OPEN ACCESS
Journal of Multidisciplinary Applied Natural Science Vol. 6 No. 1 (2026) https://doi.org/10.47352/jmans.2774-3047.334

Check for updates

Research Article

Evaluation and Comparison of Three Mixed-Effect Models for Household Food Insecurity Classification in West Java

Herlin Fransiska, Agus Mohamad Soleh*, Khairil Anwar Notodiputro, and Erfiani Erfiani

Received : May 21, 2025	Revised : August 30, 2025	Accepted : August 31, 2025	Online : December 27, 2025

Abstract
Food insecurity is a complex and multidimensional issue that requires accurate predictions or classifications to design effective interventions. The availability of large and complex datasets has enabled the use of machine learning approaches. Moreover, because food insecurity data are often hierarchical/clustered, mixed effects modelling is well suited for capturing intergroup variation. This study compared three models: the generalized linear mixed model (GLMM), a parametric model that accounts for random effects but is limited to linear relationships; the generalized mixed effects tree (GMET), a flexible decision tree framework with random effects; and the generalized mixed effects random forest (GMERF), an ensemble of trees with random effects. The analysis used data from 25,890 households in West Java, Indonesia in 2021. Model evaluation showed that GMERF provided the best prediction results compared to the other models. This study concludes that integrating random forests with mixed-effects modelling offers a robust and effective approach for predicting household food insecurity. The primary predictors of food insecurity identified by the random forest in GMERF model were the age of the household head (years) and house size (m^2).

Keywords: food insecurity; GLMM; GMERF; GMET; hierarchical data

1. INTRODUCTION

Food insecurity is a complex condition in which individuals or households lack consistent access to sufficient, safe, and nutritious food for a healthy lifestyle. It encompasses not only inadequate food quantity but also unreliable access, poor nutritional intake, and psychological distress due to food uncertainty, all of which can adversely affect physical health, mental well-being, and overall quality of life [1][2]. Recognizing its urgency, the second Sustainable Development Goal (SDG 2), zero hunger [3], underscores the global commitment to eliminating food insecurity in all its forms.

Food insecurity remains a pressing concern in Indonesia. West Java warrants particular attention because, according to Statistics Indonesia (Badan Pusat Statistik, BPS) publication, based on the food insecurity experience scale (FIES), the prevalence of individuals experiencing moderate or severe food insecurity in 2021 reached 5.46%, which is higher than the national average of 4.79%. This issue is particularly noteworthy given that, according to the publication of BPS, West Java is the most densely populated province in the country, with a population density of 1,379 people per km². Compared with other highly populated provinces, West Java recorded a higher prevalence than East Java (2.98%) and Central Java (2.87%). Moreover, from 2017 to 2023, West Java consistently exceeded both the national average and the rates of other large provinces in Indonesia. These conditions position West Java as the most urgent case compared to other large provinces and as a representative region for examining how hierarchical and socioeconomic complexities drive food insecurity patterns in Indonesia.

Indonesia has incorporated the FIES into the national socioeconomic survey (SUSENAS). This survey follows a two-stage sampling design, in which census blocks are randomly selected first, followed by the selection of households within those blocks. Consequently, households within the same census block are not statistically independent, violating the assumption of independence required by most standard models. Ignoring this dependency can result in biased inferences, including underestimated standard errors [4], misidentification of vulnerable subgroups, and inappropriate variable selection [5][6]. This clustering effect highlights the need to incorporate random effects into the analytical framework of the

Publisher's Note:
Pandawa Institute stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Table 1. List of variables.

Variable	Description	Scale
Y	Household food insecurity classification	Nominal
X_1	Number of household members (persons)	Ratio
X_2	Gender of head of household	Nominal
X_3	Age of head of household (years)	Ratio
X_4	Literacy status of head of household	Nominal
X_5	Education level of head of household	Ordinal
X_6	Bank savings account ownership	Nominal
X_7	Health insurance contribution assistance recipient status	Nominal
X_8	Health insurance ownership status	Nominal
X_9	Smoking status of head of household	Nominal
X_{10}	Home ownership status	Nominal
X_{11}	House size (m^2)	Ratio
X_{12}	House wall material type	Nominal
X_{13}	House floor material type	Nominal
X_{14}	Home sanitation adequacy status	Nominal
X_{15}	Drinking water source feasibility status	Nominal
X_{16}	People's business credit recipient status	Nominal
X_{17}	Bank/cooperative loan recipient status	Nominal
X_{18}	Village-owned enterprise benefit recipient status	Nominal
X_{19}	Property/land ownership status	Nominal
X_{20}	Prosperous family card recipient status	Nominal
X_{21}	Family hope program recipient status	Nominal
X_{22}	Non-cash food assistance recipient status	Nominal
X_{23}	Other routine assistance recipient status	Nominal
R	Subdistrict (random effect)	Nominal

study. In addition, food insecurity is influenced by diverse social, economic, and demographic factors, with relationships that are often nonlinear. This complexity challenges traditional linear models. While machine learning methods have been applied to capture nonlinear patterns and interactions [7][8], conventional approaches, such as classification and regression trees (CART) and random forests, also assume independent observations. Thus, while standard machine learning techniques can capture nonlinear interactions, they fall short of addressing the data clustering inherent in large-scale socioeconomic surveys such as SUSENAS. To overcome these challenges, mixed-effects models offer a flexible solution by accommodating these structures through random effects and resulting in data analysis with higher validity and reproducibility of the experimental findings [9].

In classical statistics, this extension refers to the incorporation of random effects into the modeling framework, enabling the transition from generalized linear models (GLM), which assume independent observations, to generalized linear mixed models (GLMM). Mixed models, such as the GLMM, allow for the handling of data with a hierarchical or clustered structure and explicitly consider random effects, which is crucial in the context of food insecurity, which often varies between regions. Similarly, in the machine learning context, decision tree models such as CART can be extended to generalized mixed-effects trees (GMET) [10], while random forests can be extended to generalized mixed-effects random forests (GMERF) [11]. These tree-based mixed-effects approaches accommodate hierarchical dependencies and capture nonlinear relationships and complex interactions among predictor variables without relying on strict distributional assumptions.

GLMM has been widely applied in public health and socioeconomic research to address clustered survey data; however, the adoption of mixed-effects machine learning methods, such as GMET and GMERF, remains relatively limited. Existing applications of GMET and GMERF have primarily been applied in agriculture, education, medical research, and social sciences and economics [12]-[16], where hierarchical data structures and complex interactions are prevalent. However, their application to socioeconomic survey data, particularly in the context of food insecurity, is scarce. This gap highlights the need to evaluate whether these approaches can provide additional benefits over classical mixed models in capturing the nonlinear effects and interactions among the determinants of food insecurity. Therefore, this study aims to compare the performance of three mixed-model approaches (GLMM, GMET, and GMERF) in modeling household food insecurity in West Java Province in 2021. By evaluating the predictive performance and model characteristics of these three approaches, this study sought to identify the most effective method for the accurate classification of household food insecurity in explaining variations in food insecurity and to identify significant factors that contribute to the level of food insecurity. By understanding how each model works in the context of these data, the results of this study are expected to provide evidence-based recommendations that can be used to support more effective and targeted policies for alleviating food insecurity.

2. MATERIALS AND METHODS

2.1. Materials

This study examined household food insecurity using secondary data from BPS. The response variable was derived from the FIES for West Java, Indonesia, in 2021. A household was classified as food insecure if it responded affirmatively ("yes") to any of the eight items in the FIES. The dataset comprises 25,890 households. The analysis included one response variable, 23 fixed-effect predictor variables, and one random-effect predictor variable, all obtained from the SUSENAS, as summarized in Table 1.

2.2. Methods

2.2.1. Framework

The performances of the GLMM, GMET, and GMERF were evaluated in this study. Figure 1 illustrates the stages of the research, which consist of data, data preprocessing, modelling, evaluation, and interpretation. Stage 1 (data); the household food insecurity dataset was obtained from a survey

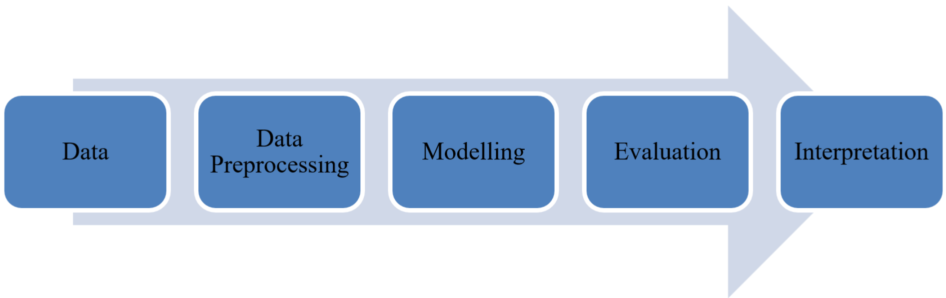

Figure 1. Stages of the research.

Table 2. The confusion matrix.

Prediction	Actual	
	0	**1**
0	True Positive	False Positive
1	False Negative	True Negative

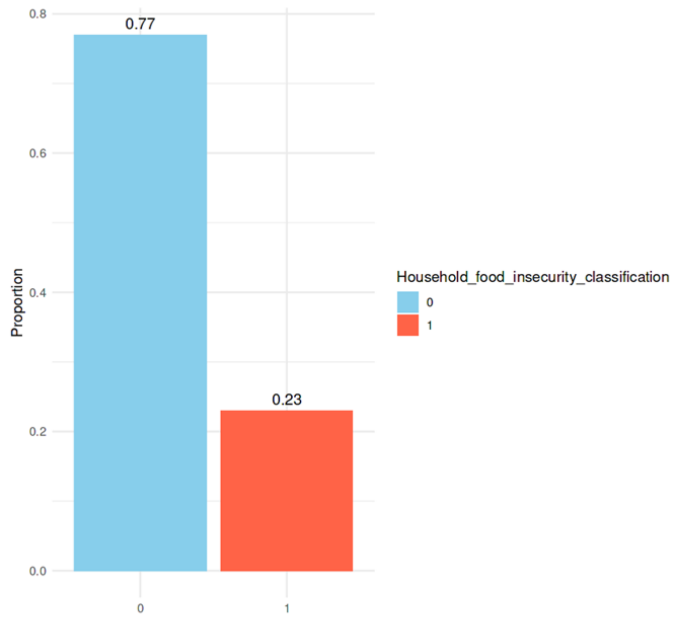

Figure 2. Shows proportion of household food insecurity (23% insecure, 77% secure).

conducted in West Java, Indonesia, in 2021. Stage 2 (preprocessing); the dataset was cleaned to address data quality problems. Because the proportion of missing data was very small, observations with missing values were removed rather than imputed to maintain data integrity and avoid introducing unnecessary bias. In addition, near-zero variance (NZV) variables were removed because they are predictors with very little variability or highly imbalanced distributions, which contribute minimal information and may reduce computational efficiency [17][18]. These variables were identified using diagnostic functions such as NZV in the R caret package and were excluded prior to modeling. Stage 3 (modelling); the GLMM, GMET, and GMERF models were employed. To ensure a robust model evaluation, a repeated train–test split procedure was applied, where the dataset was randomly divided into 70% training and 30% testing sets. This process was repeated 50 times, and the performance metrics were computed for each repetition and then averaged across all iterations. This approach reduces the variance in model evaluation and provides a more reliable estimate of predictive performance compared to a single train–test split. Parameter tuning followed the default settings of the GMET and GMERF functions, including the standard number of trees and hyperparameters. Stage 4 (evaluation); model performance was assessed using accuracy, sensitivity, specificity, balanced accuracy, kappa, and AUC to identify the optimal model. This is an important stage for the model performance and prediction precision. Stage 5 (interpretation); the best model was interpreted to identify the key variables contributing to household food insecurity, ensuring valid and reliable insights.

2.2.2. Methods

Many studies have been conducted on mixed models [4][5][6][10][11][13][15][19]-[24]. This study used the GLMM, GMET, and GMERF. The GLMM are flexible statistical models that extend traditional linear models to analyze data with responses that are not necessarily normally distributed (by selecting a distribution from the exponential family, such as binomial), accommodate hierarchical or correlated data structures by including random effects to capture between-group variability, and allow for modelling nonlinear relationships between predictors and response means [12]. The GLMM are statistical frameworks for modelling data characterized by a combination of four key components: fixed effects, which quantify the impact of predictor variables on the response variable; random effects, designed to capture intergroup variability within hierarchical or clustered data structures; a link function, which establishes the relationship between the linear combination of predictors and the mean of the response variable; and an error distribution, which specifies the probability distribution of the response variable, typically drawn from the exponential family [24].

Given $y_{ij} = (y_{ij}, y_{ij}, \ldots, y_{inj})$ with observation units i, $i = 1,2,\ldots,n_j$, in groups j, $j = 1,2,\ldots,J$. Y_{ij} is assumed to follow an exponential family distribution using Equation (1), as follows;

$$f_i(y_{ij}|b_j) = exp\left\{\frac{y_{ij}\eta_{ij} - a(\eta_{ij})}{\phi} + c(y_{ij}, \phi)\right\} \quad (1)$$

Where b_j is a random component, a and c are specific functions, η_{ij} is a natural parameter, and ϕ is a dispersion parameter. The mean and variance of y_{ij} are respectively $E(y_{ij}|b_j) = a'(\eta_{ij}) = \mu_{ij}$ and $Var(y_{ij}|b_j) = \psi a''(\eta_{ij})$ [15]. It shows the general form of distributions in the exponential family (e.g., normal, binomial, and Poisson). The key idea is that the mean μ_{ij} and variance depend on the natural parameter η_{ij}, linking the data distribution to the model structure. Given the canonical function $g(a')^{-1}$, which connects the mean with the systematic component, the GLMM formula can be expressed by Equation (2);

$$\begin{aligned} \mu_j &= E(y_j|b_j), j = 1,2,\ldots,J \\ g(\mu_j) &= \eta_j \\ \eta_j &= X_j\beta + Z_jb_j \\ b_j &\sim N_Q(0, \psi) \end{aligned} \quad (2)$$

where j is the group index, and J is the number of groups. n_j is the number of observations in the j-th group and $\Sigma_{j=1}^{J} n_j = J$. η_j is a linear predictor vector of dimension n_j, where X_j is a fixed-effect predictor variable matrix of size n_j x $P = p + 1$, and β is a coefficient vector of predictor variables of size P. Z_j is a n_j x Q matrix of random effects regression, b_j is a Q-dimensional vector of coefficients (including random intercepts), and ψ is a Q x Q matrix of the variance of random effects [25]. Fixed effects were identified using parameters related to the entire population, whereas random effects were identified using group-specific parameters. Estimation methods include maximum likelihood, restricted maximum likelihood, and penalized quasi-likelihood [15][23]. The estimation procedure iteratively maximizes the log-likelihood, and convergence is declared once the relative change in the log-likelihood or parameter estimates between iterations falls below a specified tolerance (commonly set at 1e-6). This diagnostic process ensures the numerical stability and reliability of the parameter estimates while preventing premature termination of the optimization. If convergence warnings arise, model re-specification or alternative optimizers may be required to achieve stable solutions [26].

The GMET was presented by Fontana et al. [13][23]. Fundamentally, the GMET replaces the linear function of fixed effects in a traditional GLMM with a tree-based method [19]. Given the canonical function $g(a')^{-1}$, which connects the mean with the systematic component, the GMET model can be expressed by Equation (3).

$$\begin{aligned} \mu_j &= E(y_j|b_j), j = 1,2,\ldots,J \\ g(\mu_j) &= \eta_j \\ \eta_j &= f(x_j) + Z_jb_j \\ b_j &\sim N_Q(0, \psi) \end{aligned} \quad (3)$$

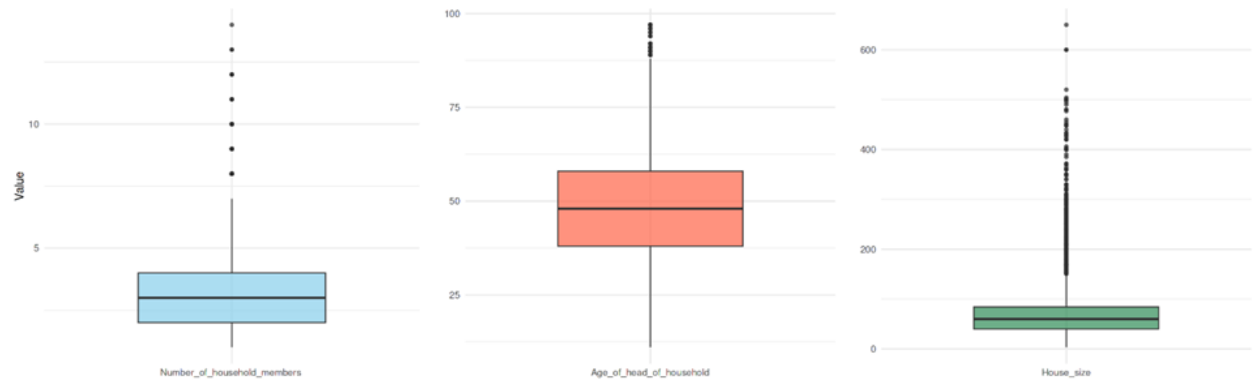

Figure 3. The boxplot of numerical variables.

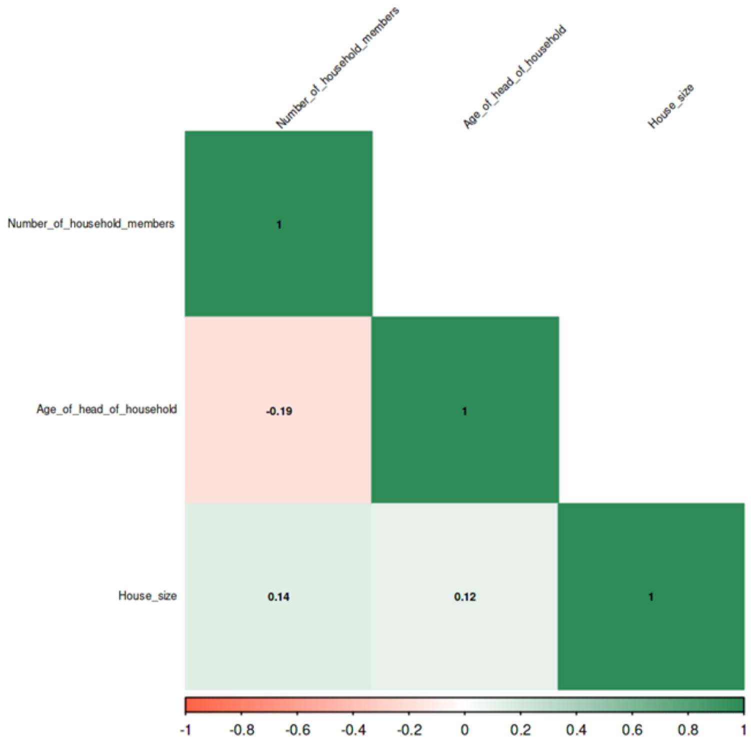

Figure 4. Headmap of numeric variable predictor.

Consequently, GMET extends the capabilities of GLMMs by enabling the modelling of nonlinear and interactive fixed effects through the integration of a tree-based structure, while still effectively addressing dependencies within grouped data via random effects. The linear function X_jb in Equation (2) of the GLMM becomes a function of $f(x_j)$ in Equation (3), which is computed using a tree-based algorithm that represents this flexible fixed-effect specification. The GMET algorithm uses a non-iterative approach. It first obtains an initial estimate using only fixed effects and then uses these estimates to build a tree structure that captures potential interactions or nonlinearities. Finally, a mixed-effects model was fitted using the tree structure as a fixed effect to estimate the random effects and refine the predictions within each segment defined by the tree.

The GMERF is an ensemble machine learning method that extends the conventional RF framework to effectively handle data with hierarchical or clustered structures by integrating principles from mixed-effects modeling. The fundamental concept underlying GMERF is rooted in the GMET, which uniquely incorporates both fixed and random effects within the structure of a single tree to model heterogeneity across different groups. GMERF can be expressed by Equation (3), where function $f(x_j)$ in $\eta_j = f(x_j) + Z_jb_j$ in Equation (3) is computed using the random forest algorithm in GMERF method. The initial approach for estimating the parameters of a GMERF involved an iterative algorithm that alternated between fitting a random forest for fixed effects and a generalized linear mixed model for random effects [15]. In the implementation of the GMERF model, the maximum number of iterations was restricted to 30, following Pellagati's R-code implementation. This upper bound was imposed to avoid indefinite looping in the estimation process, as the GMERF relies on iterative updates of fixed and random effects until convergence is reached. A threshold of 30 iterations is commonly adopted as a practical compromise, providing a balance between computational feasibility and probability of convergence. When convergence was not achieved within this limit, the model was considered unstable. Such instability may arise from factors such as excessive model complexity, strong collinearity among predictors, and weak identifiability of the random-effects structure. In these cases, convergence diagnostics and model simplification are necessary to improve the estimation stability.

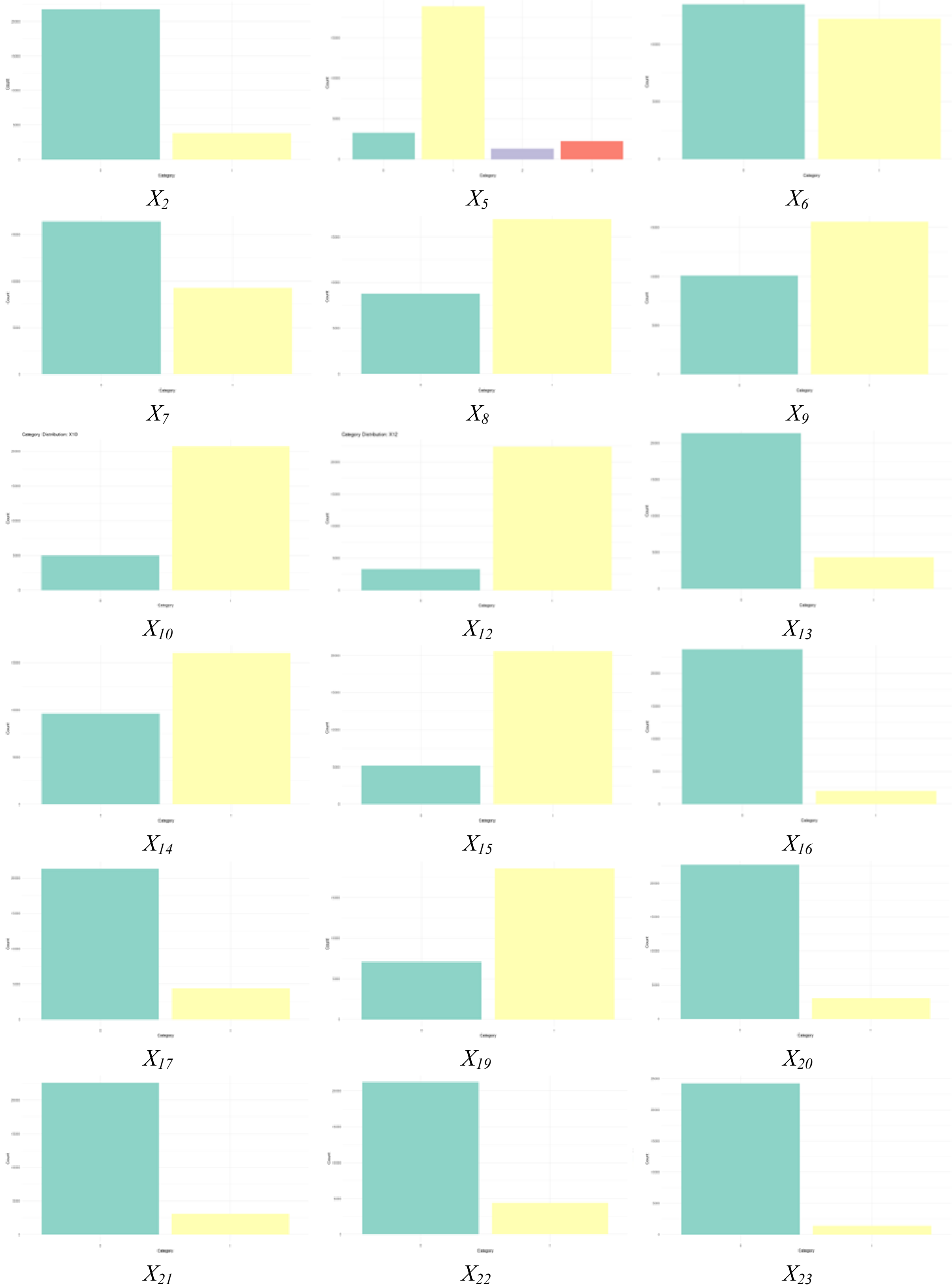

Figure 5. Frequency distribution of categorical variable predictor.

In this study, the fixed effects $X_j\beta$ or $f(x_j)$ capture the overall influence of socioeconomic characteristics such as income, household size, and education level, whereas the random effects Z_jb_j account for unobserved heterogeneity across subdistricts. Thus, two households with similar socioeconomic profiles may still differ in their food insecurity status depending on their subdistrict-specific conditions. To evaluate the performance of the classification model on the test data, we used a series of standard metrics, including accuracy, sensitivity, specificity, balance accuracy, kappa, and AUC. These metrics provide a comprehensive picture of the overall prediction accuracy, ability to identify positive and negative classes, and corrected prediction agreement, and their calculations are based on confusion matrices. The confusion matrix is presented in Table 2 [27]-[29]. Among these, AUC is particularly valuable because it evaluates the model's ability to discriminate between classes across all possible thresholds, rather than being tied to a single cutoff point. This makes AUC especially robust in contexts with class imbalance, as it captures the tradeoff between sensitivity and specificity in a single measure and reflects the model's overall discriminative power.

3. RESULTS AND DISCUSSIONS

3.1. Summary Statistics

The analysis in this study was conducted on 25,873 of the 25,890 observations (households) initially included in the dataset. This was a result of the data cleansing process, which involved the deletion of household records containing 'do not know' or 'do not respond' responses. The collected data were analyzed using advanced statistical techniques to interpret the methods' effectiveness. Figure 2 shows the distribution of the proportion of food insecurity in the data. Most households (approximately 77%) are in the category of food security (Y = 0), while only 23% are classified as food insecurity (Y = 1). This imbalance in proportion indicates that food insecurity is a minority phenomenon in the available data. The implications of this imbalance need to be considered in further analysis processes, especially when building a classification model, so that the model is not biased towards the majority group and is still able to properly identify households classified as food insecure.

Based on the results of the near-zero variance test, variables literacy status of head of household (X_4) and Village-owned enterprise benefit recipient status (X_{18}) had very low variances, indicating that the values of these variables hardly varied across observations. This condition indicates that X_4 and X_{18} do not provide sufficient information to distinguish between food insecurity and food security households. Kumar and Ramasree stated that removing uncommon features from the training set based on NZV can improve the model performance [30]. Therefore, both variables were not considered statistically significant in the context of predictive modelling and needed to be removed from further analysis to improve model efficiency and accuracy.

Figure 3 shows that the variables number of household members (X_1), age of head of household (X_3), and house size (X_{11}) have different distribution characteristics. The variable number of household members has a data distribution that tends to be concentrated between values 2 and 4, with a median of approximately 3, and several outliers above 7 indicate a right-skewed bias. Meanwhile, the variable age of the head of household has a relatively symmetrical distribution with a median of approximately 48, and the majority of the data are between 38 and 58, with a few outliers above 85. In contrast, variable house size shows a distribution that is skewed to the right, with the main data concentration below 150, but accompanied by many extreme outliers reaching more than 600. This indicates that house size has very high variability and a non-normal distribution. Outliers were retained throughout the process to preserve the original variability of the data and ensure that extreme values remained analytically meaningful.

Figure 4 shows that the correlation heatmap between number of household members (X_1), age of head of household (X_3), and house size (X_{11}) showed no strong linear relationship between these three variables. The correlation between the number of household members and the age of the head of the household was very weakly negative, whereas the correlation between the number of household members and house size and between the age of the head of the household and house size was weakly

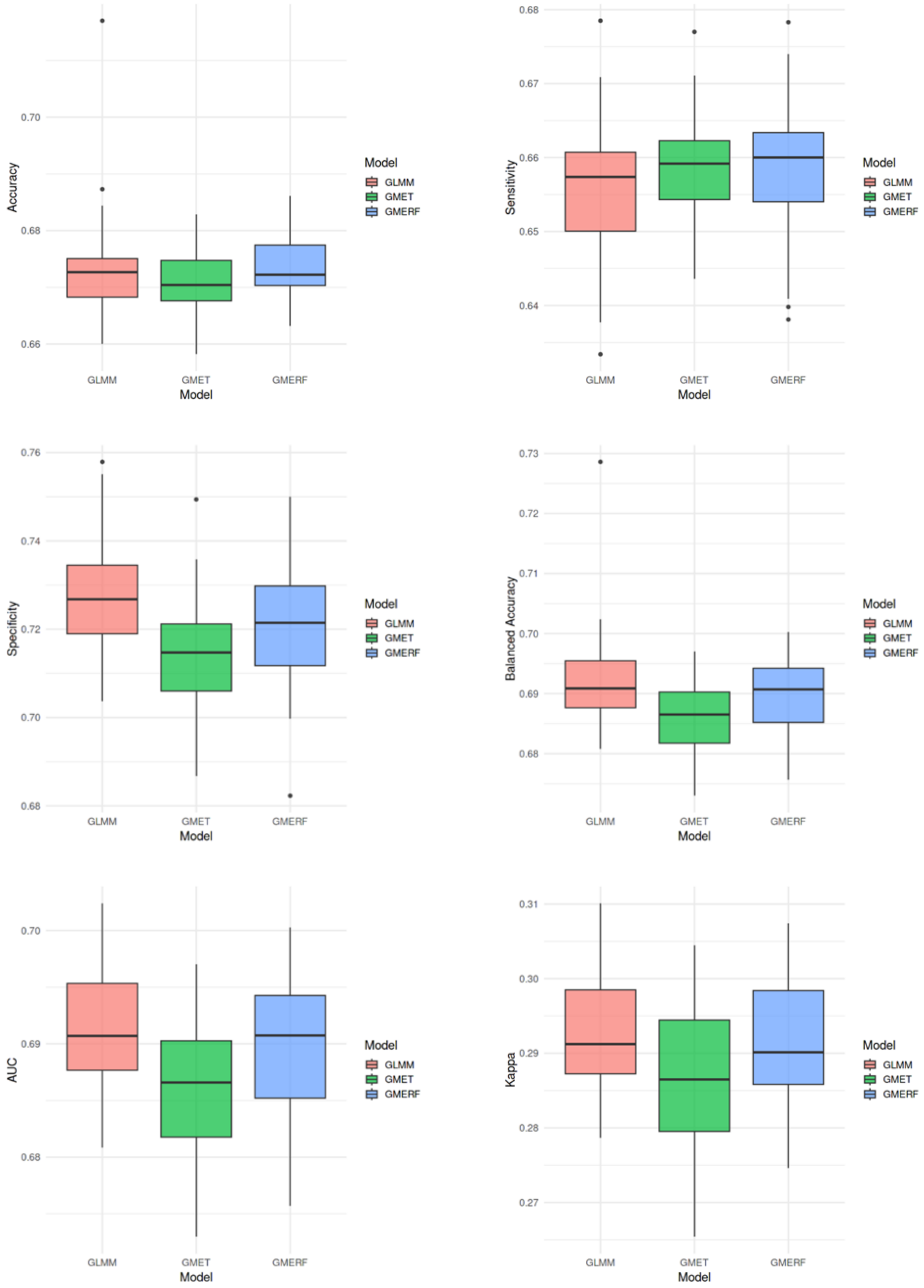

Figure 6. Shows the boxplots of the evaluation metrics three models.

positive. Thus, these three variables are relatively independent of each other in the context of a linear relationship.

Figure 5 shows that the graph visualizes the frequency distributions of the categorical predictor variables. Each bar represents the number of observations in each category for the respective variable. This allowed us to observe the dominance or rarity of specific categories, identify any imbalances in the data, and understand the general structure of the categorical inputs. Based on Figure 5, it can be seen that the dominance of category zero (0) are variables gender of head of household (X_2), house floor material type (X_{13}), people's business credit recipient status (X_{16}), bank/ cooperative loan recipient status (X_{17}), prosperous family card recipient status (X_{20}), family hope program recipient status (X_{21}), non-cash food assistance recipient status (X_{22}) and other routine assistance recipient status (X_{23}), while the dominance of category one (1) are home ownership status (X_{10}), house wall material type (X_{12}), drinking water source feasibility status (X_{15}) and property/ land ownership status (X_{19}) while others tend to be balanced. Statistical modelling was performed to understand the relationship between the predictor variables and food insecurity.

3.2. Modeling Results

This study modeled household food insecurity while accounting for the structure of the data (random effect), where households within the same sub-district shared more similarities than those from different sub-districts. The presence of subdistrict-level random effects was confirmed through a likelihood ratio test and the intraclass correlation coefficient (ICC), with a significant effect ($p < 0.05$) and an ICC of 36.06%. Based on Figure 6, the boxplot presents a comparison of the model performance across several evaluation metrics. It highlights the distribution, central tendency, and variability of the three modelling approaches: GLMM, GMET, and GMERF. The boxplots of accuracy show that all three models achieved very similar sensitivity results. GMERF had the highest median sensitivity, followed closely by GLMM, whereas GMET lagged slightly behind. The variability across the models was comparable, with a few outliers observed in the GLMM. The boxplots of sensitivity show that GMERF performed better than both GMET and GLMM. GMET showed the smallest variability, with a few scattered outliers observed in all models.

The boxplots of specificity show that the balanced accuracy of GLMM was the highest, followed by GMERF and GMET. The variability across the models was comparable, with a few outliers observed in the results of all the models. The boxplots of the balanced accuracy show that the GMERF and GLMM models achieve very similar results, but the GLMM has an outlier. The variability across the models was comparable. The boxplots of AUC and kappa show that GLMM and GMERF again display good and almost identical performances, whereas GMET records the weakest performance in these metrics. Variability across models is comparable. Overall, GMERF showed the highest predictive performance in terms of median and stability, although the differences among the three models were relatively small. GLMM remains competitive but shows slightly less consistency, whereas GMET performs worse than the other approaches.

Based on Table 3, the results indicate that the

Table 3. Performance measure.

Algorithm	Mean and Standard Deviations Performance Measure					
	Accuracy	**Sensitivity**	**Specificity**	**Balance Accuracy**	**Kappa**	**AUC**
GLMM	0.673 (0.0083)	0.727 (0.0119)	0.727 (0.0119)	0.692 (0.00743)	0.293 (0.00832)	0.691 (0.00526)
GMET	0.671 (0.0053)	0.714 (0.0124)	0.714 (0.0124)	0.686 (0.00622)	0.286 (0.00955)	0.686 (0.00622)
GMERF	0.673 (0.0055)	0.721 (0.0130)	0.721 (0.0130)	0.690 (0.00563)	0.292 (0.00870)	0.690 (0.00563)

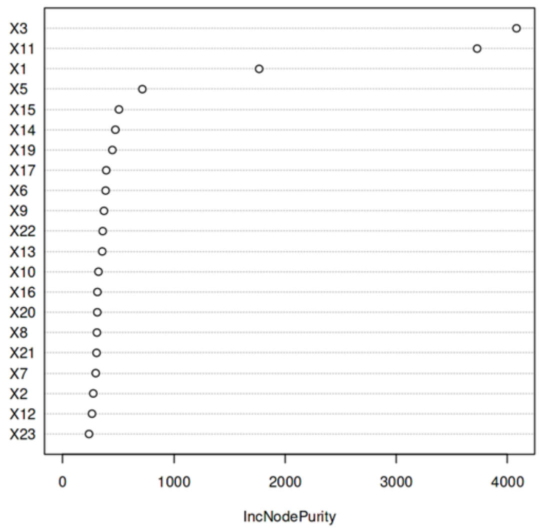

Figure 7. Shows the Increase in Node Purity (IncNodePurity) plot.

GMERF and GLMM models achieve very similar mean values across the key performance metrics, which is consistent with the patterns observed in their respective boxplots. In comparison, GMERF outperforms GMET because GMET relies on a single decision tree, which is highly sensitive to initial splits and prone to overfitting. In contrast, GMERF aggregates multiple trees, reducing the variance and producing more stable and reliable predictions. Taken together, the results from both the boxplot visualization and descriptive statistics suggest that all three models perform nearly equivalently on average, with only subtle differences across the metrics. Nevertheless, the GMERF demonstrated greater robustness. This robustness is particularly valuable in applied contexts where stability is as important as the point accuracy. Consequently, GMERF can be considered a preferable alternative when modeling objectives emphasize consistency and generalizability. These findings are consistent with those of Pellagati *et al.* [15]. who reported that the GMERF algorithm performs comparably to GLMM, particularly when fixed effects are larger than random effects, but is more robust in its estimates.

3.3. Interpretation and Discussion

For illustrative purposes, we present the results of one iteration out of 50 iterations to show the modeling and interpretation of the outputs. The GMERF model produced a relatively large random intercept variance (1.845 with a standard deviation of 1.358), indicating considerable heterogeneity between the sub-district groups (R). This shows that food insecurity is not solely shaped by household-level characteristics but is also strongly influenced by contextual factors at the sub-district level. In other words, some sub-districts face higher risks of food insecurity than others, regardless of individual household attributes. Such disparities may stem from structural differences in infrastructure, accessibility of health and education services, local food distribution systems, or socioeconomic development across sub-districts in West Java Province. These findings emphasize the importance of incorporating area-level characteristics into the analysis and suggest that equitable development policies for sub-districts are crucial for effectively reducing household food insecurity in the region.

The RF component of the GMERF model identified the most influential household-level factors for predicting food insecurity. The importance of the predictor variables was assessed using the Increase in Node Purity (IncNodePurity) measure. This metric is derived from the random forest algorithm, where each split of a decision tree aims to increase the homogeneity (or purity) of the resulting node. Variables that consistently produce larger reductions in node impurity across many trees in the forest receive higher IncNodePurity values. Therefore, a higher value indicates that the variable plays a more influential role in improving the model accuracy and explaining the variability in the response.

Based on Figure 7, the important variables of the RF model based on IncNodePurity are X_3 (4080.72) and X_{11} (3726.68). This highlights that the age of the head of household (X_3) and house size (X_{11}) appear to be the strongest discriminators and exhibit high IncNodePurity scores, implying that these characteristics substantially contribute to the classification of household food insecurity. Conversely, variables with low scores had minimal influence on explaining disparities in food insecurity. The findings indicate that household food insecurity in West Java is influenced by both individual and sub-district factors. At the household level, older household heads may have a limited capacity to adapt to economic or social changes, thereby increasing the risk of food insecurity. Similarly, smaller house sizes may reflect limited household resources and socioeconomic vulnerability, which restricts the ability to secure adequate food. In addition, the analysis revealed a significant random effect at the sub-district level, indicating that the risk of food insecurity is also shaped by broader regional contexts, such as differences in local infrastructure, public service availability, and regional development. These findings highlight the importance of a multilevel strategy. At the household level, targeted programs should support vulnerable groups, such as older household heads and families living in smaller housing conditions, for example, through social assistance, skills training, and livelihood support. At the regional level, policies should promote equitable development across sub-districts, focusing on strengthening infrastructure, improving access to health and education services, and ensuring the availability of local food markets. By addressing both household vulnerabilities and regional disparities, policymakers can foster more inclusive and sustainable efforts to reduce food insecurity in West Java.

4. CONCLUSIONS

This study shows that the GMERF model outperforms the GLMM and GMET models in predicting household food insecurity owing to its robustness and ability to capture complex patterns. Food insecurity in West Java is influenced by both household-level characteristics, particularly the age of the household head and housing size, and sub-district-level factors represented as a random effect. These findings highlight the importance of integrating household-level interventions with regionally targeted development policies to effectively reduce food insecurity in the region. Despite these contributions, this study has some limitations. The interpretability of ensemble models, such as the GMERF, remains challenging compared to simpler approaches. Future research could extend this framework to the generalized mixed effect generalized random forest (GMEGRF), which allows for richer inference and better handling of heterogeneous effects. Additionally, subsequent studies should systematically address data imbalances and outliers to further enhance predictive reliability and provide more robust guidance for policymaking.

AUTHOR INFORMATION

Corresponding Author

Agus Mohamad Soleh — Statistics and Data Science Study Program, IPB University, Bogor-16680 (Indonesia);

orcid.org/0000-0002-2732-1985

Email: agus.ms@apps.ipb.ac.id

Authors

Herlin Fransiska — Department of Mathematics, University of Bengkulu, Bengkulu -38371 (Indonesia); Statistics and Data Science Study Program, IPB University, Bogor-16680 (Indonesia);

orcid.org/0000-0002-7983-5590

Khairil Anwar Notodiputro — Statistics and Data Science Study Program, IPB University, Bogor-16680 (Indonesia);

orcid.org/0000-0003-2892-4689

Erfiani Erfiani — Statistics and Data Science Study Program, IPB University, Bogor-16680 (Indonesia);

orcid.org/0000-0001-5502-7321

Author Contributions

Conceptualization, Methodology, Validation, Funding Acquisition, Writing—Review & Editing, H. F., A. M. S., K. A. N., and E.; Software, H. F. and A. M. S.; Data Curation and Writing—Original

Draft Preparation, H. F., A. M. S., and K. A. N.; Visualization and Project Administration, H. F.

Conflicts of Interest

The authors declare no conflict of interest.

ACKNOWLEDGEMENT

This research was supported by The Center for Higher Education Funding and Assessment (PPAPT) and The Indonesian Endowment Fund for Education (LPDP), Ministry of Finance of Republic Indonesia, through the Indonesian Education Scholarship. The authors sincerely thank these institutions for their support.

DECLARATION OF GENERATIVE AI

Not applicable.

REFERENCES

[1] K. Essel and K. A. Courts. (2018). In: " H. Kersten, A. Beck, and M. Klein (Eds) Identifying and Addressing Childhood Food Insecurity in Healthcare and Community Settings". Cham: Springer. 10.1007/978-3-319-76048-3_1.

[2] K. Schroeder, A. Smaldone, and S. Smaldone. (2015). "Food Insecurity: A Concept Analysis". *Nursing Forum.* **50** (4): 274-284. 10.1111/nuf.12118.

[3] F. Sporchia, M. Antonelli, A. Aguilar-Martínez, A. Bach-Faig, D. Caro, K. F. Davis, R. Sonnino, and A. Galli. (2024). "Zero Hunger: Future Challenges and the Way Forward Towards the Achievement of Sustainable Development Goal 2". *Sustainable Earth Reviews.* **7** (1). 10.1186/s42055-024-00078-7.

[4] S. W. Raudenbush and A. S. Bryk. (2022). "Hierarchical Linear Models: Applications and Data Analysis Methods". SAGE Publications, Newbury Park, CA.

[5] M. Fokkema, N. Smits, A. Zeileis, T. Hothorn, and H. Kelderman. (2018). "Detecting Treatment–Subgroup Interactions in Clustered Data with Generalized Linear Mixed-Effects Model Trees". *Behavior Research Methods.* **50** : 2016-2034. 10.3758/s13428-017-0971-x.

[6] R. J. Sela and J. S. Simonoff. (2012). "RE-EM Trees: A Data Mining Approach for Longitudinal and Clustered Data". *Machine Learning.* **86** (2): 169-207. 10.1007/s10994-011-5258-3.

[7] H. Dharmawan, B. Sartono, A. Kurnia, A. F. Hadi, and E. Ramadhani. (2022). "A Study of Machine Learning Algorithms to Measure Feature Importance in Class-Imbalance Data of Food Insecurity Cases in Indonesia". *Communications in Mathematical Biology and Neuroscience.* **2022**. 10.28919/cmbn/7636.

[8] K. N. Khikmah, B. Sartono, B. Susetyo, and G. A. Dito. (2024). "Performance Comparative Study of Machine Learning Classification Algorithms for Food Insecurity Experience by Households in West Java". *Jurnal Online Informatika.* **9** (1): 128-137. 10.15575/join.v9i1.1012.

[9] Z. Yu, M. Guindani, S. F. Grieco, L. Chen, T. C. Holmes, and X. Xu. (2022). "Beyond t Test and ANOVA: Applications of Mixed-Effects Models for More Rigorous Statistical Analysis in Neuroscience Research". *Neuron.* **110** (1): 21-35. 10.1016/j.neuron.2021.10.030.

[10] A. Hajjem, F. Bellavance, and D. Larocque. (2011). "Mixed Effects Regression Trees for Clustered Data". *Statistics & Probability Letters.* **81** (4): 451-459. 10.1016/j.spl.2010.12.003.

[11] A. Hajjem, F. Bellavance, and D. Larocque. (2014). "Mixed-Effects Random Forest for Clustered Data". *Journal of Statistical Computation and Simulation.* **84** (6): 1313-1328. 10.1080/00949655.2012.741599.

[12] B. Suseno, K. A. Notodiputro, and B. Sartono. (2023). "GLMMTree for Modelling Poverty in Indonesia". *Proceedings of the International Conference on Data Science and Official Statistics*. 121-131. 10.34123/icdsos.v2023i1.333

[13] Sukarna, K. A. Notodiputro, and B. Sartono. (2023). "Comparison Between Binomial GLMM and Binomial GMET for Temporary Unemployment in West Java, Indonesia".

Proceedings of the 5th International Conference on Statistics, Mathematics, Teaching, and Research (ICSMTR 2023). 198-209. 10.2991/978-94-6463-332-0_22

[14] D. Kusumaningrum, H. Wijayanto, A. Kurnia, K. A. Notodiputro, and M. Ardiansyah. (2024). In: "Mathematical and Statistical Methods for Actuarial Sciences and Finance". Cham: Springer. 10.1007/978-3-031-64273-9_35.

[15] M. Pellagatti, C. Masci, F. Ieva, and A. M. Paganoni. (2021). "Generalized Mixed-Effects Random Forest: A Flexible Approach to Predict University Student Dropout". *Statistical Analysis and Data Mining: The ASA Data Science Journal.* **14** (3): 241-257. 10.1002/sam.11505.

[16] F. Asadi, R. Homayounfar, Y. Mehrali, C. Masci, S. Talebi, and F. Zayeri. (2024). "Detection of Cardiovascular Disease Cases Using Advanced Tree-Based Machine Learning Algorithms". *Scientific Reports.* **14** : 22230. 10.1038/s41598-024-72819-9.

[17] G. Y. Lee, L. Alzamil, B. Doskenov, and A. Termehchy. (2021). "A Survey on Data Cleaning Methods for Improved Machine Learning Model Performance". *arXiv Preprint.* 10.48550/arXiv.2109.07127.

[18] P. Agasthi, H. Ashraf, S. H. Pujari, M. Girardo, A. Tseng, F. Mookadam, N. Venepally, M. R. Buras, B. Abraham, B. K. Khetarpal, and M. Allam. (2023). "Prediction of Permanent Pacemaker Implantation After Transcatheter Aortic Valve Replacement: The Role of Machine Learning". *World Journal of Cardiology.* **15** (3): 95-105. 10.4330/wjc.v15.i3.95.

[19] A. Hajjem, D. Larocque, and F. Bellavance. (2017). "Generalized Mixed Effects Regression Trees". *Statistics & Probability Letters.* **126** : 114-118. 10.1016/j.spl.2017.02.033.

[20] J. Hu and S. Szymczak. (2023). "A Review on Longitudinal Data Analysis with Random Forest". *Briefings in Bioinformatics.* **24** (2): 1-11. 10.1093/bib/bbad002.

[21] P. Krennmair and T. Schmid. (2022). "Flexible Domain Prediction Using Mixed Effects Random Forests". *Journal of the Royal Statistical Society: Series C (Applied Statistics).* **71** (5): 1865-1894. 10.1111/rssc.12600.

[22] J. L. Speiser, B. J. Wolf, D. Chung, C. J. Karvellas, D. G. Koch, and V. L. Durkalski. (2020). "BiMM Tree: A Decision Tree Method for Modeling Clustered and Longitudinal Binary Outcomes". *Communications in Statistics – Simulation and Computation.* **49** (4): 1004-1023. 10.1080/03610918.2018.1490429.

[23] L. Fontana, C. Masci, F. Ieva, and A. M. Paganoni. (2021). "Performing Learning Analytics via Generalized Mixed-Effects Trees". *Data.* **6** : 74. 10.3390/data6070074.

[24] E. Setiawan, K. A. Notodiputro, and B. Sartono. (2024). "Generalized Linear Mixed-Model Tree for Modeling Dengue Fever Cases". *COGITO Smart Journal.* **10** (2): 380-392. 10.31154/cogito.v10i2.715.380-392.

[25] W. W. Stroup, M. Ptukhina, and J. Garai. (2024). "Generalized Linear Mixed Models". CRC Press, New York. 10.1201/9780429092060.

[26] D. Bates, M. Mächler, B. Bolker, and S. Walker. (2015). "Fitting Linear Mixed-Effects Models Using lme4". *Journal of Statistical Software.* **67** (1): 1-48. 10.18637/jss.v067.i01.

[27] I. Sriliana, S. Nugroho, W. Agwil, and E. D. Sihombing. (2025). "Evaluation of Multivariate Adaptive Regression Splines on Imbalanced Dataset for Poverty Classification in Bengkulu Province". *Barekeng: Jurnal Ilmu Matematika dan Terapan.* **19** (2): 1143-1156. 10.30598/barekengvol19iss2pp1143-1156.

[28] M. Heydarian, T. E. Doyle, and R. Samavi. (2022). "MLCM: Multi-Label Confusion Matrix". *IEEE Access.* **10** : 19083-19095. 10.1109/ACCESS.2022.3151048.

[29] D. Krstinić, M. Braović, L. Šerić, and D. Božić-Štulić. (2020). "Multi-Label Classifier Performance Evaluation with Confusion Matrix". *Computer Science & Information Technology.* **1** (1): 1-14. 10.5121/csit.2020.100801.

[30] S. C. Kumar and R. J. RamaSree. (2015). "Dimensionality Reduction in Automated

Evaluation of Descriptive Answers Through Zero Variance, Near Zero Variance, and Non -Frequent Words Techniques: A Comparison". *Proceedings of the 9th IEEE International Conference on Intelligent Systems and Control (ISCO)*. 1-6. 10.1109/ ISCO.2015.7282351.