Journal of Applied Informatics and Computing (JAIC)
Vol.10 , No.2 , April 2026, pp. 2066~2075
e-ISSN: 2548-6861

Explainable Machine Learning for Food Vulnerability Prediction in Indonesia

Marizka Riffiy Alfiyah [1]*, Fadiah Tussholiha [2]*, Halona Daffakhansa Nabila [3]*, Nyimas Princessa Syapeni [4]*, Ken Ditha Tania [5]*, Allsela Meiriza [6]*
* Information Systems, Universitas Sriwijaya
marizkariffiyalfiyah@gmail.com [1], tussfadiah@gmail.com [2], halonadn11@gmail.com [3], princessasyapeni@icloud.com [4], kenya.tania@gmail.com [5], allsela_meiriza@yahoo.co.id [6]

Article Info

Article history:

Received 2026-03-06
Revised 2026-04-20
Accepted 2026-04-29

Keyword:

Food Vulnerability,
Indonesia,
Machine Learning,
Random Forest,
SHAP.

ABSTRACT

Food insecurity remains a strategic multidimensional issue in Indonesia, requiring precise and transparent predictive frameworks for evidence-based policy. This study develops an explainable machine learning framework to predict interregional food vulnerability using data from the Food Security Index (IKP) and the Food Security and Vulnerability Atlas (FSVA) for the 2022–2024 period, encompassing 514 districts. To ensure model optimality and respond to the need for robust comparison, seven algorithms were evaluated, including ensemble-boosting and neural network techniques. The Gradient Boosting model demonstrated the most superior and stable performance, achieving an R^2 of 0.9770, MAE of 1.5621, and RMSE of 2.1534, outperforming Random Forest and XGBoost. Model reliability was further validated through K-fold Cross-Validation (CV $R^2 = 0.966$), confirming high generalizability and the absence of significant overfitting. Model interpretability was achieved through SHapley Additive exPlanations (SHAP), identifying the Net Consumption to Production Ratio (NCPR) as the dominant global driver, followed by clean water access and poverty levels. Localized analysis reveals that in high-risk regions like Papua, infrastructure gaps and food supply dependence are the primary catalysts for vulnerability. This study provides a high-precision, validated predictive model that enables policymakers to implement targeted mitigation strategies according to regional disparities, supporting national goals for sustainable food sovereignty.

I. INTRODUCTION

Food security stands as a strategic pillar of Indonesia's national sovereignty, legally mandated to ensure public access to safe and nutritious food. As an archipelagic nation, Indonesia faces complex challenges in production and logistical disparities. Despite achieving a historic milestone in 2025 with rice self-sufficiency and a surplus of 3.52 million tons, regional food vulnerability remains a serious threat due to climate change and global economic instability [1]. Phenomena such as agricultural land degradation and the urgent need for climate change adaptation strategies necessitate a more adaptive and precise monitoring system to achieve Zero Hunger targets under the Sustainable Development Goals (SDGs) [2]. The need for a more advanced framework has driven the application of big data and data mining techniques to deepen national food security analysis [3]. Previous studies have identified that regional vulnerability is heavily influenced by an area's adaptive capacity [4] and have utilized clustering methods to group food security indicators across the archipelago [5].

At the local level, resilience is built through diverse interventions. Recent evaluations have analyzed the effectiveness of social cash assistance in maintaining food access during disasters [6], while community empowerment through village-level programs significantly increases grassroots resilience [7]. In regions like Papua, strengthening local commodities such as sago and sweet potatoes is identified as a key driver for sustainability [8]. From a technological perspective, machine learning is increasingly used to predict the impact of food insecurity on nutrient consumption and the risk of malnutrition in children [9], as well as analyzing how recall periods affect prediction accuracy [10].

In the context of algorithm selection, model stability and accuracy are top priorities. While previous technical benchmarks often highlighted Random Forest for its competitive performance in handling complex data variances [11], [12], recent studies suggest that boosting-based ensembles can offer superior precision for socio-economic indicators [13]. To ensure the selection of the most robust framework, this study expands the scope of evaluation by comparing seven different algorithms, including bagging, ensemble-boosting (XGBoost, Gradient Boosting, LightGBM), and neural networks [14]. Such performance is crucial for predicting indicators like poverty and economic status, which are fundamental to food access [15], [16].

However, the primary challenge in implementing AI for public policy is the "black-box" nature of machine learning. To bridge this gap, Explainable AI (XAI) through the SHapley Additive exPlanations (SHAP) method has become urgent. SHAP has proven effective in providing transparency for global food price models [17], interpreting the Global Food Security Index [18], and identifying household-level food insecurity drivers in West Kalimantan [19]. Furthermore, to address concerns regarding high-performance metrics, this research emphasizes rigorous validation through K-fold cross-validation and residual analysis to ensure the model's generalizability and transparency for evidence-based policy [20].

Environmental factors, particularly critical soil moisture thresholds that trigger plant water stress, are also essential for improving ini projections [21]. The application of the SHAP analytical framework uncovers underlying driving mechanisms and improves resilience measurement precision [22]. This comprehensive modeling provides a scientific foundation for evidence-based food crisis mitigation, supported by analyses of government performance and fiscal policy [23] and strategic analysis of food demand [24].

II. METHODOLOGY

This research was conducted through several main stages, namely data collection, data preprocessing, model development or analysis, and model performance evaluation. The entire process was carried out using the Google Colab platform as the computing environment.

A. *Research Framework*

The research framework was designed to provide a high-precision and transparently interpretable analysis of food vulnerability in Indonesia. The research process was divided into four integrated main phases, as illustrated in Figure 1. The first phase began with the acquisition and synthesis of secondary data, combining regional indicators from the Food Security Index (IKP) and the Food Security and Vulnerability Atlas (FSVA) for the 2022–2024 period, covering 514 districts/cities. This longitudinal approach was selected to capture the post-pandemic food security dynamics across the archipelago. Subsequently, in the second phase, advanced data cleaning was performed, including the use of Median Imputation to handle missing data and Standard Scaling to ensure feature scale uniformity. Furthermore, a multicollinearity test was conducted via a correlation heatmap to guarantee the integrity of the predictive variables before proceeding to the modeling stage.

In the third phase, model optimization and comparative analysis were conducted by evaluating seven different algorithms, ranging from bagging methods such as Random Forest and Extra Trees, to more complex ensemble-boosting algorithms such as Gradient Boosting, XGBoost, and LightGBM, as well as artificial neural networks via the Multi-Layer Perceptron (MLP). This step was taken to ensure the selection of the most robust model architecture for this multidimensional dataset. Finally, in the fourth phase, a rigorous validation strategy was applied using K-fold cross-validation and learning curve analysis to mitigate the risk of overfitting and guarantee the model's generalizability to new data. The best-performing model was then interpreted using the SHapley Additive exPlanations (SHAP) method to provide a profound understanding of feature contributions both globally and locally, particularly in priority regions such as Papua. The entire framework was implemented using the Python ecosystem within the Google Colab environment, resulting in a transparent and reproducible pipeline to support evidence-based policy interventions.

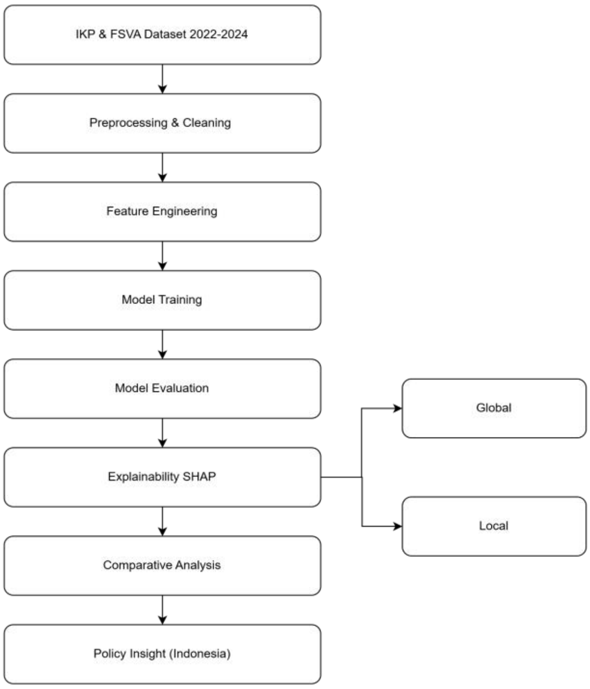

Figure 1. Research Framework Flowchart

B. Data Acquisition and Preparation

The primary dataset for this study was synthesized from the Food Security Index (IKP) and the Food Security and Vulnerability Atlas (FSVA) issued by the National Food Agency (BAPANAS). To ensure model transparency, nine independent variables representing three pillars of food security (availability, access, and utilization) were defined, as summarized in Table I.

TABLE I
OPERATIONAL DEFINITIONS OF FSVA INDICATORS

Food Security Pillar	Indicator	Brief Definition	Unit
Availability	NCPR	Ratio of total cereal consumption to total local cereal production.	Ratio
Access	Poverty	Percentage of the population whose expenditure is below the poverty line.	Percent (%)
	No Electricity	Percentage of households without access to electricity.	Percent (%)
	Food Expenditure	Ratio of food expenditure to total household expenditure.	Ratio
Utilization	No Clean Water	Percentage of households without access to safe drinking water sources.	Percent (%)
	Health Worker Ratio	Number of health workers per certain population density.	Ratio
	Stunting	Prevalence of children under five whose height-for-age is below the standard.	Percent (%)

Based on the definitions provided in Table 1, the Net Consumption to Production Ratio (NCPR) is identified as a critical predictor for the availability pillar. The NCPR formula used to measure regional cereal self-sufficiency is as follows:

$$NCPR = \frac{Total\ Cereal\ Consumption\ per\ Capita}{Total\ Local\ Cereal\ Production}$$

A higher NCPR value indicates that a region's consumption exceeds its local production, signifying a dependency on food inflows from other regions. Other key variables include the percentage of the population living below the poverty line, the percentage of households without access to clean water, the percentage of households without electricity, and the ratio of health practitioners per 1,000 population as a proxy for utilization capacity. All variables were verified for consistency across the multi-year observation period to ensure longitudinal reliability.

C. Data Preprocessing and Cleaning

Since the indicators have different units (percentages, ratios, and years), preprocessing is required to ensure comparability. The steps include cleaning data, handling missing values, data standardization, and feature selection, while the results of the high correlation analysis are presented in Table II.

TABLE II
RESULTS OF HIGH CORRELATION ANALYSIS

Features With High Correlation (> 0.8)	
Ikp	1.000000
Komposit	0.950851
NCPR	-0.805384
Ikp_ranking	-0.822056

The data preprocessing phase was conducted rigorously to ensure model integrity and prevent data leakage, which is critical given the high precision of the resulting models. Initially, a preliminary correlation analysis was performed using the abs() function to identify relationships between variables and detect potential redundancies. During this screening, it was identified that variables such as IKP, Komposit, and ikp_ranking exhibited near-perfect correlation with the target variable because they are derivative metrics. To eliminate information redundancy and prevent biased modeling, these three variables were removed from the dataset using the df.drop function.

Following feature selection, the dataset was partitioned into a training set (80%) and a testing set (20%) using train_test_split with a random_state=42 to ensure reproducibility. To handle missing values while mitigating the influence of regional outliers, median imputation was applied to maintain data continuity. Subsequently, feature scaling was performed using StandardScaler to achieve a mean of 0 and a standard deviation of 1. Crucially, the fit_transform process was applied exclusively to the training set, while the test set was only transformed using parameters derived from the training data. This specific procedure was implemented to strictly prevent data leakage and ensure that the model evaluation remains objective.

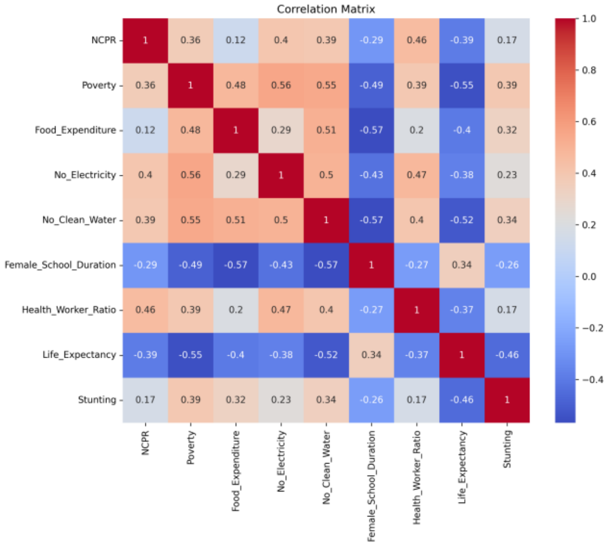

Figure 2. FSVA Indicator Correlation Heatmap

Based on the correlation matrix analysis shown in Figure 2, all FSVA indicators were retained in this study. The highest correlation coefficient value recorded was 0.57, indicating no multicollinearity between variables (below the critical threshold of 0.80). Therefore, all indicators are considered to provide unique and complementary information in comprehensively describing the food vulnerability profile of the region.

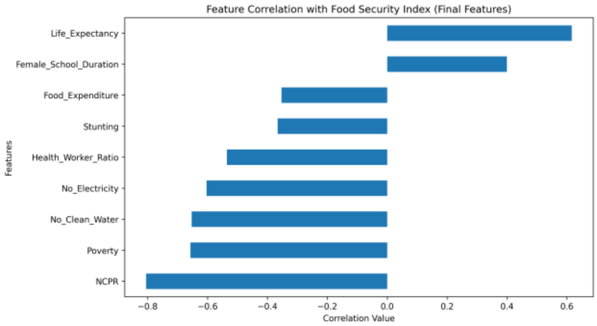

Figure 3. Final Correlation Analysis

Next, a final correlation analysis was conducted, as illustrated in Figure 3. The results of the final correlation analysis showed that after removing variables with high correlations (|r| > 0.80), the analysis results show that NCPR has the strongest negative correlation with the IKP, followed by poverty and limited access to clean water and electricity. Conversely, life expectancy and years of schooling for women are positively correlated with the IKP. Other variables show more moderate negative correlations. These results are more representative and show no indication of data leakage because the variables are independent.

D. Model Development and Comparative Analysis

This study benchmarks seven machine learning models to identify the most robust architecture for socio-economic data. These models encompass Bagging Ensembles, Boosting Ensembles, and Neural Networks. To ensure mathematical precision in evaluating these models, four primary metrics were employed:

1. Mean Absolute Error (MAE):
   Represents the average of the absolute errors between predicted and actual values.

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i|$$

2. Root Mean Square Error (RMSE):
   Measures the square root of the average of squared differences between prediction and actual observation.

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

3. Coefficient of Determination (R^2):
   Measures the proportion of variance in the dependent variable that is predictable from the independent variables.

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

4. Cross-Validation R-squared (CV_R^2):
   Measures the average R^2 score obtained across multiple folds in cross-validation, reflecting how well the model generalizes to unseen data.

$$CV\_R^2 = \frac{1}{k}\sum_{i=1}^{k} R_i^2$$

E. Hyperparameter Configuration

To ensure stability and reproducibility, several key hyperparameters were configured for each algorithm. The specific values used to balance predictive accuracy and computational efficiency are detailed in Table III.

TABLE III
MACHINE LEARNING MODEL HYPERPARAMETERS

Parameter	Value	Description
n_estimators	100	Number of trees to balance bias and variance.
max_depth	None	Nodes expanded until all leaves are pure.
random_state	42	Ensures consistency across executions.
learning_rate	0.1	Adjusting the learning rate (GB, XGBoost, LightGBM)
max_iter	500	Maximum training iterations (MLP)
hidden_layer_sizes	100,50	The structure of hidden layers in an MLP

To ensure optimal performance and model stability, several key hyperparameters for the Machine Learning regressor were configured as shown in Table 3. The number of estimators (n_estimators) was set to 100, which was determined through preliminary testing to balance predictive accuracy and computational efficiency. Additionally, a random_state of 42 was applied to ensure the reproducibility of the experimental results across different executions. These configurations are essential for capturing the complex non-linear relationships within the FSVA indicators while maintaining model robustness.

F. Validation and Anti-Overfitting Strategy

To ensure the reliability and generalizability of the predictive models, a rigorous validation framework was implemented. The primary validation strategy involved K-fold Cross-Validation, where the training dataset was partitioned into ten equal subsets. The model was trained on nine subsets and validated on the remaining one in ten iterative cycles, ensuring that every data point was utilized for both training and validation. This method provides a more stable estimate of model performance and significantly reduces the risk of overfitting compared to a single train-test split. Furthermore, Learning Curve analysis was conducted to monitor the convergence between training and validation scores. By observing the gap between these scores as the training set size increased, the study could objectively confirm that the model was not suffering from high variance (overfitting) or high bias (underfitting), thereby validating the robustness of the high-accuracy results achieved in this research.

G. Model Interpretation using SHAP (XAI)

This study employs SHapley Additive exPlanations (SHAP) as an Explainable AI (XAI) method to interpret the machine learning model. SHAP quantifies the contribution of each feature to the prediction by computing its marginal contribution across all possible feature combinations.

$$\phi i = \Sigma_{s \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} [f(S \cup \{i\} - f(S)]$$

Where ϕi represents the contribution of feature i, S is a subset of features, and $f(S \cup \{i\}) - f(S)$ denotes the marginal impact of including feature ii in the model. SHAP values indicate both the magnitude and direction of feature influence, where positive values increase and negative values decrease the predicted outcome. The results are visualized using SHAP summary (beeswarm) plots to present global feature importance and the direction of their effects across observations. This approach enables both global and local interpretation of the model.

III. RESULTS AND DISCUSSION

A. Algorithm Performance Comparison

Based on the evaluation of seven machine learning algorithms, tree-based ensemble models demonstrated significant advantages over conventional linear and kernel-based models in predicting the Food Security Index (IKP). The results of the comprehensive evaluation, summarized in Table IV, indicate that Gradient Boosting achieved the most superior performance with an R² of 0.97708, followed by Random Forest and XGBoost.

TABLE IV
COMPARISON OF MODEL PERFORMANCE

Model	**MAE**	**RSME**	**R²**	**CV_R²**
Gradient Boosting	1.562140	2.153476	0.977082	0.966736
Random Forest	1.641535	2.302913	0.973791	0.956143
LightGBM	1.497024	2.336569	0.973019	0.962087
XGBoost	1.526680	2.391826	0.971728	0.960759
ANN (MLP)	2.877353	4.310580	0.908174	0.909884
Linear Regression	3.314403	4.709530	0.890390	0.868171
SVR	3.224414	7.447270	0.725912	0.663381

The data in Table IV proves that boosting-based models successfully increased the variance explanation (R²) by approximately 9% compared to Linear Regression as a baseline. Furthermore, these models achieved up to a 14% reduction in prediction error (MAE and RMSE), confirming that the ensemble learning approach is more effective in capturing the complex, non-linear relationships inherent in Indonesia's food security indicators. The high consistency between the R² and CV_R² scores, particularly for Gradient Boosting, demonstrates the model's exceptional stability and robustness, effectively addressing concerns regarding potential overfitting in high-accuracy modeling.

B. Evaluation Model

Based on the evaluation results, the three best models were identified based on their R² values: Gradient Boosting, Random Forest, and XGBoost. All three models demonstrated excellent performance, with R² values above 0.95 and relatively low MAE and RMSE values. Among the three, Gradient Boosting performed best, achieving the highest R² value and the smallest error.

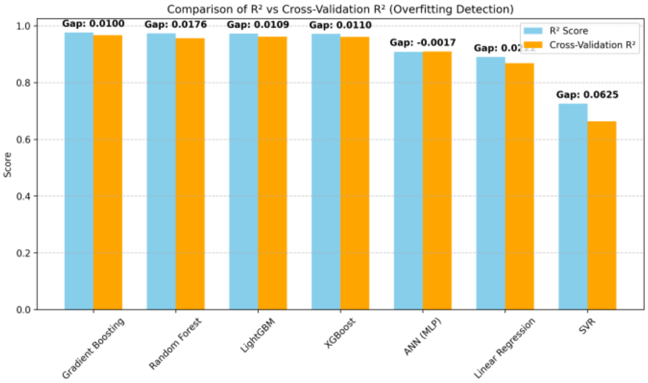

Figure 4. Results Overfitting Model

Additionally, an analysis was conducted to detect potential overfitting by comparing the R² values on the test data and the cross-validation results (CV_R²), as illustrated in Figure 5. The results showed that the gap between R² and CV_R² for all three models was relatively small (< 0.01), indicating that the models have good generalization and do not exhibit significant overfitting.

C. *Sensitivity and Residual Analysis*

To further ensure the robustness and reliability of the proposed model, sensitivity and residual analyses were conducted. These analyses aim to evaluate how variations in input variables affect model predictions and to examine the distribution of prediction errors, ensuring that the model does not suffer from bias or instability.

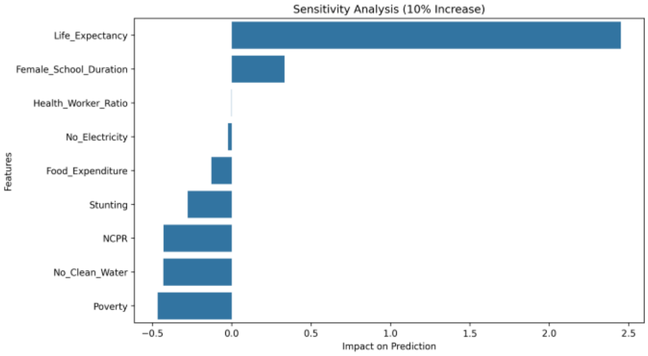

Figure 5. Sensitivity Analysis

Figure 5 presents the sensitivity analysis results by applying a ±10% change to each input variable. The findings indicate that life expectancy has the strongest positive influence on the predicted values, followed by years and female school duration. In contrast, NCPR, poverty, and lack of clean water show negative contributions, meaning that increases in these variables lead to higher food vulnerability. This result supports the multidimensional concept of food security, where both socio-economic and food availability factors significantly influence vulnerability.

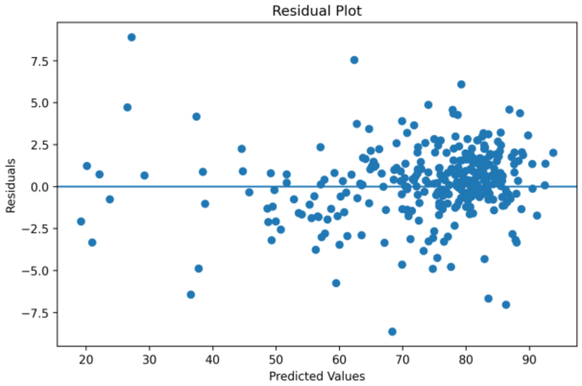

Figure 6. Residual Plot of Predicted Values vs Errors

Figure 6 illustrates the residual plot between predicted values and prediction errors. The residuals are randomly distributed around the zero line without forming a specific pattern, indicating that the model does not exhibit systematic bias and is able to capture the underlying relationships effectively. However, slightly larger deviations are observed in extreme values, suggesting that prediction accuracy may decrease in certain regions.

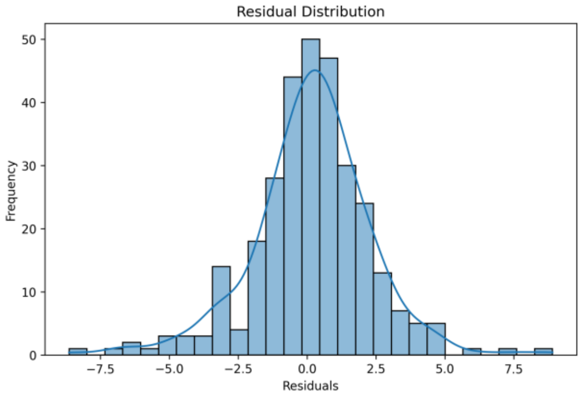

Figure 7. Distribution of Residuals

Figure 7 shows the distribution of residuals in the form of a histogram The results indicate an approximately normal distribution centered around zero, suggesting that the prediction errors are random and unbiased. Minor deviations at the tails indicate the presence of outliers, which is common in real-world socio-economic data. Overall, these results confirm that the model is stable and reliable for predicting food vulnerability.

D. *Global and Local SHAP Analysis for Food Vulnerability Prediction*

To bridge the complexity of the Gradient Boosting model with the need for transparency in policy-making, this study implements the SHapley Additive exPlanations (SHAP) framework. The analysis is conducted comprehensively, ranging from a global perspective to map national food vulnerability drivers to a local perspective serving as a region-specific diagnostic tool. The global influence of each FSVA indicator is visualized through a SHAP Summary (Beeswarm) Plot, which ranks each feature based on its average contribution to the IKP score predictions across all 514 districts and cities in Indonesia.

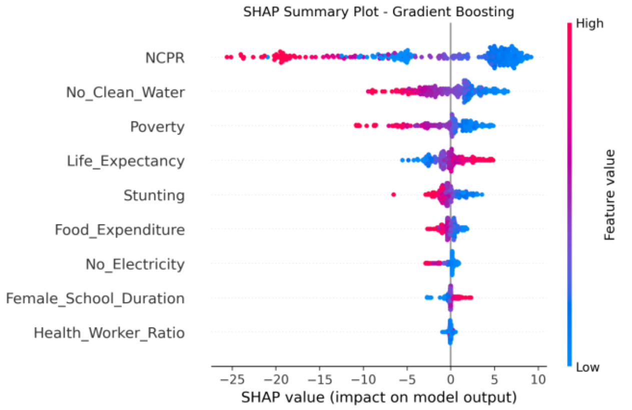

Figure 8. Global SHAP Feature Importance for Food Security Index Prediction

Based on the visualization in Figure 8, the Net Consumption to Production Ratio (NCPR) emerges as the most influential feature in determining a region's vulnerability. The distribution of red dots (indicating high values) on the right side of the SHAP axis suggests that a high dependency on food supplies from outside the region consistently increases the predicted food vulnerability. Beyond availability factors, economic and infrastructural dimensions such as poverty levels and limited access to clean water also show a significant positive impact on vulnerability scores. This proves that food insecurity in Indonesia is heavily influenced by public purchasing power and environmental health quality. Conversely, health indicators such as life expectancy and the ratio of health workers act as protective factors, where regions with better healthcare services tend to exhibit more stable food security.

In addition to national trends, the primary advantage of the SHAP method lies in its ability to perform individual audits through the SHAP Waterfall Plot. This analysis provides an overview of how the model accumulates prediction scores from the base value to the final output for a specific region. For instance, an analysis was conducted on a district in the Papua region, which historically has high vulnerability levels, to provide a more granular diagnostic overview

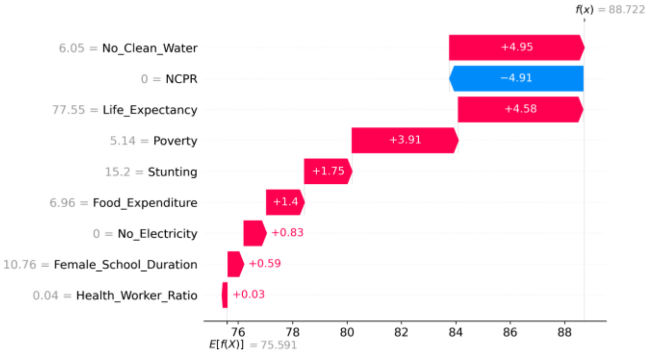

Figure 9. Local SHAP Waterfall Plot for a Selected District in Papua

The diagram in Figure 19 clearly illustrates the interaction between the drivers and inhibitors of vulnerability in that region. Extreme NCPR values and high stunting rates are the primary contributors pushing the score toward deeper vulnerability. This signals that food issues in Papua are not only caused by logistics and distribution bottlenecks but are also closely linked to chronic nutritional problems. Although several variables support resilience, their strength is insufficient to offset the pressure from poverty and supply dependency. This local diagnostic capacity is crucial for local governments, as it enables the design of highly specific interventions, such as focusing on food price stabilization and specific nutritional programs, rather than relying on generalized aid approaches.

E. Spatial Analysis (West vs East Disparities)

The results of predictions at the provincial level show a clear disparity in food vulnerability between western and eastern Indonesia, as illustrated in Figure 10. Regions in eastern Indonesia (Papua) are consistently predicted to have a higher risk profile than regions in the west, such as Java and Bali. These findings reinforce that differences in infrastructure quality, access to basic services, and inter-island logistics connectivity remain major challenges in efforts to mitigate national food insecurity evenly.

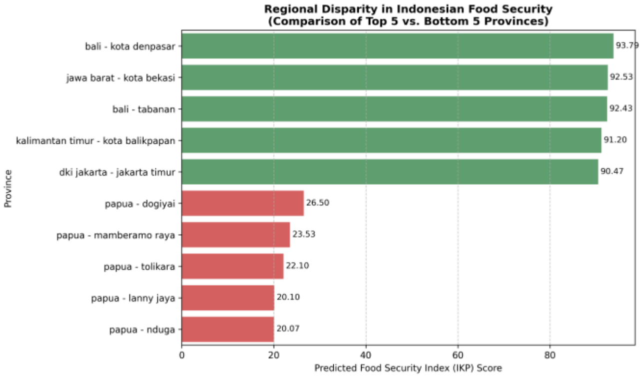

Figure 10. Food Security Index Comparison Between the Five Most and Least Resilient Provinces

Spatial analysis shows striking disparities between western and eastern Indonesia. Provinces in Java and Bali, such as Bali (predicted score of 93.79) and Central Java (90.47), are classified as highly food secure (Priority 6), while Papua (20.07) are classified as highly vulnerable (Priority 1). The extreme vulnerability in Papua is not only due to poverty rates, but is also greatly influenced by geographical conditions and inter-island logistical challenges. Papua's mountainous topography and severely limited basic infrastructure make food distribution costs very high. Dependence on air transportation to reach mountainous areas makes food supplies unstable and vulnerable to weather fluctuations. In addition, limited access to clean water and health facilities in rural Papua exacerbates the food utilization pillar, so that the risk of stunting and malnutrition remains high even though food aid may be available. These conditions underscore that mitigation strategies for Papua must include the development of connectivity infrastructure and the strengthening of local commodities such as sago and sweet potatoes, which are more adaptive to the local ecosystem than imposing a dependence on rice that must be imported from outside the island. These findings are consistent with the SHAP analysis, where variables such as NCPR, access to clean water, and poverty level show strong contributions to increasing food vulnerability in eastern regions. This confirms that the model predictions are not only statistically driven but also supported by real-world structural conditions.

F. Conclusion of Regional Disparity Analysis

To further explore the contrasting conditions between the most vulnerable and most resilient regions, a detailed comparison is presented using SHAP Force Plots for the Papua and Bali regions. This approach aims to provide a closer look at how various indicators interact within regions that possess vastly different characteristics. Through these visualizations, we can transparently identify the primary drivers and challenges unique to each area, demonstrating that the model provides deep contextual insights rather than just raw predictive figures.

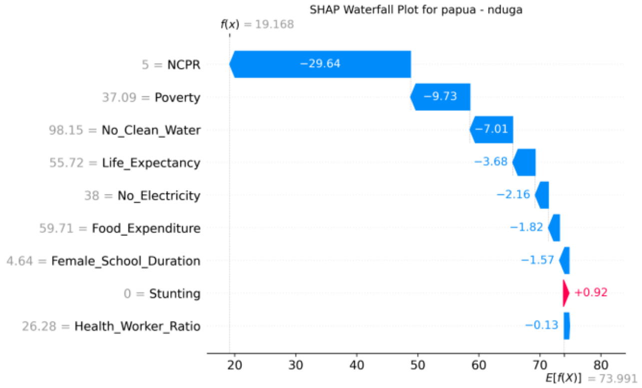

Figure 11. Vulnerability Factor Analysis in Papua (Most Vulnerable Region)

The visualization in Figure 11 illustrates the dynamics of the indicators for the Nduga district in Papua, which resulted in a very low predicted IKP score of 19.17. The score is drastically pulled down by several negative contributors. The analysis shows that the Net Consumption to Production Ratio (NCPR) is the most significant factor, reducing the score by -29.64, followed by high poverty levels (-9.73) and lack of clean water access (-7.01). This granular view explains that the extreme vulnerability in Nduga is not caused by a single issue, but rather by an accumulation of structural deficits. In stark contrast, the characteristics of a region with an optimal resilience profile are shown in the following visualization.

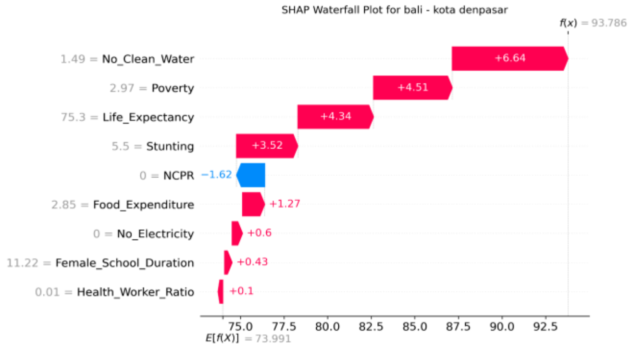

Figure 12. Vulnerability Factor Analysis in Bali (Most Resilient Region)

The contrast between Figure 11 and Figure 12 highlights the massive regional disparity, with Bali's Kota Denpasar achieving a remarkably high predicted IKP score of 93.79. As shown in Figure 13, the score is boosted significantly from the base value by positive contributors (red bars), such as clean water access (+6.54), poverty (+4.51), and high life expectancy (+4.34). These findings underscore that while eastern regions like Papua are struggling with basic supply and infrastructure, western regions like Bali have built a robust protective buffer through infrastructure and socio-economic development.

Ultimately, the significant gap between the scores of 20.07 and 93.79 confirms that national food security cannot be managed with a "one-size-fits-all" strategy. The analysis proves that the Gradient Boosting model, interpreted through SHAP, provides the necessary evidence for a differentiated policy approach. For highly vulnerable regions like Nduga, the priority must be on reducing supply dependency and building basic utilities, whereas for resilient regions like Denpasar, the focus should shift toward maintaining environmental and economic sustainability.

G. Model Generalization

Figure 13. Accuracy Stability of Gradient Boosting

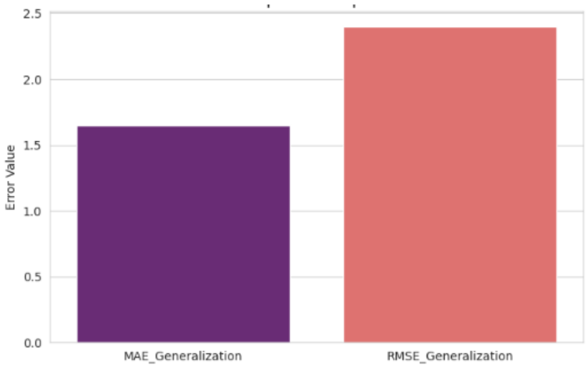

Figure 14. Error Metrics During the Generalization Phase

Based on the stability test results and error metrics shown in Figures 13 and 14, the Gradient Boosting model was selected as the primary model in this study. This decision was based on its ability to minimize the error rate, achieving the lowest RMSE value (2.157) compared to the other models. Furthermore, the generalization test results demonstrate highly consistent performance with a minimal accuracy difference (0.007) between the training and testing datasets. This proves that the model has high robustness and does not suffer from overfitting, meaning that the factors determining food vulnerability studied are universal and reliable for predicting conditions in regions or datasets outside the training data with a very high level of confidence.

H. Mitigation and Policy Recommendations

In high-risk regions such as Papua, interventions must prioritize the revitalization of local commodities. The transition back to sago (Metroxylon sagu) is supported by research demonstrating its ecological adaptability in peatlands, which is far superior to that of rice [25]. Sago is capable of growing in acidic soil without the need for land drainage, thereby preserving wetland ecosystems [25].

Regarding social safety nets, evaluations of the Family Hope Program (PKH) indicate that while cash transfers stabilize short-term consumption, the long-term impact on toddler nutrition is highly dependent on the presence of standardized health infrastructure [26]. Without adequate access to healthcare, cash assistance struggles to significantly reduce stunting rates [27].

In Bali, policy focus must be directed toward the protection of productive land. The conversion of Subak land into tourism areas threatens local food security [28]. A multi-stakeholder collaboration strategy through the "Subak Lestari" model is required to integrate sustainable tourism with the preservation of traditional agriculture [29].

IV. CONCLUSION

This study successfully fulfilled its objective of developing an accurate and interpretable predictive framework for assessing interregional food vulnerability risk in Indonesia. By evaluating 514 districts and cities using data from the 2022–2024 period, the research proved that the Gradient Boosting model outperformed other algorithms with an R^2 value of 0.9770, demonstrating its robustness in capturing complex non-linear relationships within food security indicators. The integration of SHapley Additive exPlanations (SHAP) provided critical transparency, identifying the Net Consumption to Production Ratio (NCPR), clean water access, and poverty levels as the primary drivers of vulnerability. Spatially, the model confirmed a significant score gap between western and eastern regions, highlighting the urgent need for infrastructure-based interventions in provinces like Papua. However, a notable limitation of this study is its reliance on static annual secondary data from the FSVA, which may not fully reflect sudden shocks caused by short-term climate variability or rapid food price fluctuations. To extend these findings, future research should incorporate Small Area Estimation (SAE) models to achieve village-level granularity, as planned for the FSVA 2025 update. Furthermore, integrating real-time climate monitoring and daily market prices into the framework could transform this model into a dynamic early warning system. Such advancements would strengthen the practical implementation of this interpretable framework within Indonesia's national and regional food security planning, ensuring that resource allocation remains precise, data-driven, and highly responsive to emerging local challenges.

REFERENCES

[1] N. Anwar, "Indonesia's Regional Food Security in Light Of The Impending Global Food Crisis," vol. 21, no. 2, pp. 101–110, 2022.

[2] F. Agus, E. Surmaini, and H. Lina, "Strategies for Indonesia's Agricultural Climate Change Adaptation and Mitigation," *Jurnal Sumberdaya Lahan*, vol. 16, no. 2, p. 67, Jan. 2023, doi: 10.21082/jsdl.v16n2.2022.67-81.

[3] L. Moh Arsal Fadila and N. Arsyta Putri, "Analisis Perkembangan Ketahanan Pangan di Indonesia : Pendekatan Menggunakan Big Data dan Data Mining," in *Seminar Nasional Official Statistics*, Jakarta: Badan Pusat Statistik, 2023.

[4] I. A. Juliannisa, H. Rahma, S. Mulatsih, and A. Fauzi, "Regional Vulnerability to Food Insecurity: The Case of Indonesia,"

Sustainability (Switzerland), vol. 17, no. 11, Jun. 2025, doi: 10.3390/su17114800.

[5] W. W. Prastanika and A. W. Wijayanto, "Analisis Hard dan Soft Clustering Untuk Pengelompokan Indikator Ketahanan Pangan Indonesia 2021," *Jurnal Sistem dan Teknologi Informasi (JustIN)*, vol. 11, no. 4, p. 596, Oct. 2023, doi: 10.26418/justin.v11i4.68400.

[6] B. Susantyo *et al.*, "Social Cash Assistance for Food Security During a Disaster: Lesson Learned from Indonesia," in *IOP Conference Series: Earth and Environmental Science*, Institute of Physics, 2023. doi: 10.1088/1755-1315/1180/1/012047.

[7] Anak Agung Ngurah Mayun Wirajaya, Made Sri Yuliartini, and Anak Agung Made Semariyani, "PKM for Improving Food Security at PKK Bongan Village Tabanan District, Tabanan District-Bali Province," *Asian Journal of Community Services*, vol. 2, no. 11, pp. 983–992, Nov. 2023, doi: 10.55927/ajcs.v2i11.7020.

[8] S. Saa, "Evaluation of the Policy Program for Strengthening Local Food Commodities for Sustainable Food Security in Papua," *JPPI (Jurnal Penelitian Pendidikan Indonesia)*, vol. 10, no. 2, p. 478, Jun. 2024, doi: 10.29210/020244167.

[9] R. Qasrawi *et al.*, "Machine Learning Approach for Predicting the Impact of Food Insecurity on Nutrient Consumption and Malnutrition in Children Aged 6 Months to 5 Years," *Children*, vol. 11, no. 7, Jul. 2024, doi: 10.3390/children11070810.

[10] A. H. Villacis, S. Badruddoza, A. K. Mishra, and J. Mayorga, "The Role of Recall Periods When Predicting Food Insecurity: 1 A Machine Learning Application in Nigeria 2 3," pp. 1–29.

[11] M. Al Fachrozi and K. D. Tania, "Comparison of Naïve Bayes, SVM, K-NN, Decision Tree, and Random Forest in Sentiment Analysis Based on SeaBank Application Aspects," *JURTEKSI (jurnal Teknologi dan Sistem Informasi)*, vol. 12, no. 1, pp. 69–80, Dec. 2025, doi: 10.33330/jurteksi.v12i1.4189.

[12] D. N. Fahria, K. D. Tania, and R. D. Kurnia, "Analisis Komparatif Algoritma Random Forest, XGBoost, dan CatBoost untuk Klasifikasi Tingkat Stres Pengguna Media Sosial," *Rabit : Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 11, no. 1, pp. 1843–1853, Jan. 2026, doi: 10.36341/rabit.v11i1.7449.

[13] Z. B. P. Pratama and Y. P. Astuti, "Perbandingan Metode Machine Learning (Linear Regression, Random Forest, dan XGBoost) dalam Memprediksi Kemiskinan di Jawa Tengah Tahun 2024," *Sistemasi: Jurnal Sistem Informasi*, vol. 14, no. 5, p. 2492, Jul. 2025, [Online]. Available: http://sistemasi.ftik.unisi.ac.id

[14] F. M. Saragih and W. C. Wibowo, "Analysis of Food Security Index Predictions in Indonesia Using Machine Learning Approach," *Agro Bali*, vol. 8, no. 2, pp. 377–393, 2025, doi: 10.37637/ab.v8i2.2302.

[15] D. N. Handayani and S. Qutub, "Penerapan Random Forest Untuk Prediksi Dan Analisis Kemiskinan," *RIGGS: Journal of Artificial Intelligence and Digital Business*, vol. 4, no. 2, pp. 405–412, May 2025, doi: 10.31004/riggs.v4i2.512.

[16] N. N. Sholihah and A. Hermawan, "Implementation of Random Forest and SMOTE Methods for Economic Status Classification in Cirebon City," *Jurnal Teknik Informatika (Jutif)*, vol. 4, no. 6, pp. 1387–1397, Dec. 2023, doi: 10.52436/1.jutif.2023.4.6.1135.

[17] X. Han, T. Yuan, D. Wang, Z. Zhao, and B. Gong, "How to Understand High Global Food Price? Using SHAP to Interpret Machine Learning Algorithm," *PLoS One*, vol. 18, no. 8 August, Aug. 2023, doi: 10.1371/journal.pone.0290120.

[18] Y. Wang, S. Schaub, D. Wuepper, and R. Finger, "Interpreting the Global Food Security Index: A Machine Learning Approach with SHAP Values," *Food Policy*, vol. 120, p. 102482, Oct. 2023, doi: 10.1016/j.foodpol.2023.102482.

[19] M. Usman and A. Hidayat, "Algoritma Shapley Additive Explanations untuk Analisis Kejadian Rawan Pangan Rumah Tangga di Kalimantan Barat," in *Prosiding Seminar Hasil Penelitian dan Pengabdian Kepada Masyarakat*, 2024.

[20] Rais, A. M. Soleh, and B. Susetyo, "Rotation Double Random Forest Algorithm to Predict The Food Insecurity Status of Households," *Jurnal RESTI*, vol. 8, no. 1, pp. 33–41, Feb. 2024, doi: 10.29207/resti.v8i1.5540.

[21] Z. Fu *et al.*, "Global Critical Soil Moisture Thresholds of Plant Water Stress," *Nat. Commun.*, vol. 15, no. 1, Dec. 2024, doi: 10.1038/s41467-024-49244-7.

[22] D. Zhou, J. Zhang, F. Zou, Z. Li, and H. Huan, "Optimization and Driving Mechanisms of Agricultural Resilience Measurement Based on XGBoost-SHAP Model: Evidence from China," *Front. Sustain. Food Syst.*, vol. 10, 2026, doi: 10.3389/fsufs.2026.1767684.

[23] S. R. Machieu, F. H. Y. Akib, and M. Mukhtar, "Analisis Kinerja Pemerintah atas Indeks Ketahanan Pangan Berdasarkan Kebijakan Fiskal dalam Ekonomi Pertanian pada Kabupaten/Kota di Kawasan Teluk Tomini," *Mimbar Agribisnis: Jurnal Pemikiran Masyarakat Ilmiah Berwawasan Agribisnis*, vol. 12, no. 1, pp. 901–916, Jan. 2026.

[24] M. Adith, P. Katti, and D. Gupta, "A Strategic Analysis of Food Demand Using Machine Learning and Explainable AI," in *2024 IEEE 9th International Conference for Convergence in Technology (I2CT)*, IEEE, Apr. 2024, pp. 1–7. doi: 10.1109/I2CT61223.2024.10543689.

[25] P. A. Dimara and A. Auri, "Effect of Landform on the Distribution of Metroxylon sagu Habitat in Yapen Islands, Papua Province, Indonesia," *Jurnal Sylva Lestari*, vol. 11, no. 1, pp. 79–97, Jan. 2023, doi: 10.23960/jsl.v11i1.633.

[26] N. Putri, A. A. Rachmayanti, and Y. P. Pramesti, "A Narrative Review: How to Optimize Program Keluarga Harapan in Addressing Nutritional Issues in Indonesia," *Salus Cultura: Jurnal Pembangunan Manusia dan Kebudayaan*, vol. 4, no. 1, Jun. 2024, doi: 10.55480/saluscultura.v4i1.203.

[27] N. Cahyadi, R. Hanna, B. A. Olken, R. A. Prima, E. Satriawan, and E. Syamsulhakim, "Cumulative Impacts of Conditional Cash Transfer Programs: Experimental Evidence from Indonesia," *Am. Econ. J. Econ. Policy*, vol. 12, no. 4, pp. 88–110, 2020, doi: 10.1257/pol.20190245.

[28] L. A. D. Ambarini, E. Frimawaty, and H. Herdiansyah, "Sustainable Management Strategies for Degraded Subak Rice Fields in South Denpasar, Bali," *Journal of Degraded and Mining Lands Management*, vol. 13, no. 1, pp. 9199–9209, Jan. 2026, doi: 10.15243/jdmlm.2026.131.9199.

[29] N. W. P. Rusadi, I. G. Pitana, I. N. Sunarta, and I. N. S. Arida, "Development of Subak Lestari as a Sustainable Tourism Attraction in Denpasar City," *International Journal of Green Tourism Research and Applications*, vol. 5, no. 2, pp. 114–128, Dec. 2023, doi: 10.31940/ijogtra.v5i2.114-128.