach using Random Forest and K-Means Clustering based on three key indicators. The feature importance analysis revealed that food expenditure was the most dominant factor in classifying food security clusters, followed by regional GDP at Constant Prices (GRDP AT CONSTANT PRICES) and the Human Development Index (HDI). The Decision Tree visualization illustrated a logical cluster separation based on threshold values of PDRB and the proportion of food expenditure. Although HDI did not appear explicitly in the tree structure due to its lower contribution to cluster separation, it remains relevant for interpreting the

1 Corresponding author: utomoferi8@gmail.com

socioeconomic context of the regions. The regional distribution using the K-Means method on the scatter plot further reinforced the mapping. It provided clear policy directions: Cluster 0 requires integrated development interventions, Cluster 1 can serve as a model for sustainable development, and Cluster 2 requires further analysis to determine whether its food consumption efficiency reflects resilience or conceals underlying vulnerabilities. Combining quantitative methods and spatial visualization provides a foundation for evidence-based regional development policymaking and data-driven food security strengthening in Lampung Province.

References

1. D. Susilastuti, Eur. Res. Stud. J. **20**, 164 (2017)
2. A. H. Setiawan, E. Y. A. Guanto, and B. Hayati, Econ. Dev. Anal. J. **9**, (2020)
3. S. Sugiyanto and L. B. Pintakami, Agric. Soc. Econ. J. **21**, 155 (2021)
4. N. Prawoto and R. D. Cahyani, J. Asian Financ. Econ. Bus. **7**, 489 (2020)
5. B. Arifin, AgEcon Search **7**, 18 (2007)
6. A. S. Putra, G. Tong, and D. O. Pribadi, Sustain. **12**, 1 (2020)
7. A. Annisa, B. G. D. Saputri, R. Asmara, F. Fahriyah, and N. Hanani, Agro Ekon. **33**, 91 (2022)
8. J. S. Budi, M. Munajat, and R. Efrianti, AJARCDE (Asian J. Appl. Res. Community Dev. Empower. **6**, 1 (2022)
9. G. Pradana, M. Farid, and A. Azbihardiyanti, **383**, 345 (2019)
10. F. Rahmanto, E. P. Purnomo, and A. N. Kasiwi, Caraka Tani J. Sustain. Agric. **36**, 33 (2021)
11. G. Vasishtha and S. K. Mohanty, Spat. Demogr. **9**, 213 (2021)
12. Mahyuddin, M. Arsyad, R. Darma, M. Fahmid, L. Fudjaja, R. M. Rukka, and Miftahulmawaddah, IOP Conf. Ser. Earth Environ. Sci. **807**, (2021)
13. N. Sanober, Arch. Food Nutr. Sci. **7**, 001 (2023)
14. A. Assaye, C. Seyoum, and E. Ndemo, Int. J. Agric. Ext. **8**, 85 (2020)
15. R. Çakmakçı, M. A. Salık, and S. Çakmakçı, (n.d.)
16. L. Breiman, Kluwer Acad. Publ. Manuf. **45**, 5 (2001)
17. A. Nilasari and R. Amelia, Nomicpedia J. Econ. Bus. Innov. **2**, 171 (2022)
18. C. D. Puspita and N. Agustina, Semin. Nas. Off. Stat. **2019**, 700 (2020)
19. D. D. Headey and H. H. Alderman, J. Nutr. **149**, 2020 (2019)
20. D. Chicco, M. J. Warrens, and G. Jurman, PeerJ Comput. Sci. **7**, 1 (2021)

1 Corresponding author: utomoferi8@gmail.com