Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15312
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Explainable Machine Learning for Poverty Prediction in Central Java Regencies and Cities

Wahyu Fhaldian [1], Amiq Fahmi[2] *
[1,2] Fakultas Ilmu Komputer, Universitas Dian Nuswantoro, Semarang, Indonesia
Jl. Imam Bonjol No.207, Kota Semarang, Jawa Tengah 50131
[1] 111202214572@mhs.dinus.ac.id, [2]* amiq.fahmi@dsn.dinus.ac.id

Submitted :Sep 4, 2025 | **Accepted** : Oct 2, 2025 | **Published** : Oct 4, 2025

Abstract: Poverty remains a multidimensional challenge in Central Java, necessitating robust data-driven approaches to identify its socioeconomic determinants across diverse administrative contexts. This study applied six machine learning models—Extreme Gradient Boosting (XGBoost), Random Forest, CatBoost, LightGBM, Elastic Net Regression, and a Stacking ensemble—using regency- and city-level data from Statistics Indonesia. The dataset encompasses indicators related to demographics, education, labor, infrastructure, and household welfare across 35 administrative units (29 regencies and 6 cities). Model evaluation employed an 80:20 hold-out split, 10-fold cross-validation, and noise perturbation tests to assess generalizability and robustness. XGBoost achieved the best individual performance (MAE = 2,180.01; RMSE = 3,512.07; R^2 = 0.931), while the Stacking ensemble outperformed all single learners (MAE = 2,640.99; RMSE = 3,202.79; R^2 = 0.942). Interpretability was ensured through SHAP (Shapley Additive Explanations), Partial Dependence Plots (PDP), and Accumulated Local Effects (ALE), consistently identifying Number of Households, Per Capita Expenditure, and Uninhabitable Houses as the most influential predictors. Counterfactual simulations indicated that increasing per capita expenditure by 10% could reduce the poverty index by 9.9%, while reducing household size by 10% lowered it by 11.3%. Robustness checks revealed Brebes Regency as an influential unit shaping model sensitivity. Overall, the findings demonstrate that boosting and stacking ensembles, when combined with explainable AI tools, not only enhance predictive accuracy but also provide transparent, policy-relevant evidence to strengthen poverty alleviation programs in Central Java. This study contributes both methodological advances in explainable machine learning and practical insights for targeted poverty reduction strategies.

Keywords: Poverty; XGBoost; Random Forest; CatBoost; LightGBM; Elastic Net Regression; Stacking; SHAP; PDP; ALE; Central Java

INTRODUCTION

Poverty is among the most pressing and complex issues in regional development, notably in Central Java Province. According to the Indonesian Central Bureau of Statistics (Badan Pusat Statistik BPS, 2024), roughly 3.70 million people in Central Java live below the poverty line, making the province the third-highest in the country behind East and West Java. The area poverty rate is 10.47%, higher than the national average of 9.36%. These findings highlight the crucial importance of inclusive, data-driven, and context-sensitive development initiatives that are both focused and sustainable.

Understanding poverty involves more than just an economic perspective; it necessitates a multifaceted study that includes access to education, healthcare, a respectable job, and suitable living conditions. As a result, spatial and multidimensional techniques have gained traction in efforts to detect and map poverty holistically (Chi et al., 2022). Chi and colleagues, for example, assessed wealth distribution in developing nations by combining satellite imagery and topographic data using spatial micro-estimation approaches.

Recent advancements in analytics and processing capability have elevated machine learning to a potent tool for detecting complex patterns and predicting poverty that traditional statistical methods cannot. For example, Li

*Amiq Fahmi

et al. (2022)found that XGBoost outperformed the Generalized Linear Model (GLM) in identifying poor households in Kyrgyzstan. Researchers in Latin America and Southeast Asia, including Muñetón-Santa & Muñetón-Santa & Manrique-Ruiz (2023) in Colombia and Salvador (2024) in the Philippines, have found that Random Forest and XGBoost effectively map multidimensional poverty. Recent studies have also shown that CatBoost, a gradient boosting algorithm optimized for categorical data, can deliver predictive performance comparable to XGBoost and Random Forest in poverty mapping applications (Muñetón-Santa & Manrique-Ruiz, 2023a).

In Indonesia, the use of machine learning for poverty studies has begun to garner scholarly interest, while its deployment is still limited. Izzati et al. (2024) used the Random Forest approach to categorize impoverished families in Central Java, attaining high predictive accuracy while overcoming class imbalance concerns. In a different study, Putra et al. (2024) used satellite imagery and Point of Interest (POI) data to estimate poverty levels in the same region, highlighting the potential of geospatial data to improve poverty mapping efforts.

Machine learning models must be both interpretable and predictively accurate, especially for policymakers. Explainable AI (XAI) techniques, such as SHAP (SHapley Additive Explanations), can link model results to policy decisions. Previous research, particularly Wang et al. (2024) and Hall et al. (2022), has emphasized the critical role of model interpretability in facilitating practical adoption in policy contexts. More recently, model-agnostic interpretability methods such as Partial Dependence Plots (PDP) and Accumulated Local Effects (ALE) have been employed to complement SHAP by illustrating marginal effects and feature interactions, thus enhancing transparency in ensemble learning models (Christoph Molnar, 2025).

This study attempts to fill gaps in the literature by developing a poverty prediction model for Central Java using XGBoost, Random Forest, CatBoost, LightGBM, Elastic Net Regression, and a Stacking ensemble. To strengthen interpretability, SHAP was applied to tree-based learners (particularly XGBoost). In contrast, PDP and ALE were applied to the stacking ensemble, whose meta-learner (Elastic Net Regression) does not natively support SHAP analysis. In addition to achieving high predictive accuracy, this combined strategy aims to achieve high predictive accuracy and deliver transparent, evidence-based insights that can guide policymaking at the regency and city levels.

LITERATURE REVIEW

Poverty and Spatial Approaches

Accurate poverty mapping is essential for carrying out successful spatially focused programs. Chi et al. (2022) argue that understanding the distribution of prosperity and poverty in low- and middle-income nations requires a micro-scale spatial approximation. By combining Facebook connectivity metrics, topographical data, mobile network data, and satellite photos, their study produced comprehensive wealth estimates at the 2.4 km level across 135 countries. This strategy demonstrates how geographical data can highlight regional disparities and assist decision-makers in deciding which areas require priority attention.

Poverty Modeling with Machine Learning

Several studies show that machine learning outperforms traditional methods for predicting household poverty. Li et al. (2022) found that XGBoost outperformed GLM on Kyrgyz DHS data, whereas Solís-Salazar & Madrigal-Sanabria (2022)observed reduced inclusion and exclusion errors in Costa Rica using XGBoost. In Indonesia, Izzati et al. (2024) used Random Forest and CHAID to analyze National Socioeconomic Survey (SUSENAS) 2020 data, reaching 93.95% accuracy (AUC = 0.9417) with SMOTE adjustment. Similarly, Ramayanti et al. (2023) identified low-income families in West Sumatra with 94.35% accuracy, identifying significant predictors such as lighting, cooking fuel, water availability, and occupation of the head of the household. Valentika et al. (2024) emphasized interval precision with Random Forest at the regency and city level, while Putra et al. (2024) showed that SVR outperformed Random Forest with satellite and POI data. Boosting algorithms have also demonstrated high performance. Salvador (2024) showed 89% accuracy with XGBoost in the Philippines, whereas Zamzuri et al. (2023) found poverty patterns using XGBoost and Random Forest, and Lastras Rodríguez (2024) mapped social inequalities in Madrid with XGBoost. Muñetón-Santa & Manrique-Ruiz (2023) used CatBoost, XGBoost, LightGBM, and Random Forest to map poverty in Medellin, Colombia, with equivalent results (CatBoost MAE = 0.078).

Random Forest and Ensemble Approaches in Poverty Prediction

Random Forest and other ensemble approaches are commonly used in poverty prediction because of their robustness in simulating complex nonlinear patterns. Muñetón-Santa & Manrique-Ruiz (2023) found that Random Forest and XGBoost effectively identified multidimensional poverty in Colombia based on education, sanitation, and asset ownership. Valentika et al. (2024) Random Forest's capacity to create point and interval forecasts for regency and city-level poverty in Indonesia. Similarly, Putra et al. (2024) used Random Forest and SVR with

*Amiq Fahmi

e-ISSN : 2541-2019
p-ISSN : 2541-044X

satellite and POI data to capture regional and multidimensional elements of poverty. Beyond single learners, ensemble techniques such as stacking have also been applied. Kuswanto et al. (2025) developed a stacking ensemble (KNN, Random Forest, XGBoost, SVM) for regency and city -level expenditure estimation, demonstrating a moderate R^2 improvement compared to individual models, thereby highlighting the potential of combining heterogeneous learners for complex socioeconomic datasets.

Regularized Linear Models for Poverty Prediction

While ensemble methods like XGBoost, Random Forest, and CatBoost frequently produce higher predictive accuracy, regularized linear models like Elastic Net have also shown promise in socioeconomic prediction tasks. Elastic Net integrates L1 (lasso) and L2 (ridge) penalties, allowing for feature selection and coefficient shrinking in the presence of multicollinearity. Binka et al. (2022) used Elastic Net penalized logistic regression to detect long COVID patients using population-level health data, with a sensitivity of 86%, specificity of 86%, and AUC of 93%. It highlights Elastic Net's ability to identify appropriate predictors while maintaining model stability. When it comes to poverty mapping, Elastic Net can successfully capture socioeconomic indicators, including household size, income, education, and housing conditions. Although less precise than ensemble approaches, it provides interpretable coefficients that define the direction and size of feature effects, giving it an important linear baseline for policy research.

Explainable AI and SHAP for Model Interpretation

Wang et al. (2024) used SHAP in disaster risk analysis to determine how specific attributes influenced forecasts. Hall et al. (2022) stressed the need for interpretability in building confidence with policymakers, especially when using high-dimensional data like satellite photography. Beyond technological benefits, SHAP identifies the most relevant socioeconomic elements, allowing for evidence-based legislation. It also gives information about poverty determinants at the global model level and specific regency and city or households.

Relevance to the Context of Indonesia and Central Java

Although the majority of machine learning-based poverty research has been conducted outside of Indonesia, its approaches are still very relevant to locations like Central Java, where job, infrastructure, and education difficulties continue. Ensemble models, like as Random Forest and XGBoost, which have been shown to be useful in global research, are ideal for capturing the complexities of poverty data. Recent studies (Muñetón-Santa & Manrique-Ruiz, 2023; Putra et al., 2024) show that CatBoost, LightGBM, and SVR are effective in predicting poverty across diverse socioeconomic datasets. Explainable AI technologies, such as SHAP, are especially useful in Indonesia, where policymakers need transparent and evidence-based insights. The 35 regencies and the city of Central Java, each with its own economic and human development conditions, highlight the necessity for models that account for spatial heterogeneity and variable interactions. Using these methodologies with provincial data can produce actionable evidence to help resource allocation and regency—and city-specific decisions.

Related Studies

Several previous studies have methodological parallels to this study but differ in topic or location. Izzati et al. (2024) used Random Forest to analyze Central Java poverty data without using interpretability tools such as SHAP. Ramayanti et al. (2023) and Valentika et al. (2024) used Random Forest to classify poverty and predict intervals in West Sumatra and the national context. Kuswanto et al. (2025)developed a stacking ensemble (KNN, Random Forest, XGBoost, SVM) for regency/city-level expenditure estimation, demonstrating a moderate R^2 improvement and emphasizing the need to mix methods for spatial data. Huang et al. (2023) advanced this field by combining XGBoost and SHAP to identify poverty drivers in China with an accuracy of 81.9%. Their usage of SHAP and PDP provided more profound insights into feature effects, demonstrating the value of interpretable ML. Building on previous research, this study combines XGBoost, Random Forest, CatBoost, and stacking with SHAP for Central Java, intending to connect technical modeling to practical, regency/city-level policy creation.

This study followed a planned series of steps to examine the socioeconomic factors influencing poverty in Central Java thoroughly. The procedure involved data collection, data pre-processing, modeling, model evaluation, and model interpretation. Each stage was created to maintain data integrity, increase predictive accuracy, and produce policy-relevant conclusions.

*Amiq Fahmi

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15312
e-ISSN : 2541-2019
p-ISSN : 2541-044X

METHOD

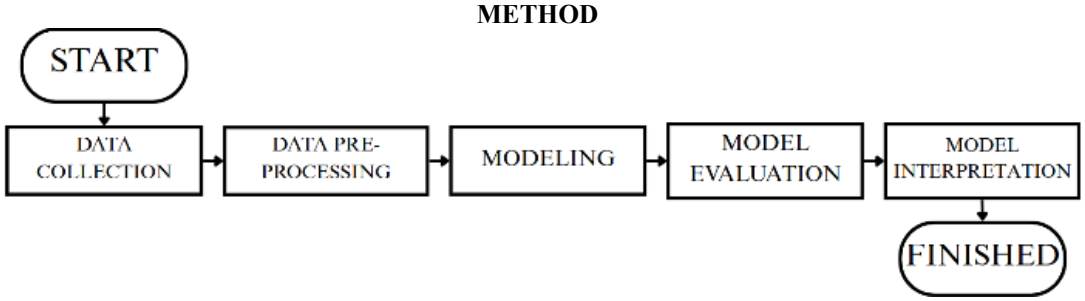

Fig. 1 Research Sequence

Data Collection

The study used socioeconomic data from the Statistics (BPS) for all 35 regencies/cities and municipalities in Central Java Province. The target variable, the Poverty Index, was calculated as the mean of three official poverty indicators: P1 (extreme poverty gap), P2 (high poverty gap), and P3 (low poverty gap). Several explanatory variables were included to capture multidimensional poverty conditions, covering demographic, economic, education, health, and infrastructure dimensions. Table 1 has a thorough explanation of the dataset.

To provide further context, descriptive statistics were computed for the main variables to illustrate their distribution and variability across regencies/cities. These indicators reflect the socioeconomic diversity of Central Java and provide a basis for subsequent modeling. The summary is reported in Table 2.

The combination of Tables 1 and 2 demonstrates the dataset's multidimensional structure, highlighting key socioeconomic conditions that underlie poverty disparities across Central Java. These descriptive results provide the foundation for the subsequent preprocessing, modeling, and interpretation stages.

Table 1. Dataset Description

Variable	Description	Type
Poverty Index (avg P1, P2, P3)	Composite indicator averaging P1 (extreme), P2 (high), and P3 (low) poverty gap indices	Numeric
Land Area (km²)	Total land area per regency/city in square kilometers	Numeric
Population	Total population per regency/city	Numeric
Number of Households	Number of households per regency/city	Numeric
Per Capita Expenditure	Average expenditure per capita in Indonesian Rupiah (IDR)	Numeric
Unemployment Rate (%)	Percentage of unemployed population	Numeric
Uninhabitable Houses	Number of houses categorized as unfit for habitation	Numeric
Minimum Wage (IDR)	Regency/city minimum wage	Numeric
Educational Attainment	Distribution of population by education levels (elementary school, junior high school, senior high school, college)	Numeric
Employment Characteristics	Proportion of working vs. unemployed residents	Numeric
Healthcare Facilities	Number of hospitals and public health centers (Community Health Centers)	Numeric
Life Expectancy (years)	Average life expectancy at birth	Numeric
Economic Growth Rate (%)	Annual growth rate of the regency/city economy	Numeric
Human Development Index	Composite index of health, education, and income	Numeric
Regency/City	Identifier of each regency/city in Central Java (not used in modeling)	Categorical

Table 2. Selected Variables and Descriptive Statistics

Variable	Mean	Standard Deviation	Min	Max
Poverty Index (avg P1, P2, P3)	17,327.56	13,701.72	1,430	57,264.67
Land Area (km²)	964.29	575.40	18.56	2,249.28
Population	1,074,524	471,332	122,150	2,043,077
Number of Households	138,875	79,230	8,990	349,073
Per Capita Expenditure	1,247,607	251,057	936,059	2,098,293
Unemployment Rate (%)	4.92	1.69	1.92	8.98

Data Preprocessing

*Amiq Fahmi

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15312
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Several preprocessing techniques were performed before creating the prediction models to improve data quality and confirm the dataset's appropriateness for machine learning. Non-numerical identifiers, such as regency/city, were omitted because they do not directly contribute to the prediction process. Furthermore, the three poverty indices (P1, P2, and P3) were combined into a single composite Poverty Index to serve as the prediction objective, reducing redundancy in the model. Missing values in various variables were handled using median imputation, a method chosen for its robustness against outliers and ability to preserve the data's underlying distribution. To ensure comparability between predictors, all numerical variables were normalized using the Standard Scaler method, translating the features into a mean of zero and a standard deviation of one. This step is crucial because, although relatively robust, tree-based algorithms can still benefit from improved feature scaling when dealing with wide-ranging socioeconomic indicators.

The dataset was then separated into training and testing subsets at an 80:20 ratio, resulting in 28 regencies/cities for training and 7 regencies/cities for testing. A fixed random state (42) was used to ensure reproducibility across runs. This technique is consistent with previous studies in socioeconomic prediction (Izzati et al., 2024; Ramayanti et al., 2023). Finally, Pearson's correlation coefficients were calculated between predictors to conduct an exploratory correlation analysis. The resulting structure was visualized as a heatmap, facilitating the detection of potential multicollinearity and providing preliminary insights into variable interactions. Including this step in preprocessing ensures transparency and creates a stronger link between the descriptive exploration and the subsequent modeling process.

Modeling

This study employed multiple ensemble-based machine learning methods, Extreme Gradient Boosting (XGBoost), Random Forest (RF), CatBoost, LightGBM, alongside Elastic Net Regression as a baseline, and a Stacking ensemble as the final model. These algorithms are widely recognized for their ability to capture complex, non-linear patterns in socioeconomic datasets while offering complementary strengths in terms of robustness, interpretability, and computational efficiency (Li et al., 2022; Muñetón-Santa & Manrique-Ruiz, 2023; Zhao et al., 2019). XGBoost applies a sequential boosting approach, where each decision tree corrects the residual errors of its predecessors. Equation (1) represents the formulation of the XGBoost.

$$\hat{y}i = \sum_{k=1}^{K} fk(x_i), \quad f_k \in \mathcal{F} \qquad (1)$$

Here, f_k denotes the k-th decision tree within the ensemble, and the final prediction $\hat{y}i$ is obtained as the cumulative contribution of all trees. A key strength of XGBoost lies in its integrated regularization terms, which help reduce overfitting and enhance generalization. Prior studies have highlighted these advantages: (Li et al., 2022) found that XGBoost beat the Generalized Linear Model (GLM) for predicting household poverty in Kyrgyzstan, while (Salvador, 2024) indicated that the algorithm obtained approximately 89% accuracy when used to multidimensional poverty in the Philippines. For this study, the hyperparameters were adjusted to: n estimators = 100, learning rate = 0.1, max depth = 6, subsample = 1, colsample bytree = 1, reg lambda = 1, reg alpha = 0, and random state = 42. Random Forest, by contrast, is based on bootstrap aggregation (bagging), generating multiple decision trees and averaging their results (2).

$$\hat{y}i = \frac{1}{T}\sum_{t=1}^{T} h_t(x) \qquad (2)$$

In this formulation, $\hat{y}i$ represents the predicted value, and $h_t(x)$corresponds to the prediction from the t-th tree. Random Forest is particularly effective in handling noisy data and high-dimensional inputs. Previous research has applied Random Forest to classify impoverished households and even generate prediction intervals for poverty analysis in Indonesia (Ramayanti et al., 2023; Valentika et al., 2024). In this study, the Random Forest hyperparameters were optimized using grid search, yielding the configuration: n estimators = 100, max depth = None, min samples split = 2, min samples leaf = 1, and random state = 42.

CatBoost was additionally implemented to strengthen the comparative analysis. CatBoost is a gradient boosting technique that was specifically intended to handle categorical variables effectively. It uses ordered boosting and permutation-driven categorical encoding to reduce prediction shift and overfitting (Dorogush et al., 2018). The categorical feature transformation in CatBoost can be formalized as follows (2).

$$X_{\sigma_p}, k = \frac{\sum_{j=1}^{p-1}\left[X_{\sigma_j}, k = X_{\sigma_p}, k\right] Y_{\sigma_j} + a \cdot P}{\sum_{j=1}^{p-1}\left[X_{\sigma_j}, k = X_{\sigma_p}, k\right] + a} \qquad (3)$$

Where X_{σ_p}, k denotes the encoded value of the k-th categorical feature for the p-th sample, $\left[X_{\sigma_j}, k = X_{\sigma_p}, k\right]$ is the Iverson bracket (equal to 1 if the categories match, 0 otherwise), Y_{σ_j} is the target value for sample j, a is a

*Amiq Fahmi

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15312
e-ISSN : 2541-2019
p-ISSN : 2541-044X

prior weight, and P is the prior distribution (often the global mean). This formulation ensures that each category is represented using prior observations in a random permutation with smoothing, thereby mitigating target leakage. For this study, CatBoost was configured with: iterations = 500, learning rate = 0.03, depth = 8, silent= True, random state = 42. (Muñetón-Santa & Manrique-Ruiz, 2023) also confirmed the competitiveness of CatBoost, reporting performance comparable to RF and XGBoost in multidimensional poverty prediction.

LightGBM, like XGBoost, is a gradient-boosting technique that improves computing efficiency and scalability by combining histogram-based feature binning and a leaf-wise tree growth strategy (Ke et al., 2017). LightGBM's prediction function has the same additive-tree shape as Equation (1). However, it is faster thanks to histogram-based techniques and feature bundling, which are especially effective when dealing with a large number of correlated predictors. For this study, LightGBM hyperparameters were set as: n estimators = 500, learning rate = 0.05, max depth = –1 (no limit), min data in leaf = 1, min_data in bin = 1, feature fraction = 0.8, bagging fraction = 0.8, and random state = 42.

Elastic Net Regression serves as the regularized linear benchmark, combining both L1 (Lasso) and L2 (Ridge) penalties:

$$\hat{\beta} = \frac{\text{argmin}}{\beta}\{\sum_{i=1}^{n}(y_i - x_i\,\beta)^2 + \lambda[(1-\alpha)\|\beta\|_2^2 + \alpha\|\beta\|_1]\} \qquad (4)$$

Here, λ controls the overall regularization strength, while α∈ [0,1] balances between Ridge (α=0) and Lasso (α=1) penalties (Zou & Hastie, 2005). This hybrid penalty allows Elastic Net to handle correlated predictors more effectively while still performing feature selection. For this study, the hyperparameters were tuned using grid search with cross-validation, resulting in: *alpha = 0.5, l1_ratio = 0.5, random_state = 42*.

Finally, a Stacking ensemble integrated the four base learners (XGBoost, RF, CatBoost, LightGBM), with Elastic Net Regression as the meta-learner. The general stacking formulation is:

$$\hat{y} = g\left(f_1(x), \widehat{f_2(x)}, \ldots, f_M(x)\right) \qquad (5)$$

where $f_M(x)$ are predictions from the m-th base learner, and g (·) is the meta-model (Elastic Net) that learns the optimal weighted combination (Sagi & Rokach, 2018; Wolpert, 1992). By integrating diverse learners, stacking captures complementary model strengths, improving accuracy and robustness.

This study focused on interpretability as well as predicted accuracy. Tree-based learners were trained using SHapley Additive Explanations (SHAP) for global and local feature attributions. In contrast, the stacking ensemble was trained with Partial Dependence Plots (PDP) and Accumulated Local Effects (ALE) to capture marginal and interaction effects. This combination provides a clear interpretability framework, linking machine learning results to practical policy insights (Lee et al., 2024; Zheng et al., 2024).

Model Evaluation

The model evaluation measured the accuracy of poverty projections across dreegeencie/cities. MAE measures average absolute errors, RMSE highlights bigger deviations through squared residuals, and R^2 quantifies explained variance, with values near 1 suggesting better performance. Corral Rodas et al. (2023) emphasized the necessity of combining statistical and visual assessments; hence, real and anticipated poverty indices were compared using scatter plots, with closeness to the diagonal line indicating greater prediction accuracy.

To guarantee consistency, the dataset was separated between training and testing subsets (80:20) with a fixed random seed of 42. This prevented models from overfitting by memorizing training data (Corral Rodas et al., 2023). 10-fold cross-validation increased robustness by using each fold once as a validation set and averaging results across folds. This minimized partitioning bias and produced more reliable generalization estimates, in line with Zamzuri et al. (2023).

Additional robustness checks included a noise perturbation study, in which Gaussian noise was injected into features at 5%, 10%, and 20% levels before retraining. This replicated the usual measurement mistakes found in socioeconomic surveys. The results showed that boosting and stacking approaches were relatively resilient; however, Elastic Net regression was more sensitive to noise, which was consistent with Satapathy et al. (2023), who stressed the importance of noise injection in poverty modeling.

Finally, a feature stability analysis looked at the consistency of important predictors across models. SHAP values determined feature priority for tree-based learners, whereas coefficient magnitudes were used to assess Elastic Net and the stacking meta-learner. Household size, per capita expenditure, and housing quality were consistently identified as important drivers across algorithms and folds, improving both interpretability and policy relevance (Christoph Molnar, 2025; Zou & Hastie, 2005).

Several evaluation criteria, as well as visual examination, helped guarantee that the models were statistically sound and useful as decision-support aids for policy development. The Mean Absolute Error (MAE) is calculated as (5).

*Amiq Fahmi

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15312
e-ISSN : 2541-2019
p-ISSN : 2541-044X

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|yi - \hat{y}i| \quad (5)$$

The Mean Absolute Error (MAE) is calculated as the average of the absolute differences between the actual value (yi) and the predicted value (ŷi), and a model demonstrates superior performance when its resulting MAE value is lower.

Root Mean Squared Error (RMSE) is calculated as (6).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(yi - \hat{y}i)^2} \quad (6)$$

Calculating the coefficient of determination (R2) is done as follows (7).

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(yi-\hat{y}i)^2}{\sum_{i=1}^{n}(yi-\bar{y}i)^2} \quad (7)$$

Where $\bar{y}i$ is mean of actual values.

Model Interpretation

Following the assessment of predictive performance, the models were further analyzed to identify the socioeconomic factors most strongly influencing poverty levels across regencies/cities. This study employed Explainable AI (XAI) methods, including SHAP (SHapley Additive Explanations), Partial Dependence Plots (PDP), and Accumulated Local Effects (ALE).

SHAP, grounded in cooperative game theory, attributes each feature a contribution value that represents its marginal effect on the model's output. The prediction for an instance can be expressed as (8).

$$f(x) = \phi_0 + \sum_{i=1}^{M}\phi_i \quad (8)$$

where f(x) is the model prediction for a given observation, ϕ0\phi_0ϕ0 is the baseline or average prediction across the dataset, and ϕi \phi_iϕi represents the marginal contribution of each feature iii to the prediction. SHAP allows for the detection of nonlinear associations between features and outputs, offering insights that linear models may overlook (Lastras Rodríguez, 2024; Lee et al., 2024; Lundberg & Lee, 2017).

To complement SHAP, PDP was applied to examine the marginal effect of a feature on the predicted poverty index by averaging over the distribution of all other features:

$$PD_j(x_j) = \frac{1}{n}\sum_{i=1}^{n} f(x_j x_{i,-j}) \quad (9)$$

where $PD_j(x_j)$ is the partial dependence of feature j, $f(x_j x_{i,-j})$ is the model prediction at fixed xjx_jxj and observed values of other features, and nnn is the number of samples. PDP provides a global interpretation of feature influence, though it can be biased when features are correlated (Greenwell et al., 2018).

To address this limitation, ALE plots were also employed. ALE measures the accumulated local effect of a feature on the prediction by averaging changes in predictions over small intervals of the feature distribution:

$$ALE_j(x) = \int_{x_{min}}^{x} E\left[\frac{\partial f(x)}{\partial x_j} x_j = z\right] dz \quad (10)$$

where $ALE_j(x)$ denotes the accumulated local effect of feature j, and the expectation is taken conditionally within local intervals of the feature's distribution. ALE provides more robust interpretability in the presence of correlated predictors, avoiding extrapolation bias (Apley & Zhu, 2020).

By integrating SHAP, PDP, and ALE, the analysis ensured a comprehensive interpretive framework: SHAP captured features' global and local contributions, PDP illustrated marginal feature–response relationships, and ALE provided robust validation against correlation bias. Together, these methods enhanced model transparency and reinforced the policy relevance of the identified determinants of poverty in Central Java.

RESULT

This study utilized a dataset comprising 35 regencies/cities and cities in Central Java Province, incorporating socioeconomic factors expected to correlate with poverty levels. The target variable, Poverty Index, was calculated as the average of three sub-indicators: P1 (Poverty Gap Index – extreme), P2 (high), and P3 (low). This composite measure offers a more comprehensive representation of poverty severity. Descriptive statistics reveal substantial regional variation, with Poverty Index values ranging from 1,430 to 57,264.7, a mean of approximately 17,327.6, and a standard deviation of 13,701.7. The median (13,186.3) and third quartile (25,695.5) indicate that, with few exceptions, most regencies/cities fall below extreme poverty levels. Table 3 identifies the five regencies/cities with

*Amiq Fahmi

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15312
e-ISSN : 2541-2019
p-ISSN : 2541-044X

the highest Poverty Index scores, which are typically associated with larger populations, infrastructure deficits, and lower average education levels.

According to Table 3, several regencies or cities with the highest poverty index values are predominantly located in areas characterized by challenging geographical conditions, inadequate infrastructure, and lower levels of educational attainment.

Fig. 2 presents a correlation heatmap constructed to investigate the inter-variable relationships among the socioeconomic indicators. Several notable correlation patterns emerge from the analysis. The number of households (3) exhibits a strong positive correlation with the number of household members (4) (r = 0.92), reflecting the demographic composition of each regency/city. Per capita expenditure (21) shows a substantial positive correlation with the Human Development Index (19) (r = 0.74), suggesting that higher consumption levels are associated with improved quality of life. Moreover, the unemployment rate (18) is negatively correlated with higher education and employment (11) (r ≈ -0.66), indicating that increased access to higher education contributes to lower unemployment. Furthermore, the poverty index (22) demonstrates a negative correlation with per capita expenditure (21) (r = -0.62), highlighting the critical role of household consumption in poverty alleviation.

Table 3. Reegncis/cities with the Highest Poverty Index

Regency	Poverty Index
Cilacap	57264.7
Purbalingga	43534.7
Grobogan	35507
Magelang	34992
Pemalang	32605

These data provide essential context for the modeling process, revealing key socioeconomic relationships that influence poverty outcomes.

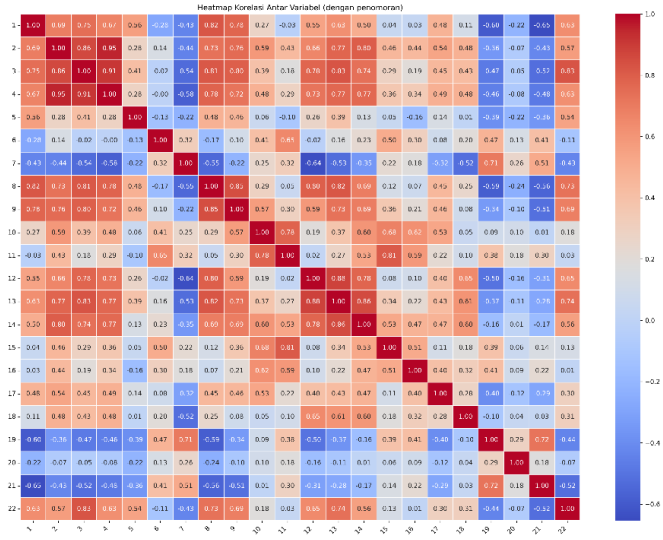

Fig. 2 Correlation heatmap of variables

Each variable is labeled with a number to facilitate visualization and interpretation of relationships in the heatmap. The numbering is as follows: 1 = Land Area, 2 = Population, 3 = Number of Households, 4 = Number of Household Members, 5 = Uninhabitable Houses, 6 = Regency/City Minimum Wage, 7 = Life Expectancy, 8 = Elementary Education & Employed, 9 = Junior High Education & Employed, 10 = Senior High Education & Employed, 11 = Higher Education & Employed, 12 = Elementary Education & Unemployed, 13 = Junior High Education & Unemployed, 14 = Senior High Education & Unemployed, 15 = Higher Education & Unemployed, 16 = Number of Hospitals, 17 = Community Health Centers (Puskesmas), 18 = Unemployment Rate, 19 = Human Development Index (HDI), 20 = Economic Growth Rate, 21 = Per Capita Expenditure, and 22 = Poverty Index.

Beyond this labeling, the correlation heatmap provides insights into how these socioeconomic factors interact. For instance, the number of households (3) and the number of household members (4) are strongly positively correlated (r ≈ 0.91), reflecting household and demographic structure. Per capita expenditure (21) and the Human Development Index (19) also show a strong positive correlation (r ≈ 0.72), indicating that higher consumption levels are associated with improved welfare outcomes. In contrast, the unemployment rate (18) and Higher

*Amiq Fahmi

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15312
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Education & Employed (11) display only a weak positive correlation (r ≈ 0.10), suggesting a limited linear association between these variables in this dataset. Moreover, the poverty index (22) exhibits a moderate negative correlation with per capita expenditure (21) (r ≈ –0.52), underscoring the role of household consumption in alleviating poverty. These patterns are important for interpreting the modeling results, as strong or weak correlations between predictors may influence variable importance and reflect the multidimensional drivers of poverty. Prior to the modeling phase, a heatmap visualization was generated to determine the correlation links between socioeconomic variables. This visualization aids in understanding the patterns of association between variables, including potential positive or negative correlations that could influence prediction outcomes. Furthermore, scatter plots of anticipated versus absolute poverty index values were used to illustrate model performance visually.

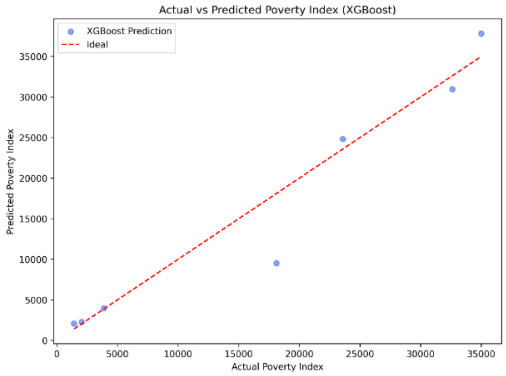

Fig. 3 Plot of Predicted vs Actual Values by XGBoost

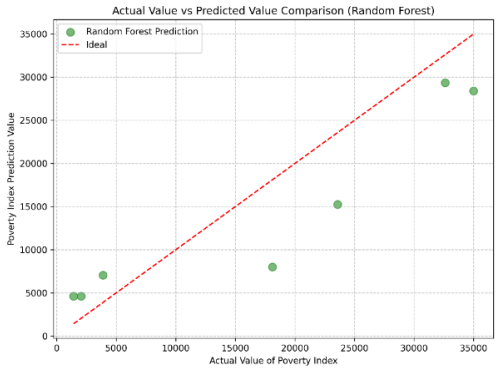

Fig. 4 Plot of Predicted vs Actual Values by Random Forest

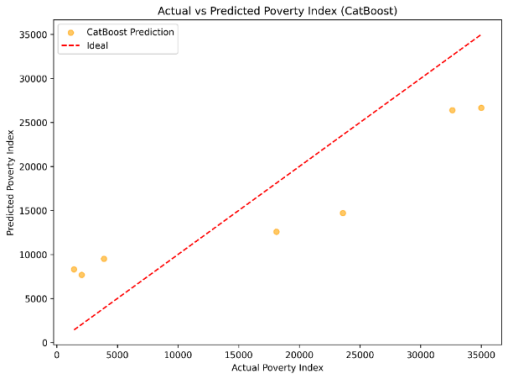

Fig. 5 Plot of Predicted vs Actual Values by CatBoost

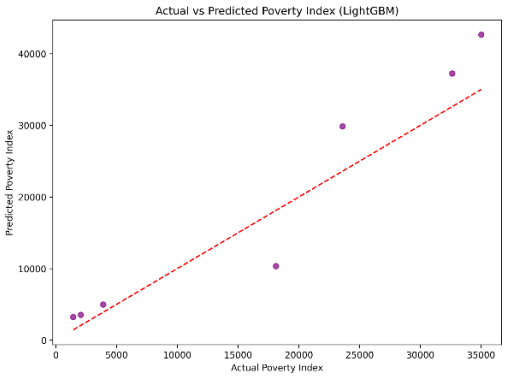

Fig. 6 Plot of Predicted vs Actual Values by LightGBM

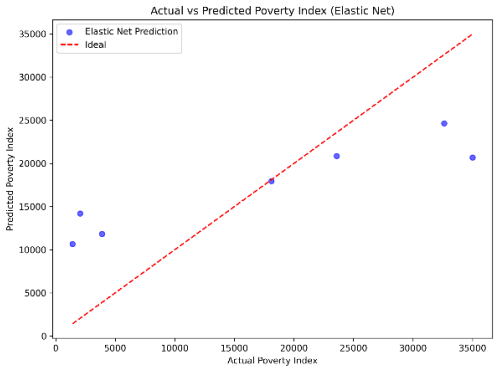

Fig. 7 Plot of Predicted vs Actual Values by Elastic Net

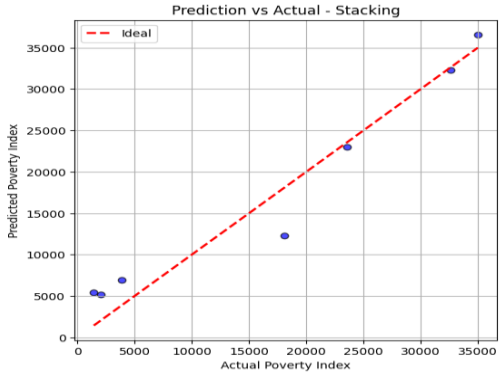

Fig. 8 Plot of Predicted vs Actual Values by Stackin

*Amiq Fahmi

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15312
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Figs 3 to 8 show scatter plots comparing actual and anticipated poverty index values for each model: XGBoost (Fig. 3), Random Forest (Fig. 4), CatBoost (Fig. 5), LightGBM (Fig. 6), Elastic Net (Fig. 7), and the Stacking ensemble (Fig. 8). Points closer to the diagonal line (y = x) have higher prediction accuracy, but greater deviations indicate more errors. Visual inspection reveals that XGBoost and Stacking are the most similar to the ideal line, as evidenced by their decreased MAE and RMSE values. LightGBM also performs competitively, albeit slightly below these versions. Random Forest and CatBoost, on the other hand, exhibit greater dispersion, while Elastic Net has the poorest fit, which is consistent with its larger error values. These findings highlight the greater generalization of boosting-based models, specifically XGBoost, in predicting multidimensional poverty indices across Central Java regions. As Corral Rodas et al. (2023) pointed out, using assessment measures such as MAE and RMSE is crucial since moderate error levels increase the robustness of data-driven strategies. However, insufficient validation risks biased outcomes and diminishes policy relevance.

Table 4. Model Evaluation Results

Model	MAE	RMSE	R^2	Top 3 most important Features / Base Learners
XGBoost	2.180,01	3.512,07	0,931	Number of Households, Uninhabitable Houses, Per Capita Expenditure
Random Forest	5.310,83	5.999,67	0,800	Number of Households, Uninhabitable Houses, Elementary School Graduates Employed
CatBoost	6,722.99	6,841.92	0.738	Number of Households, Uninhabitable Houses, Per Capita Expenditure
LightGBM	4,386.29	5,163.74	0.851	Number of Households, Uninhabitable Houses, Per Capita Expenditure
Elastic Net	7.777,85	9.031,75	0.543	Number of Households, Junior High Education & Unemployed, Population
Stacking	2,640.99	3,202.79	0.942	Random Forest, XGBoost, Catboost (base learners with strongest weights)

Table 4 compares the performance of six different models: XGBoost, Random Forest, CatBoost, LightGBM, Elastic Net, and Stacking Ensemble. XGBoost outperformed bagging approaches in multidimensional poverty modeling (MAE = 2,180.01; RMSE = 3,512.07; R^2 = 0.931), supporting Zheng et al.'s (2024) findings. This advantage comes from sequential boosting and embedded regularization, which reduce overfitting.

Random Forest yielded moderate results (MAE = 5,310.83; RMSE = 5,999.67; R^2 = 0.800), while CatBoost had lower predictive accuracy (MAE = 6,722.99; RMSE = 6,841.92; R^2 = 0.738), but successfully identified socioeconomic drivers like uninhabitable housing and household size (Muñetón-Santa & Manrique-Ruiz, 2023). LightGBM outperformed Random Forest and CatBoost, but fell short of XGBoost.

Linear regression underperformed (MAE ≈ 12,755; RMSE ≈ 13,260; R^2 ≈ 0.014), highlighting its failure to represent multidimensional poverty. Elastic Net was thus chosen as a more advanced linear baseline, with L1 (Lasso) and L2 (Ridge) penalties for feature selection and multicollinearity handling (Binka et al., 2022b). Elastic Net achieved MAE = 7,777.85, RMSE = 9,031.75, and R^2 = 0.543, indicating poorer predictive power than ensemble approaches but providing interpretable coefficients to supplement complex models.

The Stacking ensemble achieved the highest overall performance (MAE = 2,640.99; RMSE = 3,202.79; R^2 = 0.942), demonstrating the advantages of combining several learners to capture diverse data patterns. Unlike single models, its interpretability is based on meta-learner coefficients across base learners, with Random Forest, XGBoost, and CatBoost being the most influential contributors. At the same time, LightGBM had weaker and less stable weights. Feature stability was investigated further using SHAP values (tree-based learners), coefficient magnitudes (Elastic Net), and weight consistency (Stacking), as shown in Tables 8 (top three features per model) and 9 (stability of stacking contributions).

Figs. 3 to 8 visually highlight these quantitative results. XGBoost (Fig. 3) and Stacking (Fig. 8) predictions are most closely aligned with the diagonal line, indicating higher accuracy. LightGBM (Fig. 6) similarly shows competitive alignment. In contrast, Random Forest (Fig. 4) and CatBoost (Fig. 5) show larger dispersion, and Elastic Net (Fig. 7) has the lowest fit, which is consistent with its error metrics. These findings confirm that boosting-based and ensemble models provide greater generalization in modeling multidimensional poverty indices. This echoes Zheng et al. (2024), who found that boosting and stacking frameworks outperformed bagging-based techniques such as Random Forest.

To augment the hold-out evaluation, Table 5 shows the results of a 10-fold cross-validation test to check model resilience. Cross-validation had significantly lower results than the single split, with all models showing negative R^2 values. XGBoost, despite being the least unfavorable (MAE = 6,543.49 ± 3,803.45; RMSE = 7,934.74 ± 4,428.98; R^2 = -0.940 ± 2.148), failed to generalize consistently across folds. Random Forest, CatBoost, Elastic

*Amiq Fahmi

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15312
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Net, and Stacking showed significant instability, with very variable and negative R^2 values. LightGBM had lower variance (R^2 = -1.387 ± 3.315), but its predictive capacity remained limited.

These findings highlight the sensitivity of cross-validation with small-sample socioeconomic data, as fold composition has a significant impact on predictive performance. Diagnostic investigation found that the regency/city of Brebes was a significant observation. Excluding Brebes resulted in significantly worse model performance (MAE = 7,875.57; RMSE = 8,888.52; R^2 = 0.535) compared to the whole dataset (MAE = 2,640.99; RMSE = 3,202.79; R^2 = 0.942). Folds without Brebes yielded unstable and weak results, resulting in negative average R^2 values in Table 5. Thus, the instability of k-fold validation should not be regarded as a general failure of the algorithms, but rather as evidence of individual regencies/cities having disproportionate influence in small-area poverty statistics.

Table 5. Cross-Vallidation Results (10-Fold)

Model	MAE (± SD)	RMSE (± SD)	R^2 (± SD)
XGBoost	6543.49 ± 3803.45	7934.74 ± 4428.98	-0.940 ± 2.148
Random Forest	6971.40 ± 2951.92	8113.83 ± 3829.94	-4.934 ± 10.903
CatBoost	7567.74 ± 3278.84	8663.90 ± 4488.64	-6.278 ± 13.293
LightGBM	7126.8690 ± 3418.7014	8437.0971 ± 4406.1101	-1.3870 ± 3.3150
Elastic Net	7724.16 ± 3503.70	8766.01 ± 4060.00	–5.276 ± 10.092
Stacking	6894.51 ± 2828.83	8546.66 ± 3559.40	-5.478 ± 14.284

Overall, Tables 4 and 5 show a contrast between the hopeful results of hold-out evaluation and the more conservative conclusions of cross-validation. The discrepancy shows that the hold-out split benefited from influential cases, whereas cross-validation subsets frequently eliminated structurally important regencies/cities, lowering accuracy. Despite this, XGBoost and Stacking regularly outperformed other learners, supporting previous findings that boosting-based algorithms yield higher accuracy in multidimensional poverty modeling (Izzati et al., 2024; Zheng et al., 2024).

To further verify robustness, a noise perturbation analysis was performed by injecting Gaussian noise into the feature set at 5%, 10%, and 20% of each variable's standard deviation. This approach mimicked measurement mistakes or reporting inaccuracies seen in socioeconomic survey data. Table 6 summarizes model performance at different noise levels.

Table 6. Noise Test Result (Hold-out Evaluation with Added Noise)

Model	MAE (0%)	RMSE (0%)	R^2 (0%)	MAE (5%)	RMSE (5%)	R^2 (5%)	MAE (10%)	RMSE (10%)	R^2 (10%)	MAE (20%)	RMSE (20%)	R^2 (20%)
XGBoost	2180.01	3512.07	0.931	2739.45	3889.60	0.915	4210.59	6813.55	0.740	5811.99	7840.91	0.655
Random Forest	5310.83	5999.67	0.800	5338.45	6014.86	0.797	5438.35	6034.00	0.796	5171.04	5679.46	0.819
CatBoost	6722.99	6841.92	0.738	6339.39	6487.47	0.764	7878.86	8292.68	0.614	8017.81	8382.80	0.606
LightGBM	4386.29	5163.74	0.851	4136.74	4974.98	0.861	5823.37	7624.79	0.674	5421.23	7686.73	0.669
Elastic Net	7777.85	9031.75	0.543	7764.97	9330.58	0.512	7559.23	8639.86	0.582	7470.48	9796.79	0.462
Stacking	2640.99	3202.79	0.942	2633.91	3060.16	0.948	3101.74	3517.56	0.932	4082.74	4546.97	0.888

Table 6 highlights all models' robustness to incremental noise perturbations. The Stacking regressor had the highest robustness, retaining an R^2 over 0.88 even with 20% noise and only experiencing moderate increases in MAE and RMSE. XGBoost also performed well, with accuracy progressively dropping from 0.930 to 0.655 as noise rose, demonstrating that the model is still dependable with moderate measurement errors. Random Forest demonstrated sustained stability across noise levels, while having poorer baseline accuracy than boosting and stacking approaches. CatBoost and LightGBM, on the other hand, showed larger performance reductions over 10% noise, indicating a higher susceptibility to disturbances.

*Amiq Fahmi

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15312
e-ISSN : 2541-2019
p-ISSN : 2541-044X

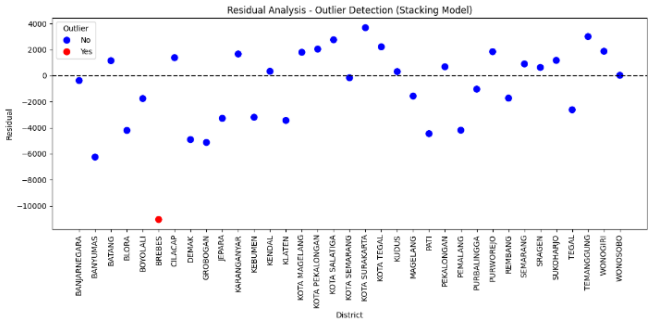

Fig. 9 Residual analysis of the Stacking model identifying Brebes as an influential regency.

Elastic Net, albeit less accurate overall than ensemble learners, showed a mixed robustness pattern. The R^2 ranged between 0.543 (0% noise) and 0.462 (20% noise), indicating that while regularization reduces variance, the linear structure is still susceptible to significant perturbations. Nonetheless, Elastic Net continues to act as a transparent linear baseline for more complicated models. Taken together, these findings highlight that boosting and stacking methodologies are not only more accurate but also more robust to real-world defects such as reporting errors and survey discrepancies, resulting in more credible information for poverty policy research.

The Stacking model's residual analysis identified Brebes as an influential outlier, having a significantly larger residual than the other regencies/cities. This insight explains the instability of cross-validation results, as folds eliminating Brebes yielded significantly lower and inconsistent R^2 values.

Table 7. Comparison of Stacking model performance with and without Brebes.

Model	MAE (All)	RMSE (All)	R^2 (All)	MAE (No Brebes)	RMSE (No Brebes)	R^2 (No Brebes)
Stacking	2640.99	3202.79	0.942	7875.57	8888.52	0.535

Excluding Brebes significantly lowered prediction accuracy, with R^2 falling from 0.942 (all data) to 0.535, indicating its influence on model stability and explaining the alarming cross-validation findings given in Table 5. Beyond this influence, it is crucial to assess the consistency of poverty determinants among algorithms. As a result, a feature stability study was conducted utilizing SHAP values for tree-based models, coefficient magnitudes for Elastic Net, and meta-learner coefficient stability for the Stacking ensemble.

Table 8. Feature Stability across Models (Top 3 Features per Model)

Model	Feature	Mean	Std	CV
XGBoost	Number of Households	7249.53	1128.96	0.156
	Uninhabitable Houses	2825.37	1244.18	0.440
	Per Capita Expenditure	2100.44	777.69	0.370
Random Forest	Number of Households	4679.76	775.42	0.166
	Uninhabitable Houses	2737.62	1207.86	0.441
	Elementary School Graduates Employed	1053.74	427.91	0.406
CatBoost	Number of Households	2272.35	221.77	0.098
	Uninhabitable Houses	1907.67	245.38	0.129
	Per Capita Expenditure	1823.14	287.97	0.158
LightGBM	Number of Households	8318.49	3544.77	0.426
	Uninhabitable Houses	2244.75	1247.98	0.556
	Per Capita Expenditure	1907.97	1300.19	0.681
Elastic Net	Number of Households	10504.28	1111.64	0.106
	Junior High Education & Unemployed	5220.00	1065.08	0.204
	Population	-3617.41	1108.86	0.307

Tabel 9. Meta-Learner Coefficient Stability (Stacking ElasticNet)

Base Learner	Mean Coefficient	Std	CV
Random Forest	0.94	0.62	0.66
XGBoost	0.21	0.32	1.50
CatBoost	0.13	0.55	4.23
LightGBM	-0.14	0.84	-5.88

*Amiq Fahmi

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15312
e-ISSN : 2541-2019
p-ISSN : 2541-044X

To assess robustness further, a feature stability study was performed utilizing SHAP values for tree-based learners, coefficient analysis for Elastic Net regression, and the stacking meta-learner. Table 8 shows each model's top three attributes, together with its mean importance, variability (standard deviation), and stability index (coefficient of variation, CV). The results reveal that Number of Households, Uninhabitable Houses, and Per Capita Expenditure are consistently the most influential predictors in tree-based algorithms. Elastic Net regression also exposes socioeconomic determinants, including the number of households, junior high education, unemployment, and population.

In parallel, Table 9 shows the coefficient stability of the stacking meta-learner. According to the data, Random Forest and XGBoost receive the most significant and most consistent weights, whereas CatBoost contributes slightly, and LightGBM has substantial variability, with an average negative coefficient. This trend implies that the stacking ensemble primarily relies on Random Forest and XGBoost as fundamental predictors, with other learners making complementing but less consistent contributions.

Negative coefficients in Elastic Net (population) and the stacking meta-learner (LightGBM) indicate inverse correlations, implying that increasing these predictors lowers the poverty index or that the ensemble applies corrective weights to less consistent base learners.

Overall, these findings show that, despite differences in modeling methodologies and regency/city-level sensitivity, there is significant convergence in identifying the key poverty determinants. This consistency improves the models' interpretability and policy relevance, confirming the claim that ensemble and boosting approaches improve predictive accuracy while also providing strong and dependable insights for poverty assessment and policy design.

Ramayanti et al. (2023) used the Random Forest method to discover key poverty indicators such as the household head's occupation and access to clean water. The SHAP summary visualization (Fig. 10) shows how characteristics influence model predictions, with red points indicating high feature values and blue points indicating low values; points on the right increase the anticipated poverty index, while points on the left decrease it. Valentika et al. (2024) highlighted prediction intervals by combining Random Forest and HDR/LM methods. This study, however, combined XGBoost's strong predictive potential with SHAP to improve interpretability. Fobi et al. (2023) emphasized the significance of interpretability, demonstrating that combining household survey data with satellite-derived characteristics enhanced model accuracy by 3.71%-4.09% via feature selection and complementing factors, providing policymakers with clearer insights.

Fig. 10 SHAP Summary Plot (XGBoost)

The SHAP summary plot showed the Number of Households, Uninhabitable Houses, and Per Capita Expenditure as the most important predictors of the poverty index. Poverty estimates rose with more households but fell with higher per capita expenditure. Huang et al. (2023) used a similar approach, combining SHAP and Partial Dependence Plots (PDP), to identify important poverty determinants. Such interpretability strategies assist in transforming XGBoost from a 'black box' to a transparent, policy-relevant instrument for evidence-based decision-making.

*Amiq Fahmi

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15312
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Fig. 10 shows how the contribution of each feature altered depending on its input value, using red (high value) and blue (low value) hues. This provides a more in-depth understanding of how specific attributes influence regency/city-level predictions. These interpretive findings help guide the development of more targeted policy initiatives, such as investing in areas with large household sizes and low per capita spending, improving housing infrastructure and essential services, and improving healthcare and educational quality. When these findings are compared to the heatmap, noticeable correlation patterns emerge. The number of households correlates favorably with the poverty index, indicating a positive SHAP contribution. Per capita expenditure has a negative connection, corresponding to its negative SHAP value, but uninhabitable buildings correlate favorably, supporting their involvement in poverty prediction. This agreement between exploratory investigation and model interpretation improves the credibility of the results. It demonstrates that a complex interplay of demographic, economic, and infrastructure factors influences poverty in Central Java.

As shown in Fig. 11, the number of households was the largest predictor of poverty in Central Java, followed by uninhabitable houses and per capita expenditure. Other characteristics, such as educational attainment, the number of hospitals, and regency/city minimum wages, made smaller but still significant contributions. This interpretation suggests that the XGBoost model's outputs might be used as a forecasting tool to construct better targeted poverty reduction strategies (Fobi et al., 2023). Areas with a higher number of households were more vulnerable to poverty, whereas higher per capita expenditure reduced poverty risk. Uninhabitable housing reflects structural and social disadvantages, exacerbating poverty rates. Additional characteristics, such as land acreage, education levels, and employment, influence community wellbeing, access to public services, and overall socioeconomic situations.

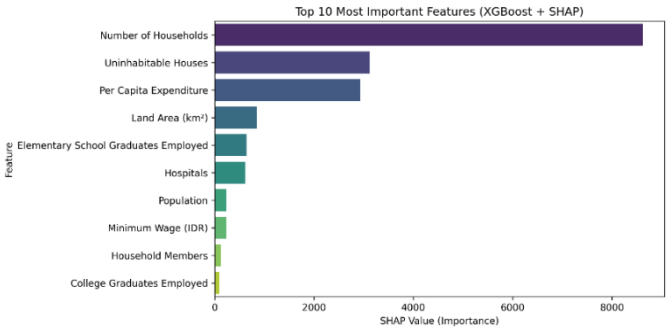

Fig. 11 SHAP Bar Plot (XG Boost)

This study extended XGBoost by applying model-agnostic interpretability techniques to the stacking ensemble. Because the stacking meta-learner (Elastic Net Regression) does not directly use the original features, SHAP cannot be usefully implemented at this level. Instead, the top three features found by SHAP on the best-performing base learner (XGBoost), namely Number of Households, Uninhabitable Houses, and Per Capita Expenditure, were evaluated further using Partial Dependence Plots (PDP) and Accumulated Local Effects (ALE) within the stacking model. This combination technique ensures that the most influential features are consistently tested across models, while also exploiting the predictive power of stacking to investigate their marginal effects.

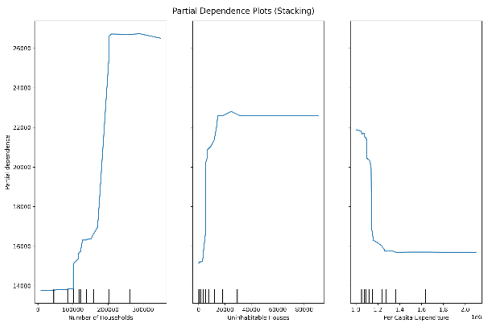

Fig. 12 PDP Plots (Stacking)

Fig. 12 depicts PDP results for the stacking model, demonstrating substantial non-linear connections between Number of Households, Uninhabitable Houses, Per Capita Expenditure, and the poverty index. Higher household counts and unfit dwellings boosted poverty estimations, whereas higher per capita expenditure decreased them.

*Amiq Fahmi

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15312
e-ISSN : 2541-2019
p-ISSN : 2541-044X

In addition to PDP, ALE plots provide a more robust interpretability framework by accounting for feature interactions and eliminating extrapolation bias. The ALE plots demonstrate that the poverty index rises quickly with the number of households and uninhabitable dwellings, but increased per capita expenditure considerably lowers poverty risks. When taken together, these findings show that a complex interplay of demographic, economic, and infrastructural elements shapes poverty in Central Java. The combination of SHAP (XGBoost) with PDP/ALE (Stacking) increases model transparency. It provides policymakers with practical information, such as prioritizing measures to improve housing conditions, assist large households, improve healthcare access, and boost per capita expenditure capacity.

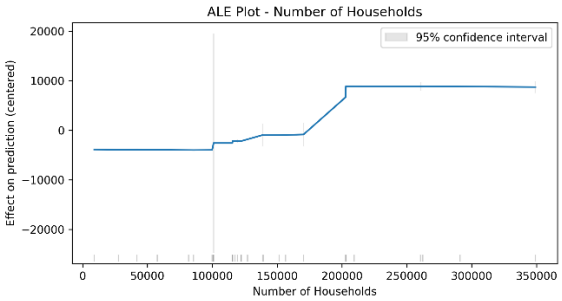

Fig. 13 ALE Plot – Number of Households (Stacking)

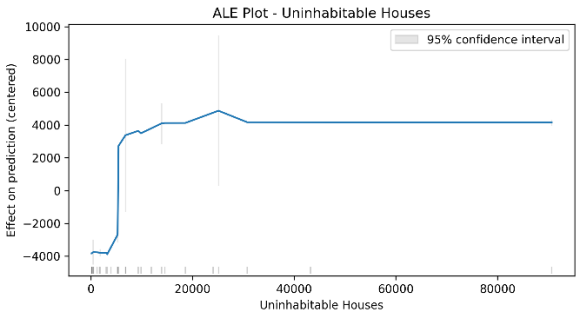

Fig. 14 ALE Plot – Uninhabitable Houses (Stacking)

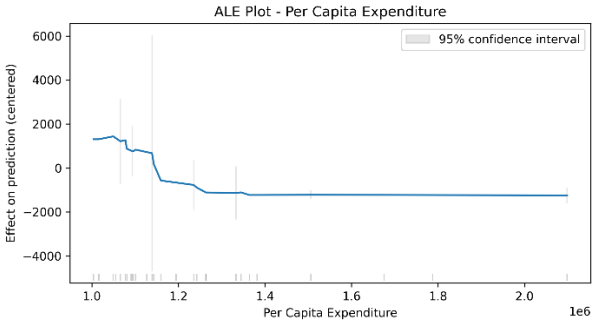

Fig. 15 Per Capita Expenditure (Stacking)

DISCUSSIONS

This study used a variety of machine learning techniques, including XGBoost, Random Forest, CatBoost, LightGBM, Elastic Net Regression, and a Stacking ensemble, to predict and assess the causes of poverty in Central Java Province. XGBoost outperformed individual learners, with MAE = 2,180.01, RMSE = 3,512.07, and R^2 = 0.931. Random Forest had a moderate predictive capacity (MAE = 5,310.83; RMSE = 5,999.67; R^2 = 0.800), while CatBoost gave inferior findings (MAE = 6,722.99; RMSE = 6,841.92; R^2 = 0.738). LightGBM outperformed Random Forest and CatBoost with an MAE of 4,386.29, RMSE of 5,163.74, and R^2 of 0.851. Elastic Net Regression was the least successful (MAE = 7,777.85; RMSE = 9,031.75; R^2 = 0.543), highlighting the limitations of solely linear approaches for predicting complex socioeconomic connections. The Stacking ensemble outperformed single models, with MAE = 2,640.99, RMSE = 3,202.79, and R^2 = 0.942. This highlights the benefit of integrating various learners to capture complicated socioeconomic trends. These findings are consistent with Zheng et al. (2024), who underlined the effectiveness of boosting and stacking strategies in multidimensional poverty modeling.

Interpretability tests proved the importance of socioeconomic and infrastructure drivers. SHAP findings for XGBoost revealed that the most influential predictors were the number of households, uninhabitable houses, and per capita expenditure. Higher household numbers and unfit housing increased poverty risk, whereas higher per capita expenditure decreased it. These findings are consistent with Ramayanti et al. (2023), who emphasized household occupation and access to clean water, and Huang et al. (2023), who used SHAP and PDP to identify poverty drivers. Complementary analyses using PDP and ALE on the Stacking ensemble demonstrated similar nonlinear effects, confirming the consistency of important determinants across models. SHAP, PDP, and ALE work together to transform machine learning models that are commonly referred to as "black boxes" into transparent, policy-relevant tools (Fobi et al., 2023).

Robustness checks highlighted the problems of small-area data. The 10-fold cross-validation yielded lower results than the hold-out split, with all models exhibiting negative R^2 values, such as XGBoost (–0.940 ± 2.148) and Stacking (–5.478 ± 14.284). Excluding Brebes reduced the Stacking performance from R^2 = 0.942 to 0.535,

*Amiq Fahmi

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15312
e-ISSN : 2541-2019
p-ISSN : 2541-044X

resulting in unstable cross-validation outcomes. Noise perturbation experiments revealed that boosting and stacking remained stable under moderate measurement errors; however, Elastic Net constantly underperformed. These findings support previous research (Izzati et al., 2024a; Zheng et al., 2024), which suggests that the instability is caused by small-sample sensitivity to influential regencies/cities rather than algorithmic restrictions.

From a policy standpoint, counterfactual simulations provided quantitative evidence of intervention effects. A 10% rise in per capita expenditure reduced the poverty index by 9.9%, a 10% decrease in household numbers by 11.3%, and a 20% reduction in uninhabitable housing by 2.1%. These findings indicate that interventions aimed at increasing household consumption and addressing demographic constraints have the most excellent aggregate effects. However, housing rehabilitation, while socially important, contributes only modest direct reductions. Such findings highlight the significance of consumption-oriented programs, including Program Keluarga Harapan and Bantuan Langsung Tunai, housing rehabilitation initiatives (Kartu Jateng Sejahtera), and family planning techniques. Furthermore, contributions from health and education emphasize the importance of investing in rural healthcare and vocational training, echoing previous requests to integrate predictive modeling into social protection policies (Fobi et al., 2023; Valentika et al., 2024).

Despite these advances, a few restrictions should be recognized. The dataset only contains 35 regencies/cities, which limits generalizability and makes results susceptible to outliers. The cross-sectional approach limits temporal validation, and the lack of external data precludes replication across areas. Furthermore, SHAP, PDP, and ALE improve interpretability without establishing causality. Future research should use larger longitudinal datasets and alternate sources, such as satellite images and socioeconomic time series, as indicated in previous studies.

In conclusion, this work demonstrates that boosting and stacking models not only improve predicted accuracy but, when combined with interpretability techniques, also produce transparent and actionable results. These findings bolster the rationale for incorporating Explainable AI into poverty policy design, allowing for more effective solutions to combat multidimensional poverty in Central Java.

CONCLUSION

This study demonstrates that interpretable machine learning algorithms—particularly boosting and stacking ensemble methods—can effectively predict and elucidate poverty levels in Central Java. Among the individual models, XGBoost had the best prediction performance, while the Stacking ensemble outperformed all learners, demonstrating the advantages of mixing multiple algorithms to capture complex socioeconomic phenomena. In contrast, Elastic Net regression performed poorly, showing the limitations of linear models when applied to multidimensional poverty data. According to interpretability evaluations, regencies/cities with bigger household sizes, a higher share of uninhabitable housing, and lower per capita expenditure are the most prone to poverty.

Counterfactual simulations demonstrated the potential impact of targeted interventions: a 10% increase in per capita expenditure reduced predicted poverty levels by 9.9%, a 10% decrease in household numbers reduced them by 11.3%, and a 20% reduction in uninhabitable houses reduced poverty by 2.1%. These findings emphasize the policy importance of increasing household consumption capacity, managing population pressures, and improving housing quality. The findings give empirical support for improving national programs such as Program Keluarga Harapan (PKH), Bantuan Langsung Tunai (BLT), and Bantuan Pangan Non-Tunai (BPNT), as well as strengthening local efforts such as Kartu Jateng Sejahtera and regency/city-level housing restoration.

Despite these achievements, many restrictions should be considered. The dataset only included 35 regencies/cities in Central Java, with no external validation from other provinces, restricting generalizability. The cross-sectional design also precludes temporal validation. Furthermore, while SHAP, PDP, and ALE increased transparency, they focused on connections rather than causal consequences. Future studies should broaden the dataset's regional and temporal coverage, including complementary data sources such as satellite images, and use longitudinal validation to improve robustness.

Overall, this work shows that combining boosting and stacking with explainable AI tools results in accurate, transparent, and actionable models. Including these approaches in decision-support systems may allow policymakers to develop more focused, evidence-based poverty alleviation initiatives, ensuring that resources are distributed more efficiently to the most vulnerable groups. These findings not only push the technical boundaries of poverty prediction, but they also provide practical decision-support evidence for focusing on multidimensional poverty remedies.

ACKNOWLEDGMENT

The authors sincerely thank the Faculty of Computer Science at Universitas Dian Nuswantoro for its invaluable support, guidance, and conducive academic environment throughout this study. We also gratefully acknowledge the Central Java Statistics Office for providing the essential datasets that enabled the successful completion of this study.

*Amiq Fahmi

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15312
e-ISSN : 2541-2019
p-ISSN : 2541-044X

REFERENCES

Apley, D. W., & Zhu, J. (2020). Visualizing the effects of predictor variables in black box supervised learning models. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, *82*(4), 1059–1086. https://doi.org/10.1111/rssb.12377

Badan Pusat Statistik BPS. (2024). Kemiskinan 2024 Jawa Tengah. In *Badan Pusat Statistik Provinsi Jawa Tengah*. https://jateng.bps.go.id/id/statistics-table/2/MzQjMg==/kemiskinan.html

Binka, M., Klaver, B., Cua, G., Wong, A. W., Fibke, C., Velásquez García, H. A., Adu, P., Levin, A., Mishra, S., Sander, B., Sbihi, H., & Janjua, N. Z. (2022). An Elastic Net Regression Model for Identifying Long COVID Patients Using Health Administrative Data: A Population-Based Study. *Open Forum Infectious Diseases*, *9*(12). https://doi.org/10.1093/ofid/ofac640

Chi, G., Fang, H., Chatterjee, S., & Blumenstock, J. E. (2022). Microestimates of wealth for all low- and middle-income countries. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(3), 1–11. https://doi.org/10.1073/pnas.2113658119

Christoph Molnar. (2025). *Interpretable Machine Learning A Guide for Making Black Box Models Explainable* (3rd ed.).

Corral Rodas, P., Henderson, H., & Segovia, S. (2023). Poverty mapping in the age of machine learning. *Available at SSRN 4587156*, *May*.

Dorogush, A. V., Ershov, V., & Gulin, A. (2018). *CatBoost: gradient boosting with categorical features support*. 1–7.

Fobi, S., Cardona, M., Collins, E., Robinson, C., Ortiz, A., Sederholm, T., Dodhia, R., & Ferres, J. L. (2023). Poverty rate prediction using multi-modal survey and earth observation data. *COMPASS 2023 - Proceedings of the ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies*, 23–29. https://doi.org/10.1145/3588001.3609359

Greenwell, B. M., Boehmke, B. C., & McCarthy, A. J. (2018). *A Simple and Effective Model-Based Variable Importance Measure*. http://arxiv.org/abs/1805.04755

Hall, O., Ohlsson, M., & Rögnvaldsson, T. (2022). A review of explainable AI in the satellite data, deep machine learning, and human poverty domain. *Patterns*, *3*(10). https://doi.org/10.1016/j.patter.2022.100600

Huang, W., Liu, Y., Hu, P., Ding, S., Gao, S., & Zhang, M. (2023). Heliyon What influence farmers ' relative poverty in China : A global analysis based on statistical and interpretable machine learning methods. *Heliyon*, *9*(9), e19525. https://doi.org/10.1016/j.heliyon.2023.e19525

Izzati, F., Masjkur, M., & Afendi, F. M. (2024). Comparison of Chi-Square Automatic Interaction Detector (CHAID) and Random Forest Methods in the Classification of Household Poverty Status in Central Java. *Indonesian Journal of Statistics and Its Applications*, *8*(1), 1–13. https://doi.org/10.29244/ijsa.v8i1p1-13

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). *LightGBM: A Highly Efficient Gradient Boosting Decision Tree*. https://github.com/Microsoft/LightGBM.

Kuswanto, H., Rouhan, A. A., Qori, M., Hia, S., & Fithriasari, K. (2025). *Per capita expenditure prediction using model stacking based on satellite imagery*. *14*(2), 1220–1231. https://doi.org/10.11591/ijai.v14.i2.pp1220-1231

Lastras Rodríguez, C. A. (2024). Predicting social welfare in Madrid neighbourhoods using machine learning. *Regional Studies, Regional Science*, *11*(1), 496–522. https://doi.org/10.1080/21681376.2024.2380890

Lee, E., Ong, T. S., & Lee, Y. (2024). Evaluating Household Consumption Patterns: Comparative Analysis Using Ordinary Least Squares and Random Forest Regression Models. *HighTech and Innovation Journal*, *5*(2), 489–507. https://doi.org/10.28991/HIJ-2024-05-02-019

Li, Q., Yu, S., Échevin, D., & Fan, M. (2022). Is poverty predictable with machine learning? A study of DHS data from Kyrgyzstan. *Socio-economic Planning Sciences*, *81*, 101195. https://doi.org/https://doi.org/10.1016/j.seps.2021.101195

Lundberg, S., & Lee, S.-I. (2017). *A Unified Approach to Interpreting Model Predictions*. http://arxiv.org/abs/1705.07874

Muñetón-Santa, G., & Manrique-Ruiz, L. C. (2023). Predicting Multidimensional Poverty with Machine Learning Algorithms: An Open Data Source Approach Using Spatial Data. *Social Sciences*, *12*(5). https://doi.org/10.3390/socsci12050296

Putra, I. K. P. P. W., Irhamah, Iriawan, N., & Fithriasari, K. (2024). Predicting Poverty Percentage Based on Satellite Imagery and Point of Interest Using Support Vector Regression and Random Forest Regression (Case Study of Central Java Province). In D. Adzkiya & K. Fahim (Eds.), *Applied and Computational Mathematics* (pp. 309–323). Springer Nature Singapore.

Ramayanti, F., Dodi Vionanda, Dony Permana, & Zilrahmi. (2023). Application of Random Forest to Identify for Poor Households in West Sumatera Province. *UNP Journal of Statistics and Data Science*, *1*(2), 97–104. https://doi.org/10.24036/ujsds/vol1-iss2/31

*Amiq Fahmi

Sagi, O., & Rokach, L. (2018). Ensemble learning: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *8*, e1249. https://doi.org/10.1002/widm.1249

Salvador, E. L. (2024). *Use of Boosting Algorithms in Household-Level Poverty Measurement: A Machine Learning Approach to Predict and Classify Household Wealth Quintiles in the Philippines*. http://arxiv.org/abs/2407.13061

Satapathy, S., Saravanan, S., Shruti, M., & Mohanty, S. (2023). A Comparative Analysis of Multidimensional COVID-19 Poverty Determinants: An Observational Machine Learning Approach. *New Generation Computing*, *41*. https://doi.org/10.1007/s00354-023-00203-8

Solís-Salazar, M., & Madrigal-Sanabria, J. (2022). Una propuesta de aprendizaje automático para predecir la pobreza. *Revista Tecnología En Marcha*, *35*, 84–94. https://doi.org/10.18845/tm.v35i4.5766

Valentika, N., Notodiputro, K. A., & Sartono, B. (2024). Performance Study of Prediction Intervals with Random Forest for Poverty Data Analysis. *Jurnal Aplikasi Statistika & Komputasi Statistik*, *16*(1), 80–88. https://doi.org/10.34123/jurnalasks.v16i1.542

Wang, N., Zhang, H., Dahal, A., Cheng, W., Zhao, M., & Lombardo, L. (2024). On the use of explainable AI for susceptibility modeling: Examining the spatial pattern of SHAP values. *Geoscience Frontiers*, *15*(4), 101800. https://doi.org/10.1016/j.gsf.2024.101800

Wolpert, D. H. (1992). Stacked generalization. *Neural Networks*, *5*(2), 241–259. https://doi.org/10.1016/S0893-6080(05)80023-1

Zamzuri, M. H. A., Sofian, N., & Hassan, R. (2023). Forecasting of Poverty using the Ensemble Learning Classification Methods. *International Journal on Perceptive and Cognitive Computing*, *9*(1), 24–32. https://doi.org/10.31436/ijpcc.v9i1.326

Zhao, X., Yu, B., Liu, Y., Chen, Z., Li, Q., Wang, C., & Wu, J. (2019). Estimation of poverty using random forest regression with multi-source data: A case study in Bangladesh. *Remote Sensing*, *11*(4), 1–18. https://doi.org/10.3390/rs11040375

Zheng, X., Zhang, W., Deng, H., & Zhang, H. (2024). County-Level Poverty Evaluation Using Machine Learning, Nighttime Light, and Geospatial Data. *Remote Sensing*, *16*(6). https://doi.org/10.3390/rs16060962

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. In *J. R. Statist. Soc. B* (Vol. 67, Issue 2). https://academic.oup.com/jrsssb/article/67/2/301/7109482

*Amiq Fahmi