JDEP Vol. 8, Special Issue: Call for Paper Pusaka Jateng (2025) 19-40

JDEP

Jurnal Dinamika Ekonomi Pembangunan

https://ejournal.undip.ac.id/index.php/dinamika_pembangunan

Building an Intelligent System for Food Distribution Empowered by AI: Tackling Surplus–Deficit Inequality in the Barlingmascakeb Agglomeration

Harun Al Azies[1, 2, *], Ananta Surya Pratama[1], Taufiqul Umam[1], and Muhammad Diva Irnanda[1]

[1]Study Program in Informatics Engineering, Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang, Indonesia

[2]Research Center for Quantum Computing and Materials Informatics, Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang, Indonesia

[*]Corresponding Email: harun.alazies@dsn.dinus.ac.id

Received: 30th November 2025; Last Revised: 27th December 2025;
Accepted: 28th December 2025; Available Online: 17th January 2026

Abstract

Food distribution disparities remain a persistent challenge in the Barlingmascakeb region (Banyumas, Cilacap, Purbalingga, Banjarnegara, and Kebumen), where socio-economic and infrastructural factors drive regional inequalities. This study applies a machine learning–based classification approach to identify sub-districts categorised as food surplus or deficit. The dataset, initially imbalanced, was balanced using the Synthetic Minority Oversampling Technique (SMOTE), followed by training and evaluating four ensemble algorithms: AdaBoost, Gradient Boosting, XGBoost, and CatBoost. Among the tested models, AdaBoost demonstrated the best overall performance with an accuracy of 0.9565, precision of 1.00, recall of 0.8333, and F1-score of 0.9091. Gradient Boosting achieved a more balanced recall (0.8333) than XGBoost and CatBoost, although with lower precision. Based on the Gradient Boosting model, Feature importance analysis identified the Food Security Index as the most critical determinant of food status, followed by clean water access, morbidity rate, health workforce availability, and poverty levels. This study offers a novel contribution by providing a high-resolution, sub-district-level classification of food surplus and deficit conditions using interpretable ensemble machine learning models integrated with multidimensional socio-economic and health indicators. Practically, the model supports targeted and data-driven food distribution policies; theoretically, it reinforces the multifaceted nature of food security beyond production alone; and for future research, it opens opportunities to extend the framework to spatio-temporal and optimization-based food distribution models.

Keywords: *Food Distribution, AdaBoost, SMOTE, and Ensemble Algorithms*

JEL Classification: *Q18, C45, and R58* https://doi.org/10.14710/jdep.8.0.19-40

Introduction

Food is an essential basic need for the survival of every individual, and its efficient management directly impacts the welfare of society (Hariram et al., 2023). Although the agricultural sector in Indonesia, particularly the production of key commodities such as rice, corn, cassava, sweet potatoes, and sago, has seen significant development, food distribution issues remain a major challenge. The disparity between surplus-producing and deficit-consuming regions often leads to price instability and supply shortages in certain areas, impacting the sustainability of food supply and the welfare of farmers. Central Java faces significant food distribution imbalances, particularly in the Barlingmascakeb agglomeration area (comprising Banyumas, Cilacap, Purbalingga, Banjarnegara, and Kebumen). Some regions of Barlingmascakeb have a surplus of food commodity production, while others experience a supply shortage, resulting in inefficient distribution. This imbalance could exacerbate food price fluctuations, affecting farmers' welfare and consumer purchasing power.

The causes of this distribution imbalance are multifaceted, including but not limited to inadequate distribution infrastructure, suboptimal logistics management, and limited access to markets for farmers (Sharma & Sharma, 2025). The lack of technological applications that could improve distribution efficiency also contributes to this issue. Therefore, this study aims to conduct a comprehensive analysis to identify and classify regions based on key food commodities' surplus and deficit status. A deeper understanding of these distribution patterns is expected to provide new insights into optimizing regional food distribution(Granillo-Macías, 2021).

This study aims to classify regions in Barlingmascakeb based on the surplus and deficit status of key food commodities, including rice, corn, cassava, sweet potatoes, and sago. The analysis uses an artificial intelligence-based approach to map regions with excess supply and those requiring additional supply to create a more efficient food distribution system(Yousaf et al., 2023). Through this approach, the study is expected to generate policy recommendations that will improve the food distribution system, stabilize prices, and enhance farmers' welfare. By utilizing data on production, stocks, consumption, and socio-economic conditions, this study is expected to provide a more detailed overview of food distribution patterns in Barlingmascakeb. The results of this research are expected to serve as a foundation for more efficient food distribution management policies, ultimately supporting price stability and improving farmers' welfare in the region.

Despite the growing body of literature on food security and agricultural resilience, previous studies have predominantly focused on macro-level analysis, spatial modelling, or single-dimensional indicators such as production or poverty. Recent applications of machine learning in this field have demonstrated promising predictive capabilities; however, most remain limited to provincial or national aggregation and rarely address food surplus–deficit classification at the sub-district level using interpretable, policy-relevant indicators. Moreover, existing studies often emphasise model accuracy without sufficiently linking results to actionable food distribution strategies. To address these gaps, this study develops a high-resolution, sub-district–level classification framework for food surplus and deficit conditions in the Barlingmascakeb agglomeration using interpretable ensemble machine learning models integrated with multidimensional socio-economic and health indicators. By bridging advanced data-driven methods with practical policy needs, this research provides a localised decision-support tool to inform more targeted, equitable, and evidence-based food distribution policies.

Related Works

Table 1 summarises recent studies on food security and the analytical methods applied. While these studies provide valuable insights into availability, accessibility, and sustainability dimensions, most remain limited in their scalability, adaptability, or ability to predict food security status at granular administrative levels. Furthermore, few studies address the issue of surplus–deficit food inequality in agglomerated rural–urban regions, such as Barlingmascakeb.

While previous studies have widely applied statistical and machine learning approaches to food security, most have focused on spatial modelling or macro-level trends, often neglecting the classification of food surplus and deficit status at the sub-district level using structured socio-economic data. This study addresses that gap by developing a supervised machine learning model, Gradient Boosting, for predicting food distribution status across 115 sub-districts in the Barlingmascakeb region. The research offers a novel, scalable framework to support more targeted and equitable food distribution policies by focusing on interpretable, non-spatial indicators.

Materials and Methods

This section details the framework used in this study. The research is divided into four main stages: data collection, data preprocessing, modeling, and final model evaluation. Figure 1 below illustrates the steps involved in this research.

Data Collection

The data used in this study were obtained from the National Food Agency of Indonesia (Badan Pangan Nasional) and encompass 115 sub-districts (kecamatan) across the Barlingmascakeb agglomeration area, which includes the regions of Banyumas, Cilacap, Purbalingga, Banjarnegara, and Kebumen. The dataset contains eight independent variables (X1–X8) representing socioeconomic, health, and food security aspects at the sub-district level.

The variables include: percentage of population living in poverty (X1), percentage of household expenditure on food (X2), percentage of households without access to clean water (X3), average years of schooling for women (X4), ratio of health workers to population (X5), morbidity rate (X6), stunting prevalence among children under five (X7), and the Food Security Index (IKP) as a composite index reflecting the dimensions of food availability, access, and utilization (X8).

Table 1. Summary of Previous Studies and Analytical Approaches Related to Food Distribution and Agricultural Vulnerability

No	Author(s) (Year)	Location	Method	Key Findings	Notes
1	(Virtriana et al., 2022)	West Java, Indonesia	CA-Markov, Multivariable Linear Regression (MLR)	Land cover prediction accuracy 93%, MLR explains 50.6% of rice productivity variation, projected food energy deficit in 2030 of 14 million Mcal	Integration of environmental factors from remote sensing strengthens food security prediction; linear methods limited in capturing complexity
2	(Riptanti et al., 2022)	East Nusa Tenggara	Structural Equation Modeling (SEM)	Food security mediates policy input effects on sustainability	Has not integrated machine learning prediction models
3	(Yuliani et al., 2022)	West Java	Random Forest + SHAP + PFI	SHAP and PFI feature importance differ; SHAP more stable across populations	Single province study; no integration with policy-driven predictive models
4	(Nur Aziza et al., 2024)	Central Java	K-Nearest Neighbors (KNN)	Low RMSE, improved food security	Did not analyze causal factors
5	(Bondansari et al., 2023)	Banyumas	Dynamic Model & MICMAC	Rice deficit projection for 2026; recommendations: diversification and land protection	Focus only on availability aspect
6	(Pratama et al., 2023)	Kulon Progo	Google Earth Engine (GEE) & Availability Prediction	Significant paddy land conversion; potential rice self-sufficiency only until 2068	Does not consider intensification and policy interventions
7	(Azies, 2023)	Indonesia	Machine Learning (XGBoost Classification)	91% accuracy; safe drinking water (X1) most influential; XGBoost effective	No causal testing or SEM integration
8	(Nisa et al., 2023)	Kulon Progo	ML (NDVI + GEE + SPSS)	Paddy field conversion of 126 ha/year (2015–2020); food security still low in Pengasih	Does not consider land quality and actual planting patterns
9	(Dharmawan et al., 2022)	Indonesia	ML (RF, XGB, SVM, NN + SHAP + SMOTE)	SMOTE-N best; RF excels; key features: poor water quality, small houses, low education	Not integrated with SEM or spatial methods
10	(Fransiska et al., 2025)	West Java, Indonesia	Random Forest (RF) & Generalized Random Forest (GRF)	GRF outperformed RF in terms of specificity and balanced accuracy; key determinants include housing conditions, sanitation adequacy, education level, and financial access	Focused on household-level food insecurity using FIES; does not address food surplus–deficit classification or regional food distribution planning

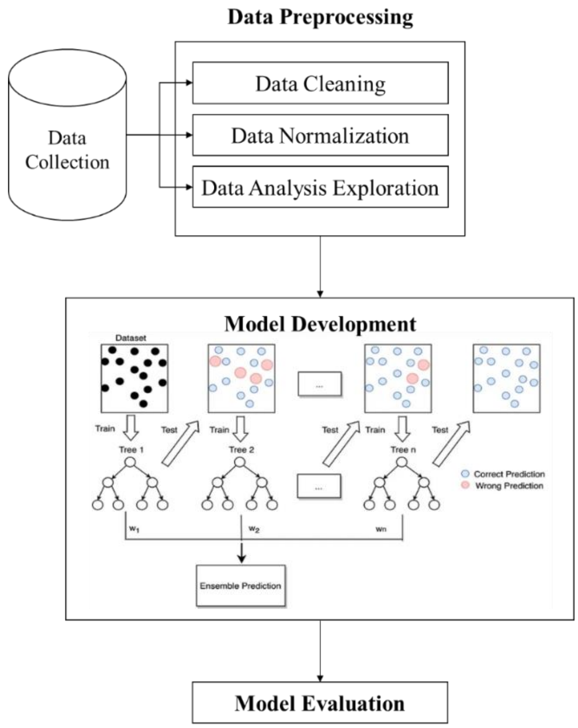


Figure 1. Research Framework

The target variable (Y) in this study is the food distribution status, categorized into two classes: *surplus* and *deficit*. This classification is based on the variable NCPR (Normative Consumption per Capita to Net Cereal Availability Ratio), in accordance with the National Food Agency guidelines. The classification is defined as follows. If NCPR ≥ 1, meaning normative consumption is equal to or greater than net cereal availability, the sub-district is classified as food-deficit (label: 1). If NCPR < 1, meaning net cereal availability exceeds normative consumption, the sub-district is classified as food-surplus (label: 0). Hence, the NCPR value, derived from official government data, serves as an objective basis for constructing the target labels in the food distribution classification model.

This data was collected for each of the 115 districts (kecamatan) within the Barlingmascakeb region, which includes the districts of Banyumas, Cilacap, Purbalingga, Banjarnegara, and Kebumen. These districts exhibit diverse agricultural, infrastructural, and socioeconomic conditions, making them ideal for studying food surplus and deficit dynamics. The research area comprising these districts is depicted in Figure 2. These data points provide a comprehensive snapshot of the local food systems and socioeconomic conditions across the Barlingmascakeb region, forming the foundation for analyzing food distribution patterns in the study.

Data Preprocessing

The data collected from BPS and Badan Pangan Nasional were subjected to preprocessing steps to ensure quality and consistency (Ghosh, 2022). Missing data were handled through imputation techniques, where feasible, or by removing incomplete records when necessary. Numerical features were normalized to standardize the scale of the data and avoid bias in machine learning algorithms(Li et al., 2021). Categorical variables were encoded into numerical representations to ensure compatibility with machine learning models.

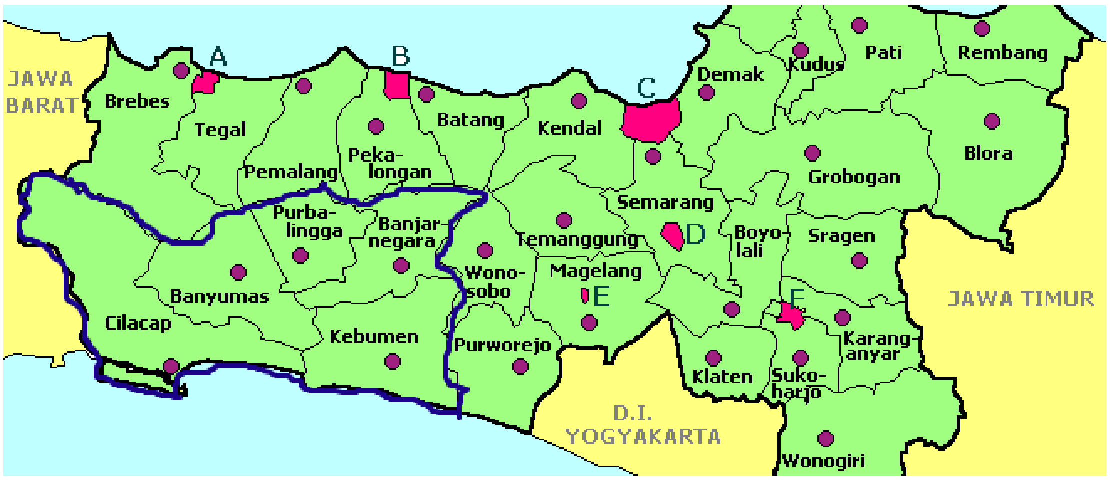


Figure 2. Map of the Barlingmascakeb Region, Highlighting the 115 Districts Included in the Study

Modeling

The modeling stage focuses on classifying the 115 districts of the Barlingmascakeb region based on their food surplus or deficit status. Ensemble Learning methods were employed due to their proven ability to enhance predictive performance by combining multiple models (Tuysuzoglu & Birant, 2020). Specifically, the study utilized Gradient Boosting, a powerful technique within the boosting family of ensemble methods, to classify the regions effectively.

Ensemble Learning and Gradient Boosting

Ensemble learning improves model performance by integrating predictions from several base learners(Ganaie et al., 2022). As illustrated in Figure 3, the two primary ensemble strategies are bagging and boosting. While bagging reduces variance by training multiple models on bootstrapped subsets of the data, boosting focuses on reducing bias and variance by sequentially training models to correct the mistakes of the previous ones.

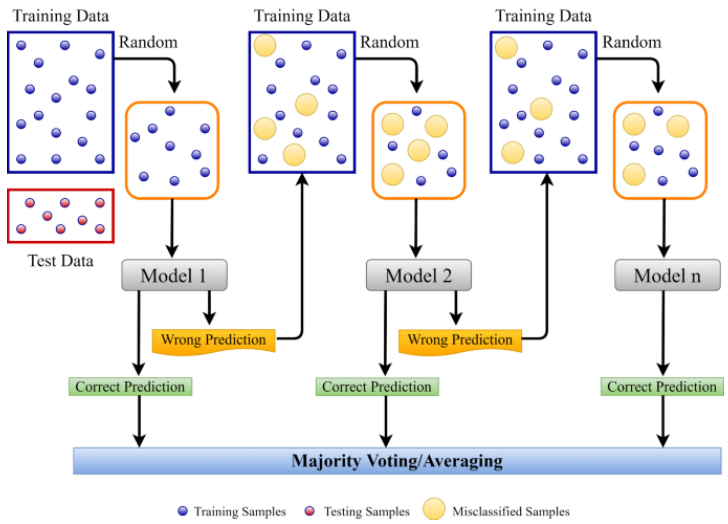


Figure 3. General Mechanism of Ensemble Learning
Source: Kavzoglu & Teke (2022)

Gradient Boosting was selected for this study as it builds a series of decision trees, where each tree is trained to correct the residual errors of the previous tree using gradient descent (Nie et al., 2021). This iterative approach allows the model to capture complex patterns and non-linear relationships in the data, making it particularly effective for classification tasks like this one. The final prediction is derived from the weighted sum of all the individual models, and the algorithm's performance improves with each additional tree. This study used Gradient Boosting (depicted in Figure 4) as the core algorithm..

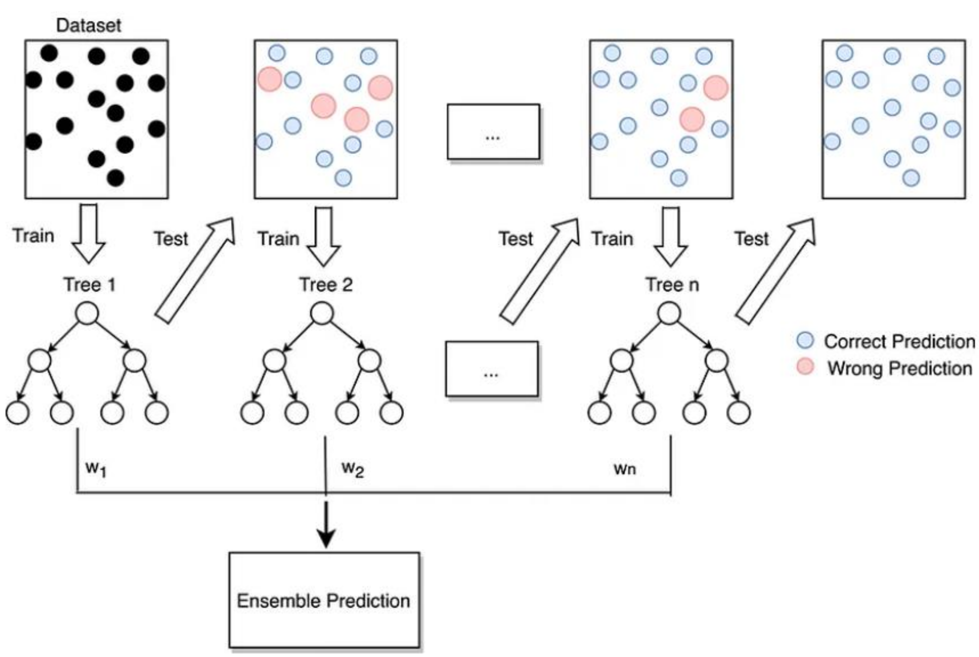


Figure 4. Working Principle of Gradient Boosting Algorithm
Source: Zhang et al. (2021)

Model Training and Hyperparameter Tuning

In this stage, the Gradient Boosting model was trained on the preprocessed dataset to classify each sub-district as either surplus or deficit, based on indicators such as agricultural production, infrastructure availability, and socioeconomic conditions. To optimize its predictive performance, hyperparameter tuning was carried out by adjusting several key parameters. These included the number of trees, which determines how many decision trees are combined in the ensemble; the learning rate, which controls how quickly the model updates its predictions; the maximum depth of the trees, which affects how complex the decision paths can be; and the subsample ratio, which introduces randomness by using only a portion of the data for each tree to help reduce overfitting. Both grid search and random search techniques were employed to identify the most optimal combination of these parameters, ensuring the model achieved reliable and accurate classification performance.

Model Evaluation

Cross-validation was employed to evaluate the model's performance, where the dataset was split into multiple subsets (folds). The model was trained on one fold and tested on the remaining data, ensuring the performance metrics were robust and generalized. Evaluation metrics included accuracy, precision, recall, and F1-score to measure the model's ability to classify regions as surplus or deficit correctly.

Model Interpretation

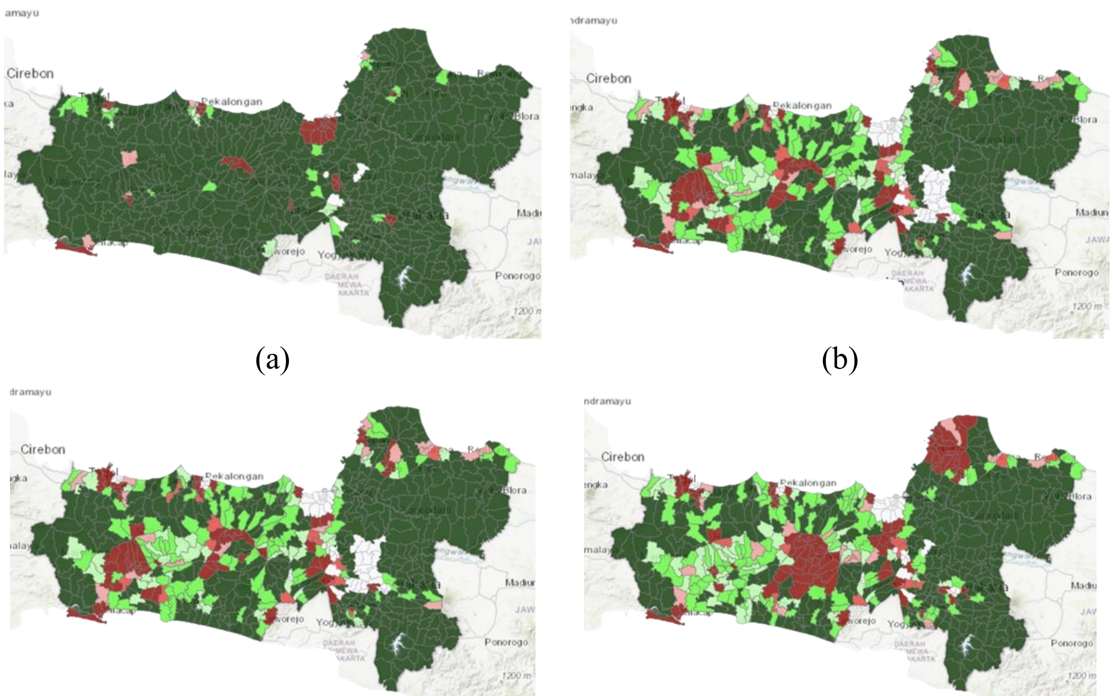
An advantage of Gradient Boosting is its ability to provide insights into the importance of various features in the classification process. Feature importance metrics were analyzed to understand which factors contributed most significantly to determining whether a district experienced a surplus or deficit in food. Variables such as net agricultural production, poverty rates, and access to irrigation emerged as key drivers in the classification process, providing valuable insights for policymakers to target regions needing interventions.

Final Model Evaluation

The final model was evaluated on the entire dataset after training the Gradient Boosting model and optimizing its hyperparameters. The results were analyzed to determine how well the model classified districts into surplus and deficit categories. Based on this evaluation, the model generated recommendations for policy interventions to optimize food distribution across the Barlingmascakeb region.

Results and Discussion

The Barlingmascakeb region covering Banyumas, Purbalingga, Banjarnegara, Cilacap, and Kebumen experienced notable shifts in food surplus and deficit distribution across its sub-districts during the 2018–2021 period. These shifts illustrate the changing landscape of food availability and distribution, influenced by environmental factors, agricultural practices, and infrastructure conditions.



(a) (b)

(c) (d)

Figure 5. Food Surplus and Deficit Status by Sub-district in Barlingmascakeb Region (a. 2018, b. 2019, c. 2020, d. 2021)

In 2018 (Figure 5.a), the majority of sub-districts showed a surplus in food production. These surpluses were predominantly found in the southern and western areas of Cilacap, central Banyumas, and parts of Kebumen. This pattern was supported by favorable seasonal conditions and relatively stable agricultural outputs in key commodities such as rice and corn. In 2019 (Figure 5.b), there was a noticeable shift, with more areas transitioning into a food deficit status. Deficit zones emerged in northern Banjarnegara, eastern Cilacap, and parts of Purbalingga. Factors contributing to this trend likely included disruptions in rainfall, reduced access to agricultural inputs, and growing pressure on productive land.

By 2020 (Figure 5.c), the pattern became more polarized. The number of sub-districts experiencing food deficits increased sharply, while surplus areas became more isolated. The impact of the COVID-19 pandemic further complicated the situation, particularly by limiting labor mobility, restricting market access, and delaying the distribution of agricultural support programs. In 2021 (Figure 5.d), the trend of food deficits continued in more than half of the region’s sub-districts. Surplus areas remained primarily in parts of Cilacap and central Banyumas. Meanwhile, districts like Banjarnegara and Purbalingga exhibited persistent deficits, suggesting challenges in maintaining agricultural production levels or improving distribution efficiency. These findings indicate that the regional food balance has undergone significant spatial and temporal changes, highlighting the importance of localized data analysis to inform more responsive and equitable interventions.

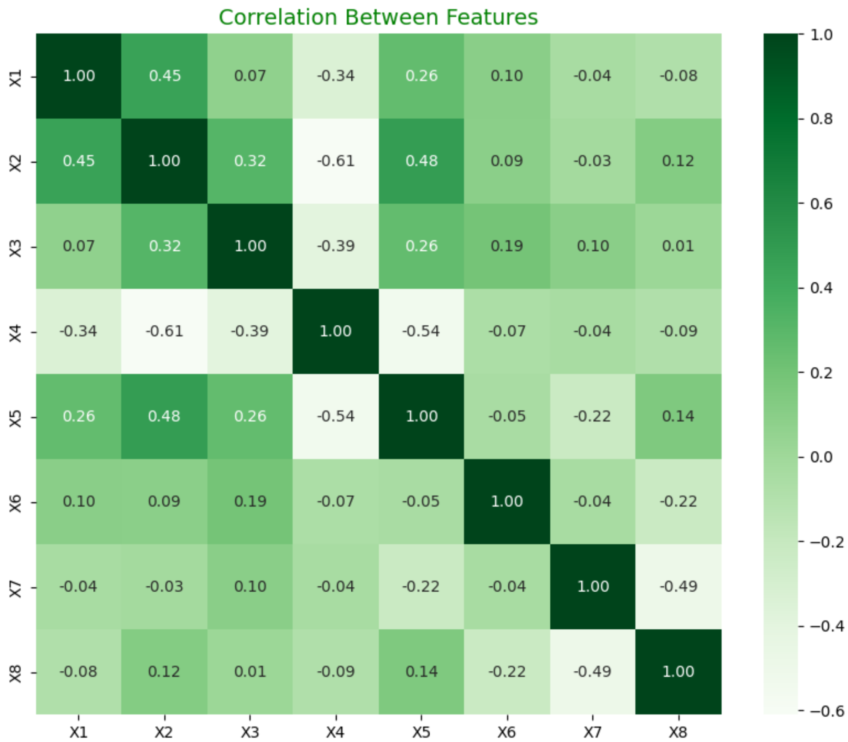


Figure 6. Pearson Correlation Heatmap of Socioeconomic Features in the Barlingmascakeb Region

Initial analysis explored the interrelationships between features to understand the relationship structure underlying food security conditions in the Barlingmascakeb region as illustrated in Figure 6. Through Pearson correlation analysis, it was found that several socio-economic variables exhibited a relatively strong linear relationship with each other. For example, the Poverty variable (X1) was positively correlated with No Clean Water (X3) and Stunting (X7), indicating that areas with high poverty rates tend to experience limited basic infrastructure and poor health conditions. Conversely, the Food Expenditure variable (X2) showed a negative correlation with Female Years of Schooling (X4) and the Health Worker Ratio (X5), suggesting that areas with high food expenditure burdens generally have lower access to education and health services. These findings reinforce the understanding that food security does not exist sectorally, but is influenced by interrelated underlying social conditions.

Because the target variable is a binary category (food surplus or deficit), a biserial correlation was also used to measure the strength of the relationship between each numeric feature and the classification label. The results, as shown in Table 2, indicate that variable *X8 (Food Security Index)* has a robust negative correlation with class status (coefficient =

-0.635, $p < 0.001$), suggesting that areas with a lower food security index are more likely to fall into the deficit category. This supports the validity of the classification labels used. The variable Without Clean Water (X3) also shows a significant negative correlation (coefficient = -0.392, $p < 0.001$), confirming the importance of sanitation as a key differentiator in food security status.

Table 2. Biserial Correlation between Socioeconomic Variables and Food Security Status

Variable	Correlation Coefficient	Significance of Correlation	Variable	Correlation Coefficient	Significance of Correlation
X1	-0.002	0.9852	X5	-0.199	0.0335
X2	-0.222	0.0169	X6	-0.103	0.2723
X3	-0.392	0.0000	X7	0.222	0.0171
X4	0.195	0.0364	X8	-0.635	0.0000

Other features, such as Food Expenditure (X2), Health Worker Ratio (X5), and Years of Schooling for Women (X4), show significant but moderate correlations ($p < 0.05$), indicating that social factors such as health and education also play a role in shaping food status. On the other hand, the variable Poverty (X1) shows a nearly zero correlation (coefficient = -0.002, $p = 0.985$), reinforcing previous findings that poverty is not necessarily a primary indicator in determining food surpluses or deficits. In contrast, access to clean water, education, and healthcare demonstrated greater significance in differentiating conditions between regions. This reinforces the premise that food security results from complex, multidimensional interactions.

Pairplot visualisations were also used to illustrate the distribution between classes for each feature, as shown in Figure 7. Several variables, such as Stunting (X7) and Poverty (X1), showed distribution patterns that tended to separate the surplus and deficit classes, indicating their potential classification power. However, most other features exhibited high overlap between classes, necessitating machine learning methods to capture non-linear interactions that univariate analysis could not capture.

Furthermore, boxplot visualisations were used to evaluate the distribution of values for each feature by food status class, as presented in Figure 8. In general, regions with deficit status had higher medians for indicators such as Poverty (X1), No Clean Water (X3), and Stunting (X7), confirming less supportive social conditions. Conversely, surplus areas generally show higher scores on the Female Years of Schooling (X4) and Health Worker Ratio (X5) indicators, indicating education's positive influence and health services' availability on food security. The geographic context of Barlingmascakeb further strengthens this explanation. The region encompasses diverse landscapes—from highlands to coastal areas and from urban to rural areas. Inequality in access to basic services is a key factor explaining differences in food status, even in regions with relatively similar macroeconomic conditions. Therefore, the results of this exploration provide an important foundation for building an artificial intelligence-based predictive model capable of comprehensively capturing the complexity of these interactions, which will be discussed in the next section.

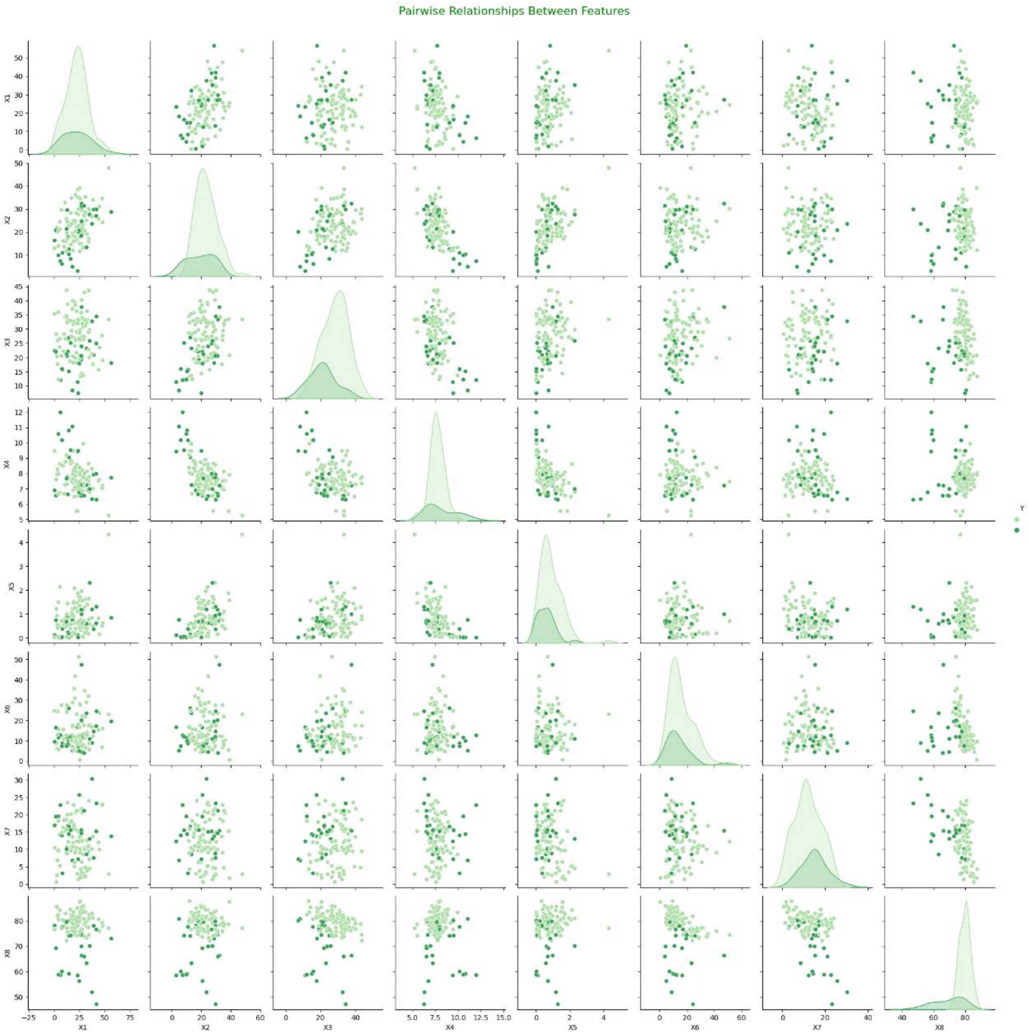


Figure 7. Pairplot Visualisations of Feature Distributions by Class

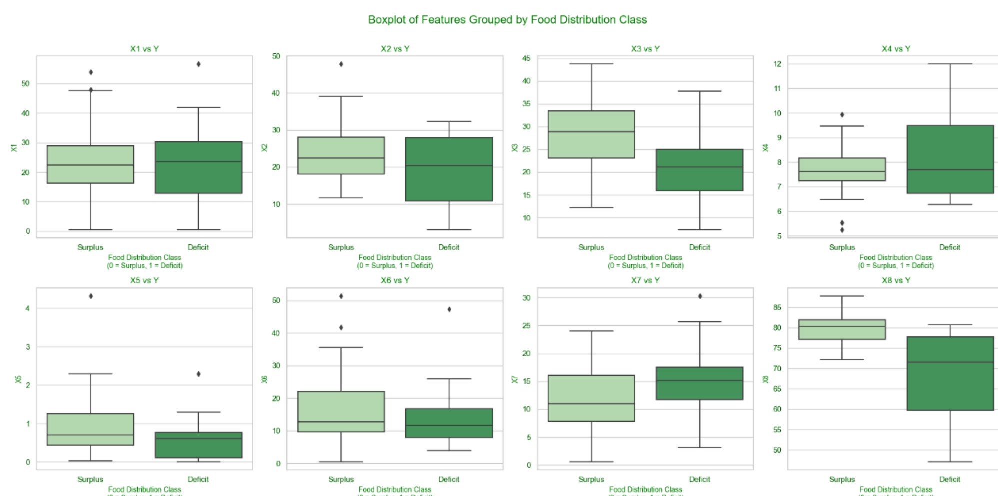


Figure 8. Boxplot Visualisations of Feature Distributions by Status

While developing a food status classification model for the Barlingmascakeb region, an imbalance in the class distribution between surplus and deficit areas was discovered. This phenomenon (Figure 9) is visible in the initial distribution graph, showing a predominance of surplus class samples. This imbalanced data can impact model performance, particularly in detecting minority (deficit) classes, as the model tends to be biased toward the majority class.

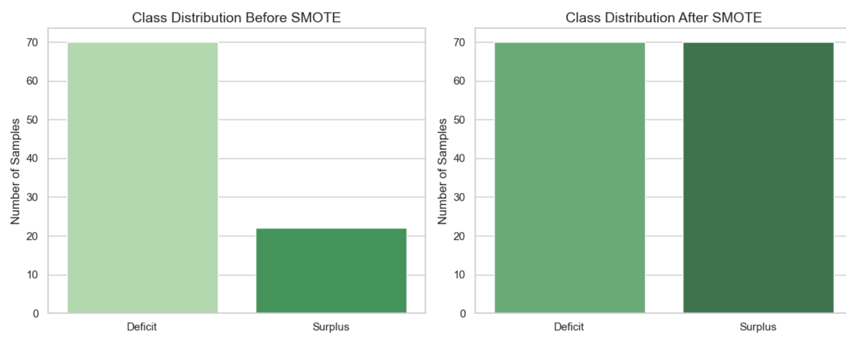


Figure 9. Class Distribution Before and After SMOTE Oversampling

To address this issue, the Synthetic Minority Oversampling Technique (SMOTE) was used. This technique synthetically replicates samples from the minority class to balance the class distribution more. The visualisation after applying SMOTE shows a more proportional class distribution, allowing the model to learn a fairer representation of both classes.

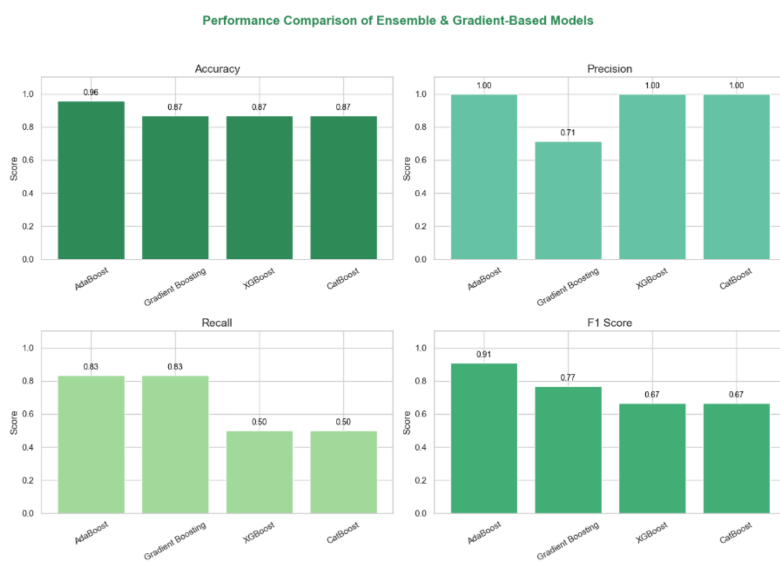
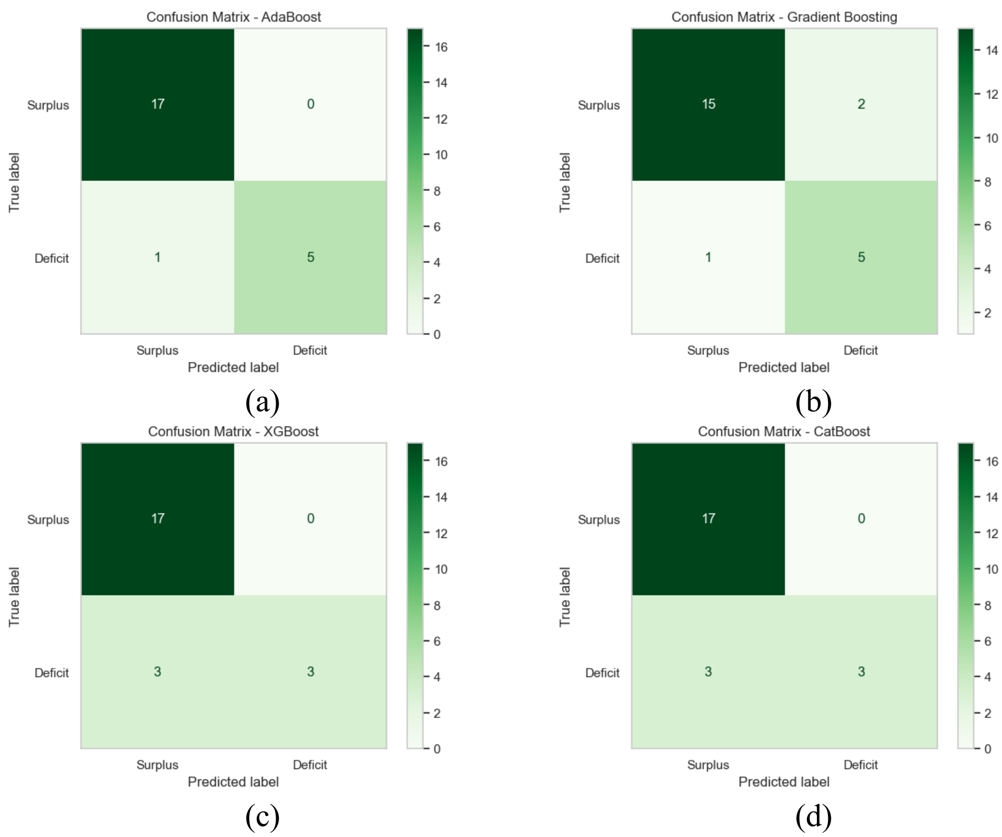


Figure 10. Comparative Model Performance Based on Evaluation Metrics

After the data balancing stage, training and evaluation were conducted on four ensemble and gradient-based algorithms: AdaBoost, Gradient Boosting, XGBoost, and

CatBoost. The comparative performance of these models is illustrated in Figure 10. The performance evaluation results showed that AdaBoost achieved the highest accuracy (0.9565), a perfect precision value (1.00), a recall of 0.8333, and the highest F1-score of 0.9091. This indicates that the model is accurate overall and balanced in recognising both classes. Meanwhile, Gradient Boosting provides balanced performance with a recall of 0.8333 and an F1-score of 0.7692, although its precision is lower (0.7143). The XGBoost and CatBoost models exhibit high precision (1.00) but lower recall (0.5), indicating a tendency to classify surplus cases while partially ignoring deficit cases correctly.



(a) (b) (c) (d)

Figure 11. Confusion Matrix Visualization of Four Ensemble Models for Food Status Classification: (a) AdaBoost, (b) Gradient Boosting, (c) XGBoost, and (d) CatBoost

Figure 11 presents the confusion matrices of the four tested models, providing a clear visual representation of their classification performance on balanced data. In Figure 1a, the AdaBoost confusion matrix demonstrates a relatively balanced classification ability, successfully identifying a substantial number of deficit cases while maintaining overall accuracy. In contrast, Figure 1c and Figure 1d, representing XGBoost and CatBoost respectively, show a noticeably lower number of correctly predicted deficit instances, supporting the previously reported low recall values and indicating a bias toward the surplus class. Figure 1b, corresponding to Gradient Boosting, displays a more balanced result than XGBoost and CatBoost, although it still falls short compared to AdaBoost in both precision and overall F1-score.

These visual findings reinforce the quantitative results and highlight the crucial role of data balancing using SMOTE. The ensemble learning approach, particularly with AdaBoost, significantly enhanced model robustness in identifying both surplus and deficit areas. Therefore, among all evaluated models, AdaBoost stands out as the most effective algorithm in handling balanced food status classification in the Barlingmascakeb region.

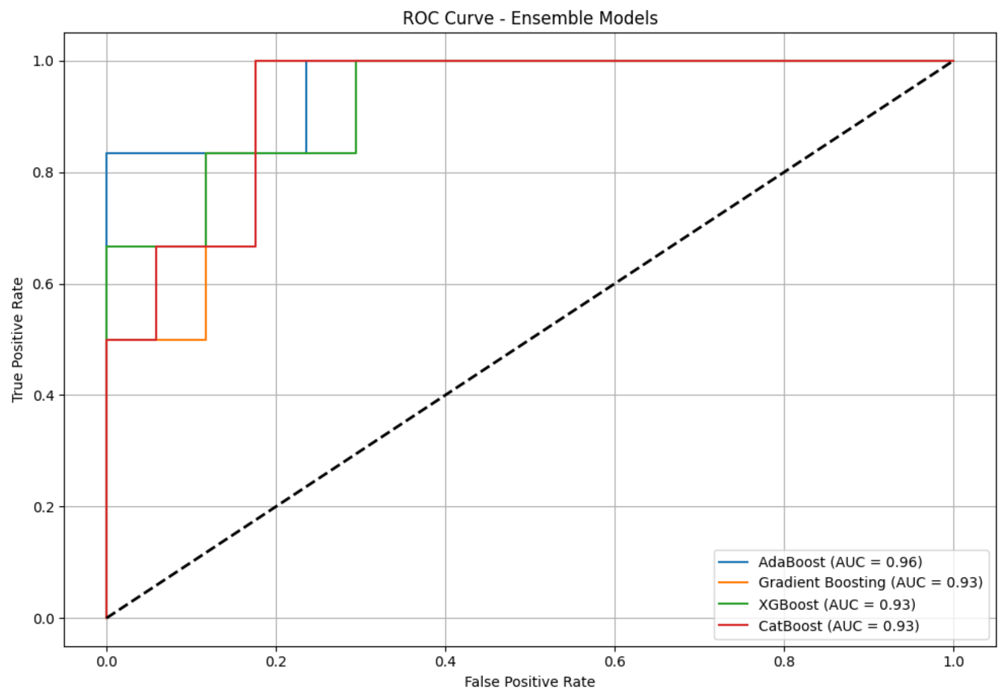


Figure 12. ROC Curve Comparison of Ensemble Models in Classifying Food Distribution Status in the Barlingmascakeb Region

The Receiver Operating Characteristic (ROC) in Figure 12 curve visually represents each model's ability to distinguish between food surplus and food deficit sub-districts. Among all models, AdaBoost's ROC curve is closest to the upper left corner, with the highest AUC value, indicating its best ability to minimize classification errors. This means AdaBoost can identify food deficit sub-districts more accurately, which is crucial considering these areas are prioritized for addressing food inequality.

Meanwhile, XGBoost and CatBoost's ROC curves are lower, indicating limited sensitivity to the food deficit class. While still competitive, Gradient Boosting performs below AdaBoost, but still better than XGBoost and CatBoost. This finding reinforces previous evaluation results, indicating that the model with the highest AUC is statistically superior and more responsive to policy needs in food-insecure areas.

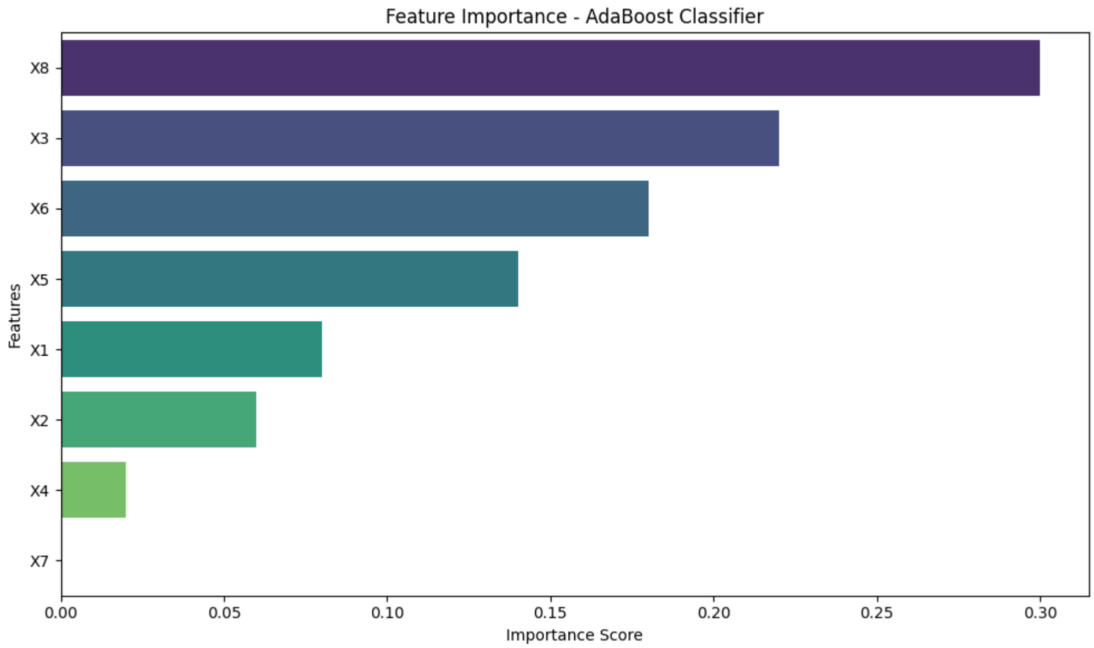


Figure 13. Feature Importance of AdaBoost Model for Food Distribution Status Classification in Barlingmascakeb Region

Based on the results of the feature importance analysis of the AdaBoost model (Figure 13), the most influential variable in determining the food distribution status of a sub-district is the Food Security Index (IKP). This indicates that, in general, a region's food security condition is closely related to the likelihood of that region experiencing a food surplus or deficit. As a composite index encompassing dimensions of food availability, access, and utilization, the IKP effectively represents a region's position within the local food system. After the IKP, the proportion of households without access to clean water emerged as the next important factor. This finding confirms that food distribution planning cannot ignore sanitation and fundamental infrastructure issues. Access to clean water is correlated with a household's ability to manage food safely and healthily. It is also related to health aspects that ultimately affect food stability at the micro-level. This support is reinforced by the importance of the morbidity rate variable, which contributes significantly to the classification of food distribution status. High morbidity rates can reflect unfavorable socioeconomic and environmental conditions, hindering regional food production and distribution.

On the other hand, the health worker ratio variable is also included in the group of variables with a significant contribution. This indicates that the distribution and availability of health services play a role in maintaining stable food distribution, both directly through public health interventions and indirectly by strengthening the community's adaptive capacity to food disruptions.

Poverty and food expenditure also influence food distribution status, although not as strongly as the health and food security dimensions themselves. This suggests that purchasing power is not the sole factor determining the balance of food distribution between regions. In the context of Barlingmascakeb, limited purchasing power tends to be structural and overlaps with other factors such as infrastructure and the quality of public services. Meanwhile, the last two variables, namely, years of schooling for girls and stunting rates, contribute less than the other features. Nevertheless, their role remains essential in the long term. Years of education for girls are related to nutritional awareness, child health, and household consumption patterns. At the same time, stunting reflects a chronic condition that can indicate future food insecurity.

The results of this analysis underscore that food distribution inequality is not simply a matter of production or consumption, but is heavily influenced by multidimensional structural factors, particularly in health, sanitation, and local food security. Therefore, an effective food distribution intervention strategy in the Barlingmascakeb area needs to be designed integratively, prioritizing a cross-sectoral approach that targets increased production and improved living conditions for the community as a whole.

The findings of this study are consistent with and extend previous research on food security and machine learning–based classification in Indonesia. Earlier studies have emphasised that structural and socio-economic factors beyond agricultural production alone influence food insecurity and food vulnerability. For example, Dharmawan et al. (2022) and Yuliani et al. (2022) identified sanitation quality, housing conditions, education, and access to basic services as key determinants in food insecurity classification using ensemble learning models. Similarly, the recent study by Fransiska et al. (2025) demonstrated that sanitation adequacy, housing characteristics, education level, and financial access were the dominant predictors of household food insecurity in West Java, using the Generalised Random Forest approach.

The present study corroborates these findings by showing that access to clean water, health conditions, and socio-economic resilience significantly influence food status. However, unlike previous studies that focus primarily on household-level food insecurity or macro-level vulnerability, this research advances the literature by shifting the analytical focus to sub–district–level classification of food surplus and deficit as a regional food distribution problem. In addition, while Fransiska et al. (2025) emphasise predictive stability and balanced accuracy at the household level, this study highlights the Food Security Index (IKP) as the most influential composite indicator in explaining surplus–deficit conditions across regions. This suggests that composite indices play a more prominent role when food security is examined from a territorial distribution perspective rather than an individual household experience.

These results indicate that although the underlying determinants of food insecurity and food distribution share common structural roots, differences in the scale of analysis and the definition of outcomes substantially affect which variables emerge as dominant. By integrating multidimensional indicators within an interpretable ensemble framework, this study complements and extends previous findings, offering new insights for regional food distribution planning and evidence-based policy intervention.

Conclusions and Recomendation

Conclusion

This study presents a data-driven and interpretable approach to analysing food distribution disparities across sub-districts in the Barlingmascakeb region (Banyumas, Cilacap, Purbalingga, Banjarnegara, and Kebumen). By applying data balancing through SMOTE and evaluating ensemble-based classification models, this research contributes to understanding the spatial distribution of food security conditions, particularly in identifying sub-districts classified as "deficit" or "surplus."

The implementation of SMOTE successfully addressed the class imbalance problem, which previously caused bias in model predictions toward surplus classifications. After balancing, four machine learning models, AdaBoost, Gradient Boosting, XGBoost, and CatBoost, were evaluated. Among these, AdaBoost emerged as the best-performing model, achieving the highest F1 Score, precision, and recall, particularly in accurately predicting deficit areas.

Feature importance analysis revealed that the Food Security Index (IKP) was the most influential factor in determining food distribution status, followed by access to clean water, health conditions, and healthcare infrastructure. Other variables, such as poverty levels, food expenditure, and female education, also played important but secondary roles. These findings suggest that food distribution disparities in Barlingmascakeb are closely linked to food availability and multidimensional aspects of public health, infrastructure, and socio-economic resilience.

For future research, the proposed framework can be extended by incorporating spatio-temporal data to capture dynamic transitions between surplus and deficit conditions, integrating optimisation or logistics models to support operational food distribution planning, and expanding the analysis to other regions to enhance generalizability and comparative insights.

Recommendations

Based on the analytical findings, the following strategic recommendations are proposed to strengthen food security and improve equitable food distribution across the Barlingmascakeb region:

1. Integrate Predictive Modelling into Local Food Distribution Policy (Pamisetty, 2024; Yang et al., 2024)
    a) Implement machine learning–based predictive models, such as AdaBoost, to enhance the accuracy and responsiveness of food distribution systems.
    b) These models have proven effective in detecting food-deficient areas early, enabling faster and more targeted government intervention.
    c) Moreover, they should be updated periodically with new data and expanded to cover other regions, promoting evidence-based and adaptive policymaking.

2. Prioritise Areas with Low Food Security Index (IKP) (Chukwuma et al., 2024; Smith et al., 2022)
   a) Areas with low IKP scores should be prioritised for distribution and intervention, as this variable is the strongest predictor of food deficit status.
   b) Concrete measures include improving logistics, food storage infrastructure, and access to essential food supplies in vulnerable regions.

3. Expand Access to Clean Water and Sanitation(Fotio & Nguea, 2022; Young, 2021)
   a) Access to clean water significantly influences food security outcomes.
   b) Infrastructure development for clean water supply and sanitation systems must be accelerated in underserved and rural communities.

4. Strengthen Healthcare Services in Vulnerable Areas (Tohit et al., 2025; Wang et al., 2022)
   a) High morbidity rates and limited healthcare personnel negatively affect food access and nutritional resilience.
   b) Expand healthcare access through mobile clinics, recruitment of community-based health workers, and rural health facility support.

5. Promote Nutrition Education and Empower Women (Pradhan et al., 2023; Sarker et al., 2024)
   a) Women's education, especially regarding family nutrition and health, correlates positively with household food security.
   b) Launch community training programs focused on food preparation, childcare nutrition, and household health management, targeting women in food-insecure areas.

6. Monitor Poverty and Food Expenditure as Socioeconomic Indicators (Marchetti & Secondi, 2022)
   a) High poverty rates and low household food expenditure proportions are critical socioeconomic vulnerability indicators.
   b) Introduce social assistance programs, food subsidies, and employment-based interventions in poverty-concentrated districts.

These recommendations are designed to guide local governments and stakeholders in creating a more resilient, inclusive, and data-driven food distribution ecosystem that is in line with sustainable development goals and regional equity.

References

Azies, H. Al. (2023). Integrating water indicators in a data-driven artificial intelligence model for food security classification. *The Journalish: Social and Government*, *4*(5), 168–180. https://doi.org/10.55314/TSG.V4I5.607

Bondansari, Widiatmaka, Machfud, Munibah, K., & Ambarwulan, W. (2023). The alternative policies to sustain the level of sufficient domestic rice staple food supply for the population (A case study in Banyumas Regency, Central Java, Indonesia). *IOP Conference Series: Earth and Environmental Science*, *1266*(1), 012032. https://doi.org/10.1088/1755-1315/1266/1/012032

Chukwuma, E. C., Afolabi, O. O. D., Okonkwo, C. C., Olamigoke, O. O., & Okonkwo, C. E. (2024). Application of geospatial technology and decision model in the development of improved food security index. *Scientific Reports*, *14*(1), 1–22. https://doi.org/10.1038/S41598-024-77845-1;SUBJMETA=166,172,639,704;KWRD=ENGINEERING,ENVIRONMENTAL+SCIENCES

Dharmawan, H., Sartono, B., Kurnia, A., Hadi, A. F., & Ramadhani, E. (2022). A study of machine learning algorithms to measure the feature importance in class-imbalance data of food insecurity cases in Indonesia. *Commun. Math. Biol. Neurosci.*, *2022*(0), Article ID 101. https://doi.org/10.28919/CMBN/7636

Fotio, H. K., & Nguea, S. M. (2022). Access to water and sanitation in Africa: Does globalization matter? *International Economics*, *170*, 79–91. https://doi.org/10.1016/J.INTECO.2022.02.005

Fransiska, H., Soleh, A. M., Notodiputro, K. A., & Erfiani. (2025). Evaluation of machine learning models based on household food insecurity data in Indonesia. *BIO Web of Conferences*, *171*, 02011. https://doi.org/10.1051/BIOCONF/202517102011

Ganaie, M. A., Hu, M., Malik, A. K., Tanveer, M., & Suganthan, P. N. (2022). Ensemble deep learning: A review. *Engineering Applications of Artificial Intelligence*, *115*, 105151. https://doi.org/10.1016/J.ENGAPPAI.2022.105151

Ghosh, C. (2022). Data Pre-processing. *Data Analysis with Machine Learning for Psychologists*, 55–85. https://doi.org/10.1007/978-3-031-14634-3_3

Granillo-Macías, R. (2021). Logistics optimization through a social approach for food distribution. *Socio-Economic Planning Sciences*, *76*, 100972. https://doi.org/10.1016/J.SEPS.2020.100972

Hariram, N. P., Mekha, K. B., Suganthan, V., & Sudhakar, K. (2023). Sustainalism: An integrated socio-economic-environmental model to address sustainable development and sustainability. *Sustainability 2023, Vol. 15, Page 10682*, *15*(13), 10682. https://doi.org/10.3390/SU151310682

Kavzoglu, T., & Teke, A. (2022). Predictive performances of ensemble machine learning algorithms in landslide susceptibility mapping using random forest, extreme gradient boosting (XGBoost) and natural gradient boosting (NGBoost). *Arabian Journal for Science and Engineering*, *47*(6), 7367–7385. https://doi.org/10.1007/S13369-022-06560-8/METRICS

Li, B., Wu, F., Lim, S.-N., Belongie, S., & Weinberger, K. Q. (2021). *On feature normalization and data augmentation* (pp. 12383–12392). https://github.com/Boyiliee/MoEx.

Marchetti, S., & Secondi, L. (2022). The economic perspective of food poverty and (in)security: An analytical approach to measuring and estimation in Italy. *Social Indicators Research*, *162*(3), 995–1020. https://doi.org/10.1007/S11205-021-02875-5/METRICS

Nie, P., Roccotelli, M., Fanti, M. P., Ming, Z., & Li, Z. (2021). Prediction of home energy consumption based on gradient boosting regression tree. *Energy Reports*, *7*, 1246–1255. https://doi.org/10.1016/J.EGYR.2021.02.006

Nisa, Z. K., Pradipta, A. G., Sholikah, L. N., Pratama, B. F., Prihanantya, A. S., Ngadisih, Susanto, S., & Arif, S. S. (2023). Recognition of agricultural land-use change with machine learning-based for regional food security assessment in Kulon Progo Plains area. *International Journal on Advanced Science, Engineering and Information Technology*, *13*(1), 54–61. https://doi.org/10.18517/IJASEIT.13.1.16550

Nur Aziza, L., Yuli Astuti, R., Akbar Maulana, B., & Hidayati, N. (2024). Penerapan algoritma K-nearest neighbor untuk klasifikasi ketahanan pangan di Provinsi Jawa Tengah. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, *4*(2), 404–412. https://doi.org/10.57152/MALCOM.V4I2.1201

Pamisetty, A. (2024). Integration of artificial intelligence and machine learning in national food service distribution networks. *SSRN Electronic Journal*. https://doi.org/10.2139/SSRN.5233353

Pradhan, M. R., Unisa, S., Rawat, R., Surabhi, S., Saraswat, A., Reshmi, R. S., & Sethi, V. (2023). Women empowerment through involvement in community-based health and nutrition interventions: Evidence from a qualitative study in India. *PLOS ONE*, *18*(4), e0284521. https://doi.org/10.1371/JOURNAL.PONE.0284521

Pratama, B. F., Sholikah, L. N. M., Nisa, Z. K., Pradipta, A. G., Ngadisih, Susanto, S., Prihanantya, A. S., Tirtalistyani, R., & Arif, S. S. (2023). Machine learning application for identifying the agricultural land use change to support regional food security estimation in Southern area of Kulon Progo. *AIP Conference Proceedings*, *2596*(1). https://doi.org/10.1063/5.0119219/2893424

Riptanti, E. W., Masyhuri, Irham, & Suryantini, A. (2022). The sustainability model of dryland farming in food-insecure regions: Structural equation modeling (SEM) approach. *International Journal of Sustainable Development and Planning*, *17*(7), 2033–2043. https://doi.org/10.18280/IJSDP.170704

Sarker, T., Roy, R., Yeasmin, S., & Asaduzzaman, M. (2024). Enhancing women's empowerment as an effective strategy to improve food security in rural Bangladesh: a pathway to achieving SDG-2. *Frontiers in Sustainable Food Systems*, *8*, 1436949. https://doi.org/10.3389/FSUFS.2024.1436949/BIBTEX

Sharma, R., & Sharma, S. K. (2025). Optimizing agricultural downstream supply chain: Addressing information asymmetry and losses. *Business Process Management Journal*, *ahead-of-print*(ahead-of-print). https://doi.org/10.1108/BPMJ-02-2024-0097/FULL/PDF

Smith, D. M., Rixson, L., Grove, G., Ziauddeen, N., Vassilev, I., Taheem, R., Roderick, P., & Alwan, N. A. (2022). Household food insecurity risk indices for English neighbourhoods: Measures to support local policy decisions. *PLOS ONE*, *17*(12), e0267260. https://doi.org/10.1371/JOURNAL.PONE.0267260

Tohit, N. F. M., Ya, R. M., & Haque, M. (2025). Bridging the gap: Integrating food security into healthcare for healthier futures – A scoping review. *Advances in Human Biology*, *15*(2), 177–199. https://doi.org/10.4103/AIHB.AIHB_22_25

Tuysuzoglu, G., & Birant, D. (2020). Enhanced bagging (eBagging): A novel approach for ensemble learning. *International Arab Journal of Information Technology*, *17*(4), 515–528. https://doi.org/10.34028/IAJIT/17/4/10

Virtriana, R., Riqqi, A., Anggraini, T. S., Fauzan, K. N., Ihsan, K. T. N., Mustika, F. C., Suwardhi, D., Harto, A. B., Sakti, A. D., Deliar, A., Soeksmantono, B., & Wikantika, K. (2022). Development of spatial model for food security prediction using remote sensing data in West Java, Indonesia. *ISPRS International Journal of Geo-Information 2022, Vol. 11, Page 284*, *11*(5), 284. https://doi.org/10.3390/IJGI11050284

Wang, J., Ding, X., Gao, H., & Fan, S. (2022). Reshaping food policy and governance to incentivize and empower disadvantaged groups for improving nutrition. *Nutrients 2022, Vol. 14, Page 648*, *14*(3), 648. https://doi.org/10.3390/NU14030648

Yang, Y., Zhao, C., & Lu, H. (2024). Leveraging machine learning for food waste reduction: An analysis of predictive models. *Applied and Computational Engineering*, *112*(1), 154–160. https://doi.org/10.54254/2755-2721/2025.18115

Young, S. L. (2021). Viewpoint: The measurement of water access and use is key for more effective food and nutrition policy. *Food Policy*, *104*, 102138. https://doi.org/10.1016/J.FOODPOL.2021.102138

Yousaf, A., Kayvanfar, V., Mazzoni, A., & Elomri, A. (2023). Artificial intelligence-based decision support systems in smart agriculture: Bibliometric analysis for operational insights and future directions. *Frontiers in Sustainable Food Systems*, *6*, 1053921. https://doi.org/10.3389/FSUFS.2022.1053921/BIBTEX

Yuliani, E., Sartono, B., Wijayanto, H., Hadi, A. F., & Ramadhani, E. (2022). Study of features importance level identification of machine learning classification model in sub-populations for food insecurity. *AIP Conference Proceedings*, *2668*(1). https://doi.org/10.1063/5.0111784/2832039

Zhang, T., Lin, W., Vogelmann, A. M., Zhang, M., Xie, S., Qin, Y., & Golaz, J. C. (2021). Improving convection trigger functions in deep convective parameterization schemes using machine learning. *Journal of Advances in Modeling Earth Systems*, *13*(5), e2020MS002365. https://doi.org/10.1029/2020MS002365