International Journal of Advances in Applied Sciences (IJAAS)
Vol. 15, No. 2, June 2026, pp. 872~882
ISSN: 2252-8814, DOI: 10.11591/ijaas.v15.i2.pp872-882

Predictive modeling of regional economic growth using agricultural and socioeconomic indicators

Septafiansyah Dwi Putra[1], Feri Utomo[2], Fitriani[2], Irmayani Noer[2], Heriansyah[3,4]
[1]Department of Information Technology, Politeknik Negeri Lampung, Bandar Lampung, Indonesia
[2]Master of Applied Food Security Program, Politeknik Negeri Lampung, Bandar Lampung, Indonesia
[3]Institute of Telecooperation, Johannes Kepler University Linz, Linz, Austria
[4]Department of Electrical Engineering, Institut Teknologi Sumatera, Lampung, Indonesia

Article Info

Article history:

Received Oct 26, 2025
Revised Apr 30, 2026
Accepted May 22, 2026

Keywords:

Data-driven policy
Feature importance analysis
Gross regional domestic product prediction
Machine learning
Sustainable agriculture

ABSTRACT

Economic growth and food security are closely interconnected dimensions of sustainable regional development, particularly in agrarian regions such as Lampung Province, Indonesia. However, conventional analytical approaches often fail to capture the complex and nonlinear relationships between agricultural productivity and socio-economic conditions. This study aims to analyze the determinants of regional economic growth by integrating agricultural and socio-economic indicators using a random forest–based modeling framework. Secondary panel data from 15 districts over the period 2014–2024 were analyzed, comprising 165 observations and 14 explanatory variables. The results show that agricultural production plays a dominant role, with rice production and harvested area contributing approximately 39.8% and 34.7% of total feature importance, respectively. The model achieved strong predictive performance with a coefficient of determination (R^2) of 0.68 and root mean squared error (RMSE) of 6,346.82, indicating that the selected variables explain a substantial portion of gross regional domestic product (GRDP) variation. Socio-economic factors, including poverty rate, per capita expenditure, and human development index (HDI), also contribute meaningfully to regional economic outcomes. These findings highlight the importance of integrating agricultural productivity with social development policies to achieve inclusive and sustainable economic growth. The proposed approach provides a data-driven framework to support regional policy formulation and improve food security strategies.

Corresponding Author:

Septafiansyah Dwi Putra
Department of Information Technology, Politeknik Negeri Lampung
Bandar Lampung, Indonesia
Email: septa@polinela.ac.id

1. INTRODUCTION

The agricultural sector plays a pivotal role in the economic stability and growth of nations, particularly in developing economies, such as Indonesia, by providing raw materials, ensuring food security, and improving livelihoods [1]–[3]. Indonesia, an agrarian country with a tropical climate, possesses substantial potential for agricultural development that contributes significantly to regional economic performance and gross regional domestic product (GRDP) [4]–[6]. Agriculture consistently ranks among the major contributors to the national economy, often ranking behind manufacturing and the wholesale–retail sectors [7], [8]. This economic importance underscores the need for analytical frameworks that can understand and predict the contribution of agricultural productivity to regional economic development [9], [10]. However, the agricultural sector is characterized by complex interactions among environmental conditions, production capacity, infrastructure, and socio-economic factors, making accurate economic forecasting challenging.

Traditional approaches used to analyze agricultural productivity and regional economic growth often rely on conventional econometric or statistical methods that assume linear relationships between variables [11], [12]. These models frequently fail to capture the nonlinear, multidimensional interactions among agricultural production, socioeconomic indicators, and regional development factors [13]. As a result, the predictive capacity of such models may be limited, potentially leading to incomplete or less effective policy recommendations. The increasing complexity of agricultural systems and regional economic dynamics, therefore, requires more advanced analytical methods capable of integrating diverse datasets and uncovering hidden patterns within them [14]. The development of robust predictive models that combine agricultural productivity metrics with socio-economic indicators is essential for designing sustainable agricultural policies and improving economic resilience, particularly in regions that rely heavily on agriculture for economic stability [15]. In this context, modern computational approaches and data-driven techniques offer promising alternatives for improving the accuracy and reliability of regional economic predictions.

Recent advancements in machine learning have opened new opportunities for analyzing complex systems in agriculture and economics [16]–[18]. Machine learning models, including random forest, deep learning architectures such as long short-term memory (LSTM), and graph neural networks, have demonstrated superior predictive capabilities compared to traditional statistical methods by capturing nonlinear relationships and high-dimensional data structures [13]. These techniques have been increasingly applied in agricultural studies to forecast crop yield, analyze agricultural productivity, and model economic indicators using large-scale datasets derived from environmental sensors, satellite imagery, and socio-economic databases [19]. In addition, artificial intelligence–driven analytical frameworks have shown potential to improve supply chain efficiency, mitigate risks associated with climate variability, and support data-driven decision-making in the agricultural sector [20].

Despite the growing adoption of machine learning in agricultural research, most existing studies focus primarily on predicting agricultural outputs or crop yields without integrating broader socio-economic indicators that influence regional economic performance [21]. Meanwhile, economic growth and food security are closely interconnected dimensions of sustainable development, particularly in developing countries, such as Indonesia [22], [23]. Economic growth can enhance welfare, generate employment opportunities, and reduce poverty; however, growth that is not accompanied by improvements in food security may increase social inequality and exacerbate household vulnerability [24]. Food security extends beyond food availability to include access, utilization, and the stability of the food supply across society. When communities experience limited access to food or unstable food systems, labor productivity may decline, public health outcomes may deteriorate, and socio-economic pressures may intensify, creating a persistent cycle of poverty and food insecurity [25].

Lampung Province represents a particularly important case for examining the relationship between agricultural productivity and regional economic growth. As one of Indonesia's major agricultural regions and a national food production center, Lampung plays a significant role in supporting the national food supply and regional economic development. Nevertheless, the province continues to face structural challenges to food security, including regional development disparities, limited agricultural infrastructure, and fluctuations in food commodity prices. These challenges often disproportionately affect smallholder farmers and rural communities, making the regional economy vulnerable to external shocks such as climate change and global market volatility [26]. Although Lampung has experienced relatively positive economic growth trends in recent years, food security challenges—particularly those related to access, distribution, and price stability—remain unresolved.

Existing research examining the relationship between economic growth and food security at the regional level in Indonesia remains relatively limited, especially studies that employ data-driven and machine learning approaches to analyze integrated socio-economic and agricultural indicators. The lack of comprehensive provincial-level empirical analyses often leads to policy decisions that lack sufficient data-driven support. Consequently, policies designed to address food security and economic growth simultaneously may fail to achieve optimal outcomes. This situation highlights the importance of developing analytical models that incorporate local data and capture the multidimensional relationships among agricultural productivity, socio-economic welfare, and regional economic performance to address this research gap, this study aims to analyze the relationship between economic growth and food security by integrating agricultural productivity indicators with socio-economic variables using a machine learning approach. Specifically, this research seeks to: i) examine the relationships among food security indicators and socio-economic conditions in Lampung Province, ii) identify the key determinants influencing regional GRDP using random forest–based feature importance analysis, and iii) evaluate the predictive performance of selected indicators in modeling regional economic growth.

The findings of this study are expected to provide empirical evidence supporting more inclusive and sustainable regional development strategies. By identifying the most influential determinants of economic growth and food security, the results can inform data-driven policy interventions to strengthen agricultural productivity, improve social welfare, and enhance economic resilience. Ultimately, this research provides valuable insights for policymakers, researchers, and development practitioners seeking to promote sustainable food systems and equitable economic development in agrarian regions such as Lampung Province.

2. METHOD

This study adopts a quantitative and data-driven approach to examine the relationship between economic growth, agricultural productivity, and socio-economic conditions in Lampung Province, Indonesia. The analysis utilizes secondary panel data obtained from the Indonesian Central Bureau of Statistics (Badan Pusat Statistik (BPS)) covering the period 2014–2024 across 15 regencies and municipalities. The dataset consists of economic, agricultural, and socio-economic indicators that represent the multidimensional aspects of food security and regional development.

The dependent variable in this study is GRDP, which serves as an indicator of regional economic growth. The independent variables include fourteen indicators related to food security and socio-economic conditions. These indicators were selected based on the global food security framework and its local adaptation in Indonesia. The variables include rice production, rice productivity, harvested rice area, percentage of food expenditure, access to safe drinking water, implicit price index, poverty rate, unemployment rate, per capita expenditure, Gini ratio, human development index (HDI), non-food expenditure, percentage of non-food expenditure, and percentage of food expenditure.

The dataset contains 165 observations derived from the combination of 15 districts and 11 years of observation. Each observation includes 15 variables (1 dependent variable and 14 independent variables), resulting in a total of 2,475 data entries. Before model development, the dataset underwent a preprocessing stage to ensure consistency and comparability across variables and time periods. Figure 1 shows the research methodology framework.

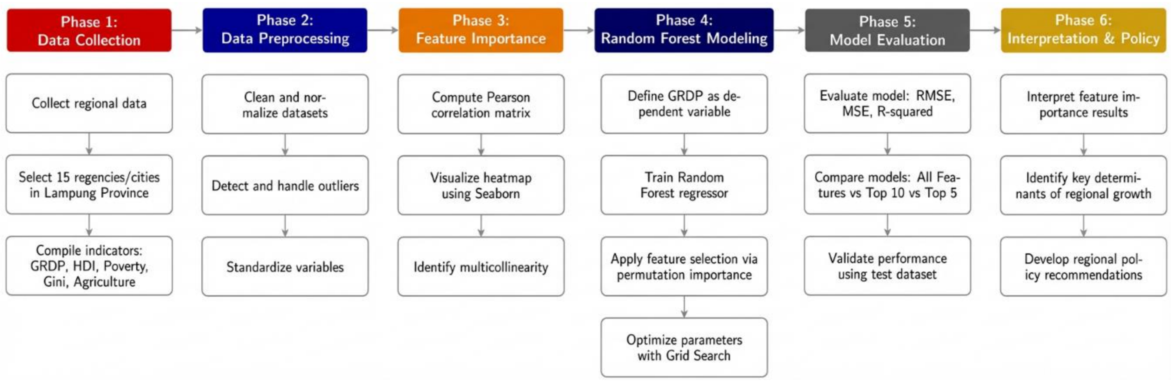

Figure 1. Research methodology framework

2.1. Data preprocessing

Data preprocessing is an essential step in machine learning analysis to ensure that the dataset is suitable for modeling. First, the dataset was examined for missing values and inconsistencies across the observation period. Because the data were obtained from official statistical publications, the dataset showed minimal missing values. Any minor inconsistencies were handled through interpolation using adjacent year values to preserve temporal continuity. Second, all variables were standardized using z-score normalization to ensure that differences in measurement units did not bias the machine learning model. Standardization transforms the original data into a normalized scale with a mean of zero and a standard deviation of one. The normalization process can be expressed as (1).

$$X' = \frac{X-\mu}{\sigma} \tag{1}$$

Where X represents the original data value, μ is the mean of the data, σ is the standard deviation of the data, and X' is the value of the data after standardization. In addition, multicollinearity among predictors was examined using a correlation matrix. Highly correlated variables, particularly those related to household expenditure composition, were carefully evaluated to avoid redundancy in the modeling process. The correlation analysis served as an exploratory step to understand the relationships among variables before the machine learning modeling stage. The research indicators are presented in Table 1.

Table 1. Variables and research indicators

No	Variable	Indicator	Variable code
1	GRDP	The value of GRDP as an indicator of economic growth	Y
2	Food security	Rice production	X1
		Rice productivity	X2
		Harvested rice area	X3
		Percentage of food expenditure	X4
		Access to clean water	X5
		Implicit price index	X6
3	Socio-economic factors	Poverty rate	X7
		Unemployment rate	X8
		Per capita expenditure	X9
		Gini ratio	X10
		HDI	X11
		Non-food expenditure	X12
		Percentage of per capita expenditure on non-food	X13
		Percentage of per capita expenditure on food	X14

2.2. Random forest modeling

To capture nonlinear relationships between agricultural productivity, socio-economic indicators, and regional economic growth, this study employs the random forest algorithm proposed by Sinuraya *et al.* [14]. Random forest is an ensemble learning method that constructs multiple decision trees using bootstrap sampling and aggregates their predictions to produce a more stable and accurate model. The prediction of the random forest model for regression problems is defined as the average of predictions from all decision trees, as in (2).

$$\hat{y} = \frac{1}{B}\sum_{b=1}^{B} T_b(x) \quad (2)$$

Where $\hat{y}$ represents the prediction from the *i*-th decision tree, B is the total number of trees in the random forest, $T_b(x)$ is the average of all tree predictions (for regression). Model development was conducted using Python 3.10 with several data science libraries, including pandas for data manipulation, NumPy for numerical computation, seaborn for visualization, and scikit-learn for machine learning. To improve model robustness and prevent overfitting, hyperparameter tuning was performed using grid search combined with k-fold cross-validation. The main parameters optimized in the random forest model include:

i) Number of trees (*n_estimators*)
ii) Maximum tree depth (*max_depth*)
iii) Minimum samples required for splitting nodes (*min_samples_split*)
iv) Minimum samples required at leaf nodes (*min_samples_leaf*)
v) Number of features considered at each split (*max_features*)

The dataset was divided into training and testing subsets using an 80:20 ratio. The training set was used to train the random forest model and tune hyperparameters, while the testing set was used to evaluate the model's predictive performance on unseen data.

2.3. Feature importance analysis

In addition to predictive modeling, this study applies permutation-based feature importance to identify the most influential variables affecting GRDP. Permutation importance measures the contribution of each predictor by randomly shuffling its values and observing the resulting change in model performance. A larger decrease in prediction accuracy indicates a more important feature. This approach was selected because it is model-agnostic and provides a more reliable measure of variable importance compared to impurity-based methods, particularly in datasets containing correlated predictors. The results of the feature importance analysis are used to identify the most relevant indicators that explain regional economic growth and food security conditions.

2.4. Model evaluation

The performance of the predictive model was evaluated using several regression evaluation metrics. These metrics provide complementary perspectives on prediction accuracy and model reliability. The root mean squared error (RMSE) is used to measure the magnitude of prediction errors and is defined as (3).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y_i})^2} \quad (3)$$

Where n represents the total number of observations, y_i denotes the actual value of GRDP, $\widehat{y_i}$ represents the predicted value from the model. In addition, the coefficient of determination (R^2) was used to evaluate the explanatory power of the model. R^2 measures the proportion of variance in the dependent variable that can be explained by the independent variables and is expressed as (4).

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \widehat{y_i})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (4)$$

Where $\bar{y}$ represents the mean of the actual values. To ensure robustness, model evaluation was conducted using both training and testing datasets, and cross-validation results were examined to assess the stability of model performance.

2.5. Limitations of data and model for interpretation or policy

Although the dataset provides valuable insights into regional economic dynamics, several limitations should be acknowledged. First, the study relies on secondary statistical data, which may not capture short-term fluctuations in agricultural productivity caused by climate variability or extreme weather events. Second, the dataset focuses primarily on rice-related indicators because rice is the dominant staple crop in Lampung; therefore, other agricultural commodities may not be fully represented in the model. Finally, while random forest is effective for capturing nonlinear relationships, the model does not explicitly provide causal interpretations, and the results should therefore be interpreted primarily as predictive and exploratory insights rather than causal conclusions.

3. RESULTS AND DISCUSSION

Building upon the analytical framework described in the previous section, this section presents and interprets the empirical findings obtained from the correlation matrix analysis and the random forest modeling. The objective of this analysis is to identify the main determinants of regional economic growth and food security in Lampung Province. The results are structured to provide both statistical evidence and substantive economic interpretation. The analysis begins with correlation matrix diagnostics to identify the strength and direction of relationships among socio-economic and agricultural indicators. This step serves as an exploratory analysis to detect potential associations and multicollinearity among predictors before implementing machine learning models. Subsequently, random forest modeling is applied to quantify the relative importance of each variable in predicting GRDP. This combined approach enables a deeper understanding of how agricultural productivity, welfare indicators, and infrastructure jointly influence regional economic performance.

3.1. Analysis of gross regional domestic product, food security, and socio-economic conditions

The correlation matrix analysis of the dataset for Lampung Province during the period 2014–2024 reveals several statistically meaningful relationships among economic, agricultural, and socio-economic indicators (Figure 2). Each value in the matrix represents the Pearson correlation coefficient (r), which measures the strength and direction of linear relationships between variables within a range of −1 to +1 [16]. The analysis highlights a strong relationship between welfare indicators and economic development. GRDP exhibits a positive correlation with several socio-economic indicators, particularly per capita expenditure (r =0.42) and the HDI. More notably, per capita expenditure shows a very strong correlation with HDI (r =0.91), suggesting that higher household income levels are closely associated with improvements in education and health outcomes. This relationship reflects a widely observed pattern in development economics where economic growth contributes to human capital development and improved living standards.

Infrastructure indicators also demonstrate meaningful relationships with welfare outcomes. Access to safe drinking water shows a moderate positive correlation with HDI (r =0.55), indicating that improved access to basic infrastructure contributes to better public health conditions and overall quality of life. Conversely, these welfare indicators display negative correlations with the percentage of the poor population, suggesting that improvements in income, infrastructure, and human development tend to coincide with reductions in poverty levels. Household consumption patterns further illustrate economic behavioral dynamics consistent with Engel's Law. The correlation matrix reveals an almost perfect negative relationship (r ≈−1.00) between the percentage of expenditure on food and the percentage of expenditure on non-food items. This pattern occurs because both variables represent complementary components of total household expenditure. Nevertheless, the relationship also reflects a well-established economic phenomenon in which rising income levels reduce the proportion of household expenditure allocated to food while increasing spending on non-essential goods and services such as education, healthcare, and recreation. As indicated by the positive relationship between per capita expenditure and the share of non-food consumption (r =0.70), higher-income households allocate a greater share of their budgets to services and investments that enhance long-term welfare. From a modeling perspective, this near-perfect correlation implies a risk of multicollinearity if both variables are included simultaneously in predictive models. Multicollinearity can inflate variance and reduce the reliability of coefficient estimates, which justifies the use of feature selection techniques during the machine learning stage [27].

In the agricultural sector, rice production exhibits a very strong positive correlation with harvested area (r =0.99) and a relatively weak correlation with rice productivity (r =0.29). This pattern indicates that variations in total rice production in Lampung are primarily driven by changes in cultivated land area rather than improvements in yield per hectare. The finding suggests that agricultural expansion has historically played a larger role than technological intensification in increasing regional rice output. Agricultural productivity also shows a modest relationship with welfare indicators. Rice productivity demonstrates a weak negative correlation with the poverty rate (r =−0.23), implying that improvements in agricultural efficiency may contribute to poverty reduction in rural areas. Although the magnitude of this relationship is relatively small, it aligns with previous studies showing that increased agricultural productivity can enhance rural incomes and strengthen food security among farming communities [28]. Overall, the correlation analysis provides important preliminary insights into the structural relationships among agricultural production, household welfare, and regional economic development. These findings serve as the empirical foundation for the subsequent machine learning analysis.

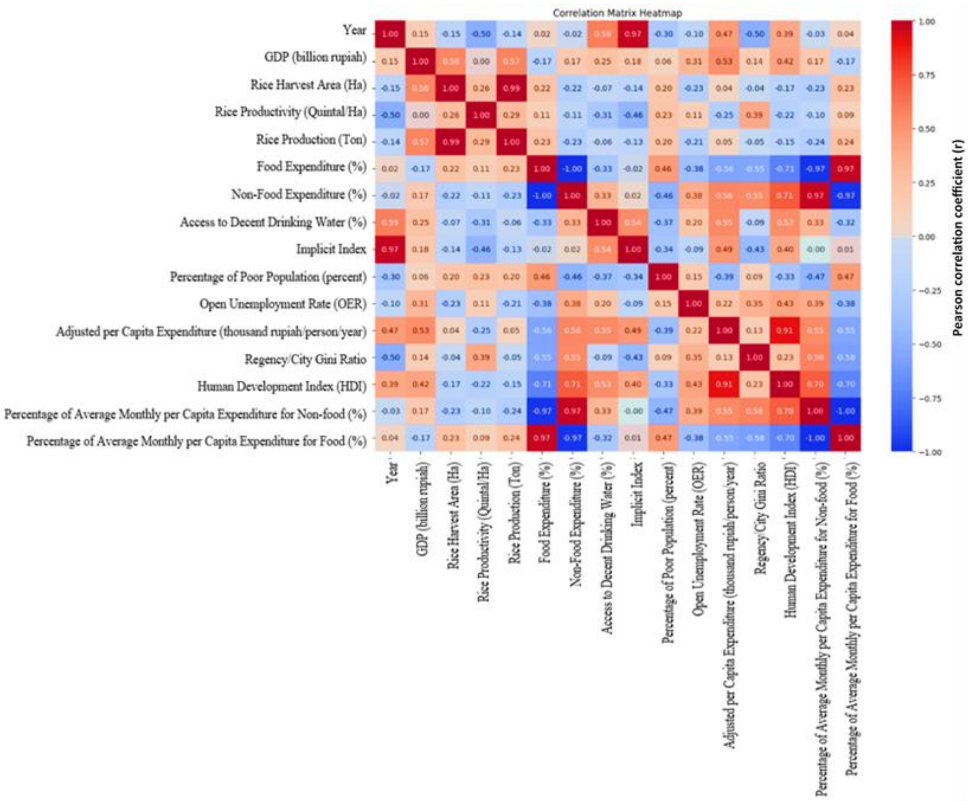

Figure 2. Heatmap of the correlation matrix of regional indicators in Lampung Province, 2014–2024

3.2. Feature importance analysis

To further investigate the determinants of regional economic growth, random forest modeling was applied to evaluate the relative importance of each predictor variable in explaining variations in GRDP. Random forest is particularly suitable for multidimensional socio-economic datasets because it can capture nonlinear relationships and interactions among variables without requiring strict distributional assumptions. We evaluated the predictive performance of the model using RMSE on the test dataset. As shown in Table 2, we compared three model configurations: a model using all features, a model using the top 5 features, and a model using the top 10 features selected through permutation importance.

Table 2. Test set RMSE evaluation

Feature permutation	RMSE
Top 10 features (permutation)	6,346.82
Top 5 features (permutation)	6,439.56
All features	6,872.47

The results demonstrate that the model using the top-ten features achieved the lowest prediction error (RMSE =6,346.82), outperforming both the full-feature model and the reduced top-five model. Compared with the model using all variables, the top-ten features model reduced prediction error by approximately 7.6%, indicating that removing less informative variables improved model generalization. This outcome is consistent with theoretical findings in machine learning research showing that feature selection can reduce model complexity, minimize redundancy, and mitigate multicollinearity, particularly when predictors are strongly correlated. Permutation importance was used to identify influential variables because it evaluates each predictor's contribution by measuring the decline in model performance when the variable is randomly permuted. The performance of the final regression model based on the selected predictors is presented in Table 3.

Table 3. Evaluate the performance of the trained regression model on the test data using appropriate metric

Evaluation methods	Score
R^2 score	0.68
Mean squared error (MSE)	40,282,062.84
RMSE	6,346.82

The model achieved an R^2 value of 0.68, indicating that approximately 68% of the variation in GRDP across districts and time periods can be explained by the selected socio-economic and agricultural indicators. In the context of regional economic modeling, this level of explanatory power is considered relatively strong given the inherent complexity and variability of socio-economic systems. The selected predictors include key agricultural indicators (rice production, harvested area, and productivity), welfare indicators (poverty rate, per capita expenditure, and HDI), labor market conditions (unemployment rate), inequality (Gini ratio), infrastructure (access to safe drinking water), and price indicators (implicit price index). These variables collectively capture multiple dimensions of regional development, including production capacity, human capital, economic distribution, and infrastructure accessibility.

3.3. Interpretation of feature importance

The ranking of feature importance derived from the random forest model is illustrated in Figure 3. The analysis reveals that rice production and harvested area are the two most influential predictors, with relative importance scores of approximately 40% and 35%, respectively. These findings highlight the dominant role of agricultural production capacity in shaping regional economic performance in Lampung.

Socio-economic indicators such as the poverty rate, per capita expenditure, and HDI also contribute meaningfully to the prediction of GRDP, although their relative influence is lower compared to agricultural production variables. This suggests that while agricultural output remains the primary driver of regional economic growth, improvements in welfare, human capital, and income distribution also play important supporting roles. Other variables, including the open unemployment rate, rice productivity, and access to safe drinking water, exhibit moderate contributions to the model. In contrast, indicators such as the Gini ratio, implicit price index, and household expenditure composition have relatively smaller importance values.

From a regional development perspective, these findings suggest that economic differentiation among districts in Lampung is strongly influenced by differences in agricultural production capacity, particularly the scale of cultivated land and total rice output. The relatively lower importance of productivity per hectare indicates that regional economic growth has historically been driven more by the expansion of cultivated land than by technological intensification. However, relying solely on agricultural expansion may pose long-term sustainability challenges, particularly in the context of land limitations and climate change. Therefore, policies aimed at strengthening regional food security and economic resilience should combine production expansion strategies with productivity improvements and socio-economic development initiatives. Overall, the results demonstrate that agricultural productivity, human development, and socio-economic welfare collectively shape regional economic performance in Lampung Province. The integration of machine learning techniques with socio-economic indicators thus provides valuable insights for designing evidence-based policies aimed at achieving sustainable and inclusive regional development.

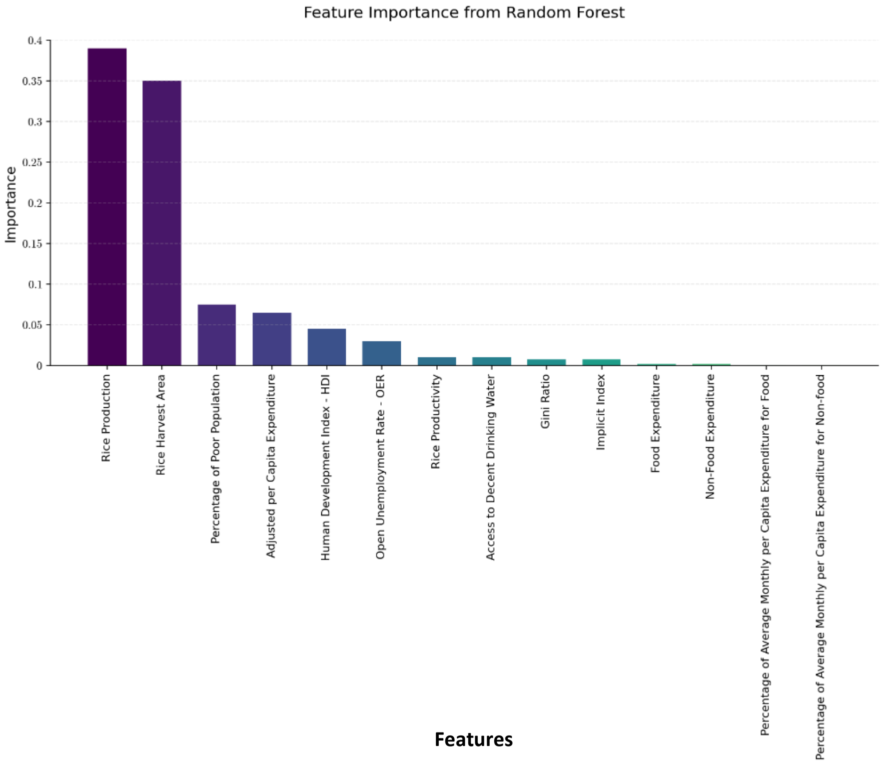

Figure 3. The ranking of feature importance derived from the random forest

3.4. Robustness analysis

To ensure the reliability and stability of the predictive results, we conducted several robustness checks. These analyses aim to verify that the predictive performance and variable importance remain consistent under different modeling conditions and are not driven by overfitting or sampling bias. First, we evaluated model stability using k-fold cross-validation (k =5) during the training phase. We partitioned the dataset into five equally sized folds, with four-folds used for training and one-fold used for validation in each iteration. The cross-validation results showed a mean RMSE of 6,412.37 with a standard deviation of 214.65, indicating that the prediction error remained stable across different data partitions.

Similarly, the average cross-validated R^2 score reached 0.66±0.03, demonstrating consistent explanatory power across validation folds. These results suggest that the model generalizes well and is not sensitive to a particular train–test split. Second, we assessed robustness by comparing model performance under different feature selection scenarios. As shown in Table 2, the model using the top ten features identified through permutation importance achieved the lowest prediction error (RMSE =6,346.82), outperforming the model using all predictors (RMSE =6,872.47). This represents a 7.6% reduction in prediction error, indicating that eliminating redundant variables improves model generalization. The model using the top five features produced a slightly higher RMSE (6,439.56), suggesting that excessive dimensionality reduction may omit relevant predictive information. Third, the stability of feature importance rankings was examined across repeated random forest runs. The importance scores for the two dominant predictors—rice production and harvested area—remained consistently above 70% of the total cumulative importance, with average scores of approximately 39.8% and 34.7%, respectively. The remaining predictors, including poverty rate, per capita expenditure, and HDI, individually contributed between 3% and 7% to the model's predictive performance. The consistency of these rankings across model iterations confirms that the dominance of agricultural production variables is not a random artifact but reflects a stable structural relationship in the dataset.

Finally, a sensitivity analysis was conducted for highly correlated predictors identified in the correlation matrix. Variables representing household expenditure composition—specifically the percentage of food and non-food expenditure—showed an almost perfect negative correlation ($r \approx -1.00$). To avoid redundancy and instability caused by multicollinearity, the feature selection process retained only the most informative expenditure-related indicators. This procedure improved the model's predictive stability while preserving the explanatory capacity of socio-economic variables. Overall, the robustness analysis confirms that the model results remain consistent across different data partitions, feature configurations, and model iterations. The relatively low variance in RMSE and the stability of feature importance rankings provide strong evidence that the identified determinants—particularly rice production, harvested area, and key socio-economic indicators—represent reliable predictors of regional economic growth in Lampung Province.

4. CONCLUSION

This study examined the relationships among agricultural productivity, socio-economic indicators, and regional economic growth in Lampung Province by integrating correlation analysis and random forest modeling using data from 15 districts during 2014–2024. The results show that agricultural production capacity is the dominant determinant of GRDP, with rice production and harvested area contributing approximately 39.8% and 34.7% to total feature importance, respectively, while poverty rate, per capita expenditure, and HDI also provide meaningful contributions. The model achieved satisfactory predictive performance, with an R^2 of 0.68 and an RMSE of 6,346.82, and robustness analysis confirmed stable feature-importance rankings and consistent model performance. These findings indicate that integrating agricultural and socio-economic indicators within a machine-learning framework can provide useful data-driven support for regional planning, particularly in strengthening agricultural productivity, infrastructure development, poverty reduction, and sustainable economic growth. Future research may extend this approach by incorporating climate variables, spatial analysis, satellite imagery, and multi-commodity agricultural data to improve predictive accuracy and policy relevance.

ACKNOWLEDGMENTS

In carrying out this study, the authors would like to express their sincere appreciation to all parties who provided valuable assistance and support throughout the research process. Special thanks are extended to the Central Bureau of Statistics (Badan Pusat Statistik – BPS RI) for supplying essential regional data, to the Directorate General of Vocational Education – Ministry of Education, Culture, Research, and Technology (Kemdiktisaintek) for continuous guidance and academic facilitation, and to the Master’s Program in Food Security, Politeknik Negeri Lampung (POLINELA) for institutional and technical support. Their collective contributions significantly enriched the quality and completion of this work.

FUNDING INFORMATION

This research was funded by the Directorate General of Research and Development, Ministry of Higher Education, Science, and Technology, Republic of Indonesia, under research contract No. 229/PL15.8/PP/2025, dated 7 August 2025, with reference to main contract No. 174/C3/DT.05.00/PL-BATCH II/2025, dated 14 July 2025. The study was conducted under the master's thesis research program within the framework of the National Competitive Research Scheme, managed by Politeknik Negeri Lampung.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Septafiansyah Dwi Putra	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓
Feri Utomo	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓			
Fitriani	✓		✓	✓			✓			✓	✓		✓	
Irmayani Noer	✓	✓		✓						✓				
Heriansyah	✓	✓				✓				✓	✓			

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, [SDP], upon reasonable request. The datasets were obtained from publicly accessible government sources, including the Central Bureau of Statistics (Badan Pusat Statistik – BPS RI), and were further processed and analyzed under the research framework of Politeknik Negeri Lampung (POLINELA).

REFERENCES

[1] H. Firdaus, U. Ciptomulyono, T. Widjaya, and D. Ramdani, "Electrifying agriculture: a systematic review of sustainable approaches to enhancing national food security in Indonesia," *BIO Web of Conferences*, vol. 192, Oct. 2025, doi: 10.1051/bioconf/202519205001.

[2] H. Rasyid and G. M. Ningsih, "Application of sustainable agricultural technology to increase productivity and food security," *Journal of World Science*, vol. 3, no. 4, pp. 500–506, Apr. 2024, doi: 10.58344/jws.v3i4.594.

[3] A. S. Azahra, M. D. Johansyah, and Sukono, "Agricultural insurance premium determination model for risk mitigation based on rainfall index: systematic literature review," *Risks*, vol. 12, no. 12, Dec. 2024, doi: 10.3390/risks12120205.

[4] D. H. S. Keefe, H. Jang, and J.-M. Sur, "Digitalization for agricultural supply chains resilience: perspectives from Indonesia as an ASEAN member," *The Asian Journal of Shipping and Logistics*, vol. 40, no. 4, pp. 180–186, Dec. 2024, doi: 10.1016/j.ajsl.2024.09.001.

[5] S. F. Purba *et al.*, "The contribution of agricultural crop production towards the economic growth of Indonesia's agricultural sector," *E3S Web of Conferences*, vol. 444, Nov. 2023, doi: 10.1051/e3sconf/202344402034.

[6] S. Susanawati and N. Noviyanti, "Supply chain efficiency of red chilies in the production center of Sleman Indonesia based on performance measurement system," *Open Agriculture*, vol. 9, no. 1, Jul. 2024, doi: 10.1515/opag-2022-0294.

[7] S. V. Hoseini, J. Suutala, J. Partala, and K. Halunen, "Threat modeling AI/ML with the attack tree," *IEEE Access*, vol. 12, pp. 172610–172637, 2024, doi: 10.1109/ACCESS.2024.3497011.

[8] T. Tom, M. M. Nirmala, N. S. Kumar, and U. P. Reshmi, "Unveiling the agricultural growth catalyst: a review on the importance of public expenditure in the agriculture sector," in *Anticipating Future Business Trends: Navigating Artificial Intelligence Innovations*, Cham, Switzerland: Springer, 2024, pp. 243–256, doi: 10.1007/978-3-031-63402-4_19.

[9] Y. M. Baion, C. Jianmin, and W. Yunxia, "The determinants of agriculture total factor productivity dynamics in Sub-Saharan Africa – evidence from 44 countries (1961-2019)," *Research Square*, Apr. 2024, doi: 10.21203/rs.3.rs-4301735/v1.

[10] Y. Maiga, "The impact of agricultural productivity on economic growth in Africa," *Journal of Agricultural Studies*, vol. 12, no. 2, Apr. 2024, doi: 10.5296/jas.v12i2.21680.

[11] S. B. Dhal and D. Kar, "Transforming agricultural productivity with AI-driven forecasting: innovations in food security and supply chain optimization," *Forecasting*, vol. 6, no. 4, pp. 925–951, Oct. 2024, doi: 10.3390/forecast6040046.

[12] M. Kulisz *et al.*, "Implications of neural network as a decision-making tool in managing Kazakhstan's agricultural economy," *Applied Computer Science*, vol. 19, no. 4, pp. 121–135, Jan. 2024, doi: 10.35784/acs-2023-39.

[13] B. S. Saruk and G. M. Rayalu, "Integrating machine learning for enhanced agricultural productivity: a focus on bananas and arecanut in the context of India's economic growth," *Journal of Statistical Theory and Applications*, vol. 23, no. 4, pp. 408–426, Oct. 2024, doi: 10.1007/s44199-024-00090-y.

[14] J. F. Sinuraya, A. Ulpah, A. Setiyanto, A. F. Astari, and F. B. M. Dabukke, "Technical efficiency of rice productivity in Indonesia," *BIO Web of Conferences*, vol. 119, Jul. 2024, doi: 10.1051/bioconf/202411901010.

[15] B. Almadani and S. M. Mostafa, "IIoT based multimodal communication model for agriculture and agro-industries," *IEEE Access*, vol. 9, pp. 10070–10088, 2021, doi: 10.1109/ACCESS.2021.3050391.

[16] S. Du, Y. Xu, and L. Wang, "Predicting economic resilience: a machine learning approach to rural development," *Alexandria Engineering Journal*, vol. 121, pp. 193–200, May 2025, doi: 10.1016/j.aej.2025.02.049.

[17] L. Maruejols, X. Yu, and L. Chenarides, "Machine learning in applied economics and agribusiness: emerging applications and integration with traditional methods," *International Food and Agribusiness Management Review*, vol. 28, no. 2, pp. 237–240, Jun. 2025, doi: 10.22434/ifamr.0001.

[18] J. Tenorio and W. Perez, "Monthly GDP nowcasting with machine learning and unstructured data," Feb. 2024, *arXiv: 2402.04165*.

[19] G. Manogaran, C.-H. Hsu, B. S. Rawal, B. Muthu, C. X. Mavromoustakis, and G. Mastorakis, "ISOF: information scheduling and optimization framework for improving the performance of agriculture systems aided by Industry 4.0," *IEEE Internet of Things Journal*, vol. 8, no. 5, pp. 3120–3129, Mar. 2021, doi: 10.1109/JIOT.2020.3045479.

[20] M. K. Bashir and S. Schilizzi, "Determinants of rural household food security: a comparative analysis of African and Asian studies," *Journal of the Science of Food and Agriculture*, vol. 93, no. 6, pp. 1251–1258, Apr. 2013, doi: 10.1002/jsfa.6038.

[21] R. R. S. M. Sudarsono and H. Wandebori, "Machine learning predictive modeling of agricultural sustainability indicators," *Indonesian Journal of Applied Statistics*, vol. 6, no. 1, Jan. 2024, doi: 10.13057/ijas.v6i1.64245.

[22] Sulaeman, H. Mirzal, M. S. Mubarok, I. F. Ulfah, M. A. Nugroho, and S. Rusgianto, "Enhancing food security through agricultural financing and loans within Indonesia's dual banking system: empirical study approaches," *Regional Statistics*, vol. 15, no. 5, pp. 852–874, 2025, doi: 10.15196/RS150502.

[23] W. Wahyudin, S. Tan, J. Junaidi, Z. Zulgani, and F. Zevaya, "Dynamic model in analyzing Indonesia's agricultural sector's macroeconomic performance," *Journal of Infrastructure Policy and Development*, vol. 8, no. 8, Aug. 2024, doi: 10.24294/jipd.v8i8.5661.

[24] J. Fanzo *et al.*, "The food systems dashboard is a new tool to inform better food policy," *Nature Food*, vol. 1, no. 5, pp. 243–246, May 2020, doi: 10.1038/s43016-020-0077-y.

[25] D. Headey and M. Ruel, "Food inflation and child undernutrition in low and middle income countries," *Nature Communications*, vol. 14, no. 1, Sep. 2023, doi: 10.1038/s41467-023-41543-9.

[26] C. Béné *et al.*, "Global drivers of food system (un)sustainability: a multi-country correlation analysis," *PLoS ONE*, vol. 15, no. 4, Apr. 2020, doi: 10.1371/journal.pone.0231071.

[27] K. P. Vatcheva and M. Lee, "Multicollinearity in regression analyses conducted in epidemiologic studies," *Epidemiology (Sunnyvale)*, vol. 06, no. 02, 2016, doi: 10.4172/2161-1165.1000227.

[28] M. A. Akudugu, K. K. Millar, and S. Mubarik, "Crop productivity and implications for food security and rural livelihood development in Africa," *Preprints*, Jan. 2020, doi: 10.20944/preprints202001.0327.v1.

BIOGRAPHIES OF AUTHORS

Septafiansyah Dwi Putra is a certified professional engineer (IPM, ASEAN Eng). His research focuses on smart agriculture, internet of things (IoT), blockchain applications, and machine learning for precision farming and sustainable development. He received the bachelor's (S.T.) degree in Electrical Engineering from Lampung University and the master's degree (cum laude) in Computer Engineering and the Dr. degree (cum laude) in Electrical Engineering and Informatics from Institut Teknologi Bandung (ITB). He is currently a lecturer and researcher with the Cybersecurity Research Group in the Department of Information Technology, Internet Engineering Technology Program at Politeknik Negeri Lampung, Lampung, Indonesia. He also holds several professional certificates in cybersecurity and computer system security. He can be contacted at email: septa@polinela.ac.id.

Feri Utomo is a postgraduate student in the Master of Applied Food Security at Politeknik Negeri Lampung. His research focuses on regional food economics, particularly the development of analytical models to support sustainable food systems and local economic resilience. Throughout his academic journey, he has conducted studies exploring the dynamics of regional food supply chains, the influence of agricultural productivity on local economies, and the application of machine learning in regional food security modeling. He has a strong interest in integrating data-driven approaches with policy analysis to improve food security and regional development planning. His dedication to research reflects his commitment to contributing to Indonesia's sustainable food and agricultural sector. He can be contacted at email: 237021010@student.polinela.ac.id.

Fitriani is a highly respected academic at Politeknik Negeri Lampung (Polinela), Bandar Lampung, Indonesia, specializing in Agricultural Economics. She is a Professor of Agricultural Economics and a lecturer in the Master of Applied Food Security Program, Politeknik Negeri Lampung. As an expert in sustainable agricultural economics and agribusiness, she is actively involved in research aimed at strengthening the local economy through sustainable agribusiness development and food security. Her expertise is reflected in numerous publications on sustainable agribusiness and production economics. She can be contacted at email: fitriani@polinela.ac.id.

Irmayani Noer is a distinguished senior lecturer in the Master of Applied Food Security Program, Politeknik Negeri Lampung (Polinela), Bandar Lampung, Indonesia, where she has been a member of the faculty since 1994. Her academic journey includes earning her undergraduate degree from the University of Lampung, followed by a Master of Science in Agricultural Economics from Bogor Agricultural University (IPB). In 2017, she further distinguished herself by completing her doctoral studies in Agricultural Sciences at Universitas Padjadjaran, graduating with cum laude honors. Currently, she holds the functional position of Associate Professor. She can be contacted at email: irmayani_noer@polinela.ac.id.

Heriansyah is affiliated with the Department of Telecommunication Engineering, Johannes Kepler University of Linz, Austria. He is also registered as a lecturer in the Department of Electrical Engineering, Institut Teknologi Sumatera, Lampung, Indonesia. He holds a background in electronic and communication systems engineering, with research interests spanning wireless communication, signal processing, and embedded systems for smart infrastructure. His current research focuses on next-generation communication networks and their integration with data-driven intelligent systems. He can be contacted at email: heriansyah.heriansyah@jku.at.