Advancements in Household Classification Using Multiclass Decision Forest: A Case Study of the Kut Bak District, Sakon Nakhon Province

Wichidtra Sudjarid,[a] Wipawan Buathong,[b] & Pita Jarupunphol[b]*

[a] Department of Environmental Science, Sakon Nakhon Rajabhat University, Sakon Nakhon, 47000 Thailand
[b]Department of Digital Technology, Phuket Rajabhat University, Phuket, 83000 Thailand

Abstract

This study examines the classification of household poverty in Kut Bak District, Sakon Nakhon Province, Thailand, using advanced machine learning techniques, specifically the Multiclass Decision Forest algorithm. The research pursued two main objectives: (1) to develop a robust model for classifying households into distinct poverty levels, and (2) to identify key socio-economic factors that significantly influence poverty classification. The dataset comprised 302 households, representing various socio-economic strata within the district. The study population consisted of 686 households, with the sample selected through a participatory process involving local community leaders and government officials. Nineteen socio-economic features were included in the model, such as household income, expenditures, and participation in local development projects. Households were classified into four poverty levels: extremely poor, moderately poor, marginally poor, and above poor. Model performance was evaluated using confusion matrices, yielding an overall accuracy of 94.4%, including perfect (100%) accuracy in classifying the extreme categories (extremely poor and above poor). The analysis further revealed that household income and participation in local projects were the most influential determinants of poverty levels. Specifically, households with low income and minimal project participation were most likely to fall into the “extremely poor” category, whereas those with higher income and multiple income sources were typically classified as “above poor.” The findings provide important insights into the socio-economic dynamics of poverty in rural Thailand and offer practical implications for designing targeted poverty alleviation strategies. More broadly, the study demonstrates the potential of machine learning approaches to address complex socio-economic challenges, thereby contributing to the field of development studies.

Keywords: Machine learning, Multiclass decision forest, Poverty classification, Socio-Economic analysis

*** Corresponding Author**
e-mail: p.jarupunphol@pkru.ac.th

Introduction

Poverty classification is crucial in understanding and addressing socio-economic disparities (Huang & Xia, 2023). This process is crucial for effectively allocating resources, policy-making, and targeted interventions aimed at poverty alleviation. The complexity of poverty, encompassing financial constraints and broader socio-economic factors, necessitates nuanced and sophisticated classification methods (Morris et al., 2018). Globally, poverty is a pervasive issue, affecting billions of people. The World Bank and other international organizations often define poverty as income (World Bank, 2023a). However, this monetary approach does not capture the multifaceted nature of poverty, which includes lack of access to education, healthcare, and other essential services (Lemanski, 2016). Regionally, the context of poverty varies significantly. In developing countries, poverty is often associated with rural areas with limited access to resources, education, and employment opportunities. In contrast, in developed countries, poverty might be more linked to systemic issues and urban inequality (World Bank, 2023b).

The classification of household poverty is of paramount importance for several reasons. For example, accurate poverty classification allows governments and organizations to tailor their policies and resources to the needs of the most vulnerable (Talingdan, 2019). It helps identify which households require immediate assistance, and what kind of support would be most effective. In addition, classifying household poverty on multiple dimensions enables a more holistic understanding of deprivation as poverty extends beyond mere financial constraints; it encompasses access to education, healthcare, sanitation, and employment opportunities. Furthermore, understanding impoverished households’ specific needs and challenges can lead to more participatory and community-driven development strategies. This empowers communities to be part of the solution, ensuring sustainable development. To address these challenges, there is a need for innovative approaches, such as leveraging technology for better data collection and analysis, adopting dynamic models that can adjust to the changing nature of poverty, and ensuring participation and input from diverse groups to mitigate subjectivity and bias. Multiclass Decision Forest and other advanced analytical tools offer promising avenues for more accurate and insightful classification, helping to navigate these complexities more effectively.

In Thailand, Kut Bak District (Wikipedia, 2023) in Sakon Nakhon Province, presents a unique and challenging context for poverty classification. This area is among the poorest in Thailand, characterized by a predominantly rural population, reliance on agriculture, and limited economic diversification. The socio-economic landscape of Kut Bak District is marked by limited access to markets, dependency on agriculture, and a lack of alternative income sources. Characterized by high poverty rates and a heavy reliance on subsistence agriculture, the district exemplifies the struggles faced by many rural areas in Thailand. In addition, comprehensive, localized economic data for Kut Bak District is lacking, which impedes the formulation and implementation of effective policies. These factors contribute to a high vulnerability to poverty, making it crucial to understand the various levels and dimensions of poverty in this area. This research aims to bridge that gap by providing detailed insights into the region's economic dynamics and poverty landscape. Enhanced data availability is expected to facilitate more informed decision-making and policy development, which are essential for the district's socio-economic development.

In this case, a multidimensional approach to poverty classification is essential (Alkire & Foster, 2011; Thorbecke, 2013). The advancement in data collection methods and analytical techniques, like the Multiclass Decision Forest, offers new opportunities for more accurate and insightful poverty classification. By leveraging detailed household-level data, this approach can uncover patterns and correlations that traditional methods might lack. In Kut Bak District, this methodology can provide a deeper understanding of the poverty dynamics. The classification based on incomes, expenses, and project participation, encompassed by various sub-features, allows for a comprehensive analysis of the factors contributing to poverty in this region. This, in turn, can inform more effective and targeted interventions. Through this extended and detailed analysis of results, including the derivation of rules for poverty classification, the study aims to offer actionable insights and a structured framework for understanding and addressing the different household poverty levels in Kut Bak District.

Objectives

The objectives of this research are:

1) To develop and demonstrate a machine learning-based system, explicitly utilizing a Multiclass Decision Forest, to classify households into distinct poverty levels based on various socio-economic parameters.
2) To construct a predictive model capable of processing and analyzing complex socio-economic data, thereby facilitating a more nuanced understanding of poverty.

Conceptual Framework

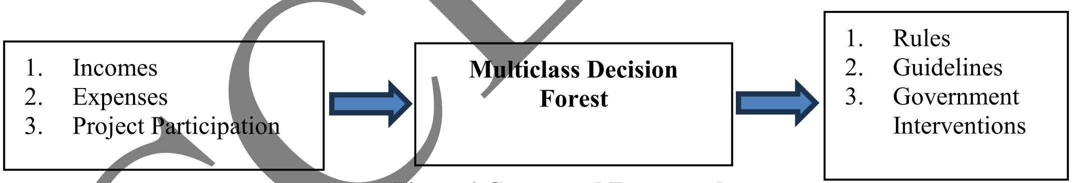

Figure 1 Conceptual Framework

Literature Review

This section delves into the extensive research developed in the field of poverty classification and the use of machine learning techniques. The exploration begins by examining the various traditional methods for classifying household poverty, highlighting their evolution and the complexities involved. Then, it discusses the emergence and integration of machine learning techniques in poverty classification, emphasizing how these advanced approaches have revolutionized the field. Finally, the review focuses specifically on Multiclass Decision Forest, a particular type of machine learning algorithm, detailing their applications and effectiveness in various domains, including but not limited to poverty classification.

1. Methods for Classifying Household Poverty

The classification of household poverty has evolved over time, incorporating various methods and approaches to capture the multifaceted nature of poverty. Traditionally, poverty has been classified based on income or consumption levels. For example, the World Bank's international poverty line is a well-known benchmark in this category (Mansi et al., 2020). In addition, multidimensional poverty index (MPI), developed by the Oxford Poverty and Human Development Initiative (OPHI), assesses poverty across multiple dimensions, including education, health, and living standards (United Nations, 2023). Household poverty classification can also be effectively achieved using composite indices. These indices amalgamate multiple indicators into a single measure, providing a comprehensive view of poverty (Gweshengwe & Hassan, 2020). Examples of such indices include the Human Poverty Index (Chakravarty & Majumder, 2005), which combines various aspects of deprivation, and the Social Vulnerability Index (Mah et al., 2023), which integrates factors contributing to a household's susceptibility to social and economic hardships. Asset-based approaches evaluate poverty levels by examining ownership of assets such as land, housing, and livestock (Carter & Barrett, 2006). This method is particularly useful in rural areas where income data might not be consistently reliable. Another method is the basic needs approach (Kharas & Dooley, 2022), which assesses whether households can fulfil fundamental requirements like food, shelter, and clothing, often by establishing a baseline standard for these necessities (Wong, 2012). Additionally, participatory approaches directly engage communities in poverty assessment, leveraging local insights and experiences to define and measure poverty, ensuring a more inclusive and representative evaluation (Islam et al., 2023).

2. Machine Learning Techniques for Poverty Classifications

In recent years, the field of poverty classification has witnessed a significant paradigm shift with the advent of machine learning techniques (Alsharkawi et al., 2021; Huang & Xia, 2023; Muñetón-Santa et al., 2022; Sihombing & Arsani, 2021). These advanced computational methods have introduced a new dimension to poverty analysis, transcending the traditional boundaries of statistical models. At its core, machine learning involves using algorithms that can learn from and make predictions or decisions based on data. This ability is particularly advantageous in the context of poverty classification, where the complexities and dynamics of socio-economic factors pose a substantial challenge. Techniques like decision trees, neural networks, and support vector machines have been widely used among the various machine learning methodologies. Each technique offers distinct advantages, depending on the nature of the data and the study's specific objectives. Decision trees are particularly useful for their simplicity and interpretability, making them an excellent tool for stakeholders who may not have advanced technical knowledge.

Huang & Xia (2023) classified rural relative poverty groups and measured the influence of land elements in China, based on a questionnaire analysis of 23 poor counties. The paper utilizes evaluation indexes, including per capita household income, education level, poverty registration situation, employment situation, critical disease situation, natural disaster frequency situation, etc. Primarily, the paper establishes a decision tree for rural relative poverty group evaluation based on an improved ID3 algorithm and quantifies the effect of different land elements using an ordered logistic model. The paper finds that a better condition of land resource endowment leads to a lower degree of rural relative poverty, but over-reliance on land increases the risk of relative poverty. Moreover, the paper finds that the asset value and capital value of rural land are not evident. The paper suggests policy implications and future research directions for rural relative poverty alleviation.

Sihombing & Arsani (2021) modeled poverty using the machine learning classification method. The data used are imbalanced data, where one of the categories is small enough for the resample of both sampling methods to be used. This study applied several machine learning methods, including the Decision Tree, Naïve Bayes, K-Nearest Neighbor (KNN), and Rotation Forest. The results show that using a resample of both samplings provides optimal results for the four machine-learning methods. If viewed from the indicators of accuracy, specificity, sensitivity, AUC, and the highest Kappa coefficient produced, the best method is the KNN method. The KNN model has an accuracy value of 0.73 percent, sensitivity of 0.68 percent, specificity of 78 percent, and AUC of 0.73.

Muñetón-Santa et al. (2022) introduced a methodology to classify poor and extremely poor people through Natural Language Processing (NLP). The approach utilizes machine learning algorithms to analyze the discourses of people affected by poverty and identify their level of deprivation. The paper proposes two strategies for document-level representations: 1) document-level features based on the concatenation of descriptive statistics and 2) Gaussian mixture models. The paper evaluates three classification methods: support vector machines, random forest, and extreme gradient boosting. The results show that the embeddings based on GloVe word vectors have the highest sensitivity of 79.6%, implying that they can correctly identify 79.6% of the poor and extremely poor people in the dataset. The paper suggests that this methodology can be useful for public policy makers to prioritize social programs for the most vulnerable groups.

Alsharkawi et al. (2021) attempt to understand the multidimensional poverty problem in Jordan by proposing an original machine learning approach to assess and monitor the poverty status of Jordanian households. The approach uses data from household expenditure and income surveys conducted since the early 2000s. The paper applies various machine learning classification models, with the LightGBM algorithm achieving the best performance with an 81% F1-Score1. In this case, the authors argue that this approach is accurate, inexpensive, and makes poverty identification cheaper and much closer to real-time. The paper also discusses data preprocessing and handling imbalanced data. The authors also suggest that investigating how machine learning can be used to monitor or evaluate the impact of social programs or interventions on poverty reduction should be conducted.

3. Multiclass Decision Forest and Its Applications

Decision Forest, commonly known as Random Forest, is an ensemble learning technique combining the predictions from multiple machine learning algorithms to make more accurate predictions than any individual model (Rokach, 2016). A Decision Forest is a collection of decision trees, each contributing to the final output, making its approach particularly effective for complex classification and regression tasks. The algorithm constructs multiple decision trees during the training phase in a Decision Forest. The forest generated is a product of random subsets of the training data and, in some cases, random subsets of features, which ensures diversity among the trees and contributes to the robustness of the model. Each tree in the forest votes when making predictions, and the most popular output is chosen as the final prediction. This technique generally outperforms single decision trees and other algorithms due to their ability to mitigate overfitting while maintaining high accuracy (Adnan et al., 2021).

Multiclass Decision Forest is an ensemble learning method for classification tasks involving more than two classes (Rokach, 2016; Sagi & Rokach, 2020). They extend the concept of Random Forest to handle complex, multi-class problems by combining multiple decision trees, each trained on a random subset of data and features. This approach offers several key advantages, including high accuracy, robustness to noise and outliers, effective handling of high-dimensional data, and the ability to capture complex feature interactions. The forest also provides feature importance measures and efficiently handle missing values (Rokach, 2016). The primary purposes of Multiclass Decision Forest include solving complex classification problems with improved accuracy and generalization, handling large and high-dimensional datasets, and providing insights through feature importance rankings. They are valued for their versatility across various domains, resistance to overfitting, and ability to work well without feature scaling (Sagi & Rokach, 2020). Additionally, the forest balances model performance and interpretability, making them suitable for applications where understanding the decision-making process is important. Their effectiveness in dealing with imbalanced datasets and non-linear relationships and the ability to parallelize computations make them popular in many real-world machine-learning scenarios. Multiclass Decision Forest find versatile applications across various fields, demonstrating their adaptability and effectiveness in healthcare, environmental science, marketing, and social sciences. In the context of household poverty classification, Multiclass Decision Forest can analyze complex and multidimensional data, providing nuanced insights into various poverty levels. They can efficiently process the large volume of socio-economic data, dealing with the variability and complexity inherent in poverty data.

Research Methodology

This section contains methodological steps for poverty classification, where households might be categorized into multiple poverty levels. The following steps are involved in classifying household poverty in Kut Bak District.

1. Population and Samples

The study utilized a sample of 302 households from a total population of 686, identified through a comprehensive, participatory process of locating and verifying impoverished households in the Kut Bak District. This process involved collaboration among representatives from households previously identified as below the poverty threshold, local community leaders, the Deputy District Officer, and local administrative organizations. The identification method aimed for 100% coverage within the district, with the resulting data recorded in the Practical Poverty Provincial Connext (PPPConnext) database system, a specialized repository for information on impoverished households. At this stage, classifying impoverished households into four distinct groups based on the poverty line involves a multifaceted analysis of economic data. The study used a questionnaire-based tool to examine economic indicators such as the target population's income, expenditure, debt, assets, and savings. The primary goals were to establish baseline data for the households, monitoring income growth, and observing changes within the identified groups over time. This approach allows comparisons against benchmarks set by the Office of the National Economic and Social Development Council and related poverty alleviation projects.

From a sample of 302 households, classification into four distinct groups was based on specified income criteria. Group 1 includes 97 households (32%) with incomes falling below the poverty line, defined as less than 2,763 baht per person per month. This categorization highlights those most in need of immediate economic support. Group 2, with 94 households (31%), features incomes slightly above the poverty line but still under 40% of the poorest segment, ranging between 2,764 and 5,347 baht per person per month. This group represents a slightly better economic standing yet is still considerably vulnerable. Group 3 consists of 21 households (7%), with income levels that surpass 40% of the poorest but remain below the median, between 5,348 and 6,532 baht per person per month, indicating a transition towards economic stability. Finally, Group 4 includes 90 households (30%), with incomes above the median, defined as more than 6,533 baht per person per month, signifying the most economically stable group among the sampled population.

Three main features contribute to the classification: 1) incomes (This feature covers various sources of household income, including but not limited to agricultural earnings, wage labor, and remittances); 2) expenses (household expenses offer a window into the living standards and economic pressures faced by the families); and 3) project participation (engagement in local projects, whether governmental or non-governmental, indicates a household's access to external support systems and community resources). These main features are further divided into sub-features, providing a layered and detailed perspective on each household's socio-economic status. The sub-features encompass a range of indicators from basic income and expenditure to involvement in specific community projects, painting a comprehensive picture of the multi-dimensional nature of poverty. The main features and their respective sub-features are outlined in Table 1 as follows.

Table 1 Features and sub-features used in research

Main Features	Sub-Features
Household Income	Annual Household Income (AHI)
	Agricultural Income (AGRI)
	Non-Tangible Income (NTI)
	Income from Land Holdings (ILH)
	Support and Assistance Income (SAI)
	Total Household Income (THI)
Household Expenditures	Household Consumption Expenditure (HCE)
	Utility Expenditure (UE)
	Essential Services Expenditure (Water, Electricity, Phone) (ESE)
	Educational Expenditure (EDUEX)
	Healthcare Expenditure (HEX)

	Insurance Premium Expenditure (IPE) Social Event Expenditure (SEX) Leisure and Entertainment Expenditure (LEX) Gambling Expenditure (GEX) Energy Drink Expenditure (EDEX) Travel Expenditure (TREX) Motorcycle Maintenance and Fuel Expenditure (MMFE) Car Maintenance and Fuel Expenditure (CMFE) Total Household Expenditure (THEX)
Project Participation	Vegetable Cultivation Project Participation (VCP) Vegetable Seedling Project Participation (VSPP) Community Welfare Project Participation (CWPP) Organic Fertilizer Project Participation (OFPP) Cost Reduction in Vegetable Cultivation Project (CRVCP) Cost Reduction in Fertilizer Usage Project (CRFUP) Cost Reduction in Insecticide Usage Project (CRIUP) Total Income from Project Participation (TIPP)

2. Collection of Data

Field surveys conducted by trained personnel and interviews with household members form the cornerstone of data collection in socio-economic research, particularly in studies like ours in Kut Bak District. Personnel conducting the surveys receive comprehensive training on the survey's objectives, the importance of ethical considerations, and how to accurately collect data. This research received ethical clearance from the ethics approval committee (full board review) of Rajabhat Sakon Nakhon University, under approval number HE 65-099. This approval was valid from August 31, 2022, to August 31, 2023. This training also ensures consistency and reliability in data collection. Field staff visited each household in person, allowing them to observe living conditions firsthand and gather detailed information. Efforts were made to speak with various household members to get a holistic view of the household's economic and social dynamics. Interviews were conducted with the utmost respect for confidentiality and sensitivity to the personal circumstances of the interviewees. The data collection spanned three months, commencing in February 2023 and concluding at the end of April 2023. This extensive period allowed researchers to gather and verify a comprehensive dataset, ensuring accuracy and depth in our analysis of household poverty levels within the Kut Bak District.

3. Research Instrument

Several critical steps were undertaken to prepare the dataset for effective multiclass decision forest algorithm use. Initially, data cleaning was conducted, which involved addressing missing values, removing duplicate entries, and correcting any errors in the dataset, ensuring the data's accuracy and reliability. Next, feature selection was carried out, where relevant features, such as income, expenses, and project participation, are identified based on their significant impact on poverty classification. This step is crucial for focusing the analysis on the most influential variables. Following this, data transformation was applied to normalize or scale the data, ensuring consistency across the dataset and mitigating issues arising from differing scales or distributions. Finally, encoding categorical data is vital, where non-numeric categories are converted into numeric formats. This conversion is essential for the algorithm to process and analyze the data effectively, as it typically handles numerical inputs. Each step is fundamental in preparing the data for accurate and efficient analysis using advanced machine learning techniques.

After that, the gathered dataset was meticulously divided into two segments: 70% for training and 30% for testing, to ensure robust model validation. This division is key to the effectiveness of the multiclass decision forest model. The dataset includes various features like income, expenses, and project participation, each contributing to the classification of poverty levels. The training set, comprising most of the data, is utilized to educate the model on the underlying patterns and nuances within the dataset. This training phase is crucial for the model to learn and interpret the socio-economic dynamics being studied accurately. The smaller, yet significant, 30% portion reserved for testing serves an invaluable purpose. It acts as an unbiased benchmark to evaluate the trained model's performance, testing its ability to predict and classify data not previously encountered accurately. This split not only safeguards against overfitting, where the model might excel in training but fail in a real-world application but also ensures that the model's predictions are reliably generalizable.

4. Data Analysis and Classification

The next crucial step was to analyze the data once the data is split into a 70% training set and a 30% testing set. Utilizing the 70% training set, the model was trained to recognize and understand the patterns and relationships within the data. This involves feeding the data into the Multiclass Decision Forest algorithm, allowing it to learn from the diverse range of features and their corresponding impact on poverty levels. The ability to handle a mix of quantitative and qualitative data makes it particularly suitable for multidimensional poverty assessment, offering a more comprehensive and accurate picture of household poverty. The construction and evaluation of Multiclass Decision Forest involves the synergistic application of several critical mathematical formulations (Rokach, 2016; Sagi & Rokach, 2020). Impurity measures, such as Entropy or Gini Impurity, are employed to quantitatively assess the efficacy of potential data partitions. At the same time, Information Gain serves as a criterion for identifying the optimal bifurcation at each decision node. This algorithmic process is recursively applied to generate a multiplicity of decision trees, each cultivated on a bootstrap sample derived from the original dataset, thereby ensuring heterogeneity within the ensemble. A collective decision-making mechanism aggregates the individual tree outputs for classification tasks to determine the ultimate class assignment. The Out-of-Bag (OOB) error estimation technique provides an unbiased metric of the forest's generalization performance by utilizing samples excluded from the training of individual trees. Equations (1-6) provide a foundation for understanding and implementing Multiclass Decision Forest.

Entropy

$$H(S) = -\sum(p(i) * log2(p(i))) \quad (1)$$

Entropy is a measure of impurity or uncertainty in a dataset. In the context of decision trees, where: 1) *S* is the set of samples at a node; 2) *p(i)* is the proportion of samples belonging to class *i*; 3) The sum is taken over all classes. A lower entropy indicates a purer node (more samples of the same class). Perfect purity (all samples of one class) gives an entropy of 0, while maximum impurity (equal distribution among classes) gives the highest entropy. In our analysis, entropy quantifies the uncertainty in household poverty levels before a split in our decision tree model, providing a baseline to assess the reduction of uncertainty after applying specific socio-economic attributes as splitting criteria, thus enhancing the precision of poverty classification in the Kut Bak District.

Information Gain

$$IG(S,A) = H(S) - \Sigma((|Sv| / |S|) * H(Sv)) \quad (2)$$

Information Gain measures how much a split on attribute A reduces the entropy, where: 1) *H(S)* is the entropy of the parent node; 2) *Sv* is a subset of *S* where attribute *A* has value *v*; 3) |*Sv*| / |*S*| is the proportion of samples in subset *Sv*; 4) *H(Sv)* is the entropy of subset *Sv*; 5) The sum is taken over all possible values *v* of attribute A. The attribute with the highest information gain is chosen for splitting at each node. We employed information gain to evaluate the effectiveness of different socioeconomic attributes, such as income or project participation, in reducing entropy, thereby determining the most informative features that best segregate households into distinct poverty categories.

Gini Impurity

$$Gini(S) = 1 - \sum(p(i)\hat{}2) \quad (3)$$

Gini Impurity is an alternative to Entropy for measuring node impurity, where: 1) *p(i)* is the proportion of samples belonging to class *i*; 2) The sum is taken over all classes. Similar to Entropy, a lower Gini Impurity indicates a purer node. It ranges from 0 (pure) to 1 - 1/*k* (most impure), where *k* is the number of classes. Gini impurity measures the frequency of misclassification within a node if randomly labeling data according to class distribution; this metric helped us optimize the splits in our Multiclass Decision Forest, ensuring more accurate poverty level predictions

Classification Error

$$Error(S) = 1 - max(p(i)) \quad (4)$$

This measures the misclassification rate if we were to classify all samples in a node as the majority class, where: 1) *p(i)* is the proportion of samples belonging to class *i*; 2) *max(p(i))* is the proportion of the most common class. This equation gauges the error rate of classifying all households in a node into the majority class, providing a straightforward assessment of potential misclassification before refining the decision tree splits, vital for enhancing model accuracy in delineating poverty levels.

Voting Mechanism

$$Class = argmax(\Sigma(I(h_t(x) = c))) \quad (5)$$

This formula is used to combine predictions from multiple trees in the forest, where: 1) $h_t(x)$ is the prediction of the *t-th* tree for input *x*; 2) *c* is a class label; 3) *I()* is the indicator function, returning 1 if the condition is true, 0 otherwise; 4) The sum is taken over all trees in the forest; 5) *argmax* returns the class with the highest vote count. By integrating the voting mechanism, our model aggregates predictions from various decision trees to determine the most likely poverty level for each household, leveraging collective decision-making to improve reliability and robustness in our results

Out-of-Bag (OOB)

$$OOB\_Error = 1/N * \Sigma(I(y_i \neq \hat{y}_i^{OOB})) \quad (6)$$

This estimates the generalization error of the forest without needing a separate test set, where: 1) *N* is the total number of samples; 2) y_i is the true label of sample *i*; 3) $\hat{y}_i^{OOB}$ is the *OOB* prediction for sample *i* (using only trees that did not use this sample in training); 4) *I()* is the indicator function; 5) The sum is taken over all samples. OOB error estimation offers a robust measure of model accuracy by using each tree's predictions on the data not included in its training set, enabling us to validate the generalization ability of our Multiclass Decision Forest without the need for a separate test set.

5. Data Validation

The model's performance was validated using a separate testing dataset to ensure its accuracy and reliability in classifying poverty. With the model trained and optimized, it was then tested using the 30% testing set. The testing phase involves running the model against this unseen data to evaluate how well it can predict poverty classifications, providing an objective measure of its effectiveness. This step was critical for assessing the model's performance and accuracy. Data validation for the model involved a comprehensive approach, utilizing six key metrics to assess its performance accurately, including overall accuracy, average accuracy, micro-averaged precision, macro-averaged precision, micro-averaged recall, and macro-averaged recall. Each metric provides a unique perspective on the model's effectiveness in classifying household poverty. After these metrics were evaluated, a confusion matrix was used for further validation. This matrix is a table used to describe the performance of the classification model on a set of test data for which the true values are known. It allows for visualizing the model's performance, showing true positives, false positives, true negatives, and false negatives. This detailed breakdown helps understand the model's strengths and weaknesses in predicting each class. The comprehensive validation process thoroughly assesses the model's ability to accurately classify households into different poverty levels, ensuring its reliability and effectiveness for practical application.

6. Analysis of Results

The final step in the analysis phase involved a comprehensive evaluation of the model's performance and extends to the derivation of rule-based definitions for the four levels of household poverty. This multifaceted approach ensures a more profound understanding and practical application of the findings. The analysis begins with thoroughly interpreting metrics like accuracy, precision, recall, and the confusion matrix. These metrics provide a quantitative measure of the model's ability to accurately classify households into different poverty levels. The results were contextualized within the socio-economic environment of Kut Bak District, and, where possible, compared with existing models or studies. This comparison helps identify unique patterns or deviations captured by the current model. Significant patterns and trends were identified by closely examining the model's output. This might involve recognizing which features (like income, expenses, or project participation) most indicate each poverty level. In addition, rule-based definitions for the four poverty levels were derived based on the insights gained. These rules were formulated by analyzing the combination of features and their thresholds that most frequently lead to a particular classification. For example, a rule might state that households with incomes below a certain level and high expenses are classified as 1 (extremely poor). The derived rules and model analysis are then translated into actionable insights.

Results

In our study, we implemented sophisticated data mining techniques to classify data, specifically focusing on utilizing the Multiclass Decision Forest algorithm. To ensure the robustness and accuracy of our analysis, we adopted a systematic approach to dividing the data: 70% is allocated as the training dataset, while the remaining 30% was used for testing. This division was accomplished using the split data method, a standard technique in machine learning that helps validate the model's performance on unseen data. As we examined each model, our primary objective was to explore and identify specific rules that can accurately indicate the household poverty level, ranging from level 1 (extremely poor) to level 4 (above poor). This involved a detailed examination of the models' outputs, assessing how socio-economic variables like income, expenses, and project participation interact and contribute to the poverty classification. By doing so, we aimed to extract meaningful patterns and rules that enhanced our understanding of the different levels of poverty and provide actionable insights for targeted interventions. These insights were crucial for developing effective poverty alleviation strategies, as they offered a granular view of the varying degrees of poverty within the community. This comprehensive analysis aimed to equip policymakers and social workers with the necessary tools and knowledge to address poverty more effectively in the Kut Bak District and similar regions.

The performance metrics of the model demonstrated its robustness and accuracy in classifying households into various poverty levels. With an overall accuracy of 94.4%, the model proved highly reliable in general predictions across all classes, suggesting its effectiveness in practical applications. The average accuracy, at an even higher 97.2%, indicated exceptional consistency and reliability of the model in predicting poverty levels across different classes, ensuring equitable performance. The micro-averaged and macro-averaged precision scores, at 94.4% and 95.3% respectively, highlighted the model's precision in correctly identifying poor households while minimizing false positives, ensuring that non-poor households were seldom misclassified. Similarly, the micro-averaged and macro-averaged recall scores of 94.4% and 92.9% underscored the model's capability to capture a high percentage of actual poor households across all categories. These high recall scores indicated the model's efficiency in identifying true positive cases, crucial for poverty alleviation efforts. Overall, these metrics collectively affirmed the model's balanced and accurate performance in poverty classification, making it a valuable tool for policymakers and social program implementers to effectively target and assist the appropriate households.

Furthermore, in the confusion matrix analysis for classifying the four levels of household poverty, the results demonstrated a notable degree of accuracy across all groups, with outstanding performance in identifying the extremes of the poverty spectrum. Groups 1 and 4, representing the ‘extremely poor’ and ‘above poor’ categories respectively, were classified with remarkable precision, each achieving a 100% accuracy rate. This indicated that the model was exceptionally adept at identifying both the most vulnerable households (Group 1) and those that were relatively more economically stable (Group 4). On the other hand, Groups 2 and 3, which represent intermediate poverty levels, showed slightly lower accuracy rates, with Group 2 (poor) being correctly identified 91.7% of the time and Group 3 (rather poor) at 80%. While these figures are still high, they suggest that the model faces more challenges in distinguishing between the nuances of the middle poverty levels. This discrepancy may be attributed to the overlapping characteristics and subtleties present in the economic conditions of these groups. To enhance the model’s effectiveness, further refinement and analysis should focus on these middle groups, possibly by incorporating more detailed socio-economic factors or adjusting the classification thresholds to better capture the subtle differences between moderately and marginally poor households. This refinement is crucial for ensuring that interventions and resources are accurately targeted, particularly for households on different poverty levels. Figure 2 presents the model's performance, evaluated using six key metrics and a confusion matrix, providing insights into its strengths and weaknesses.

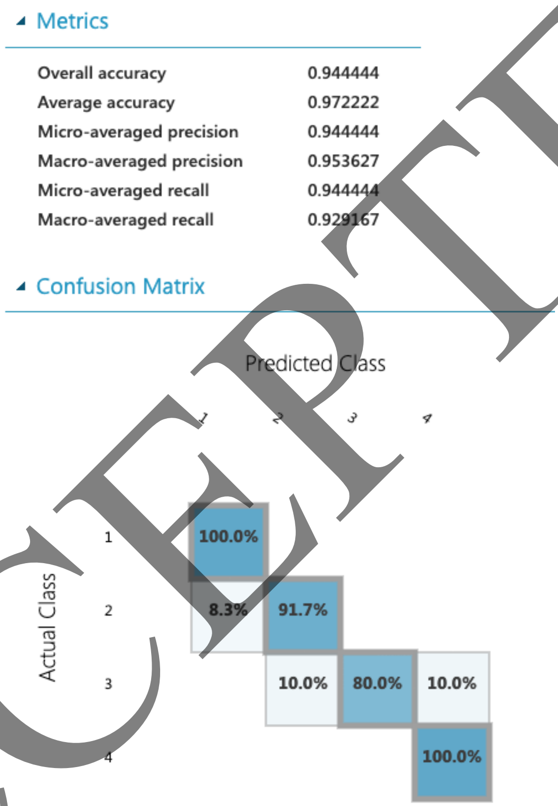

Figure 2 Model performance visualized using 6 metrics and confusion matrix

The model effectively classified four household poverty levels, utilizing 19 features, and achieved an impressive 100% accuracy rate. This exceptional accuracy was particularly evident for Groups 1 and 4, as indicated by the confusion matrix results. Consequently, four specific rules were selected for accurately identifying Group 1, while six distinct rules were chosen to classify Group 4, ensuring precise and reliable categorization of these groups. Five distinct rules were established for classifying households into poverty Group 2, while four specific rules were formulated for classifying households into poverty Group 4. Figure 3 depicts the classification of all 19 instances across Groups 1 to 4, each being accurately categorized with a 100% accuracy rate.

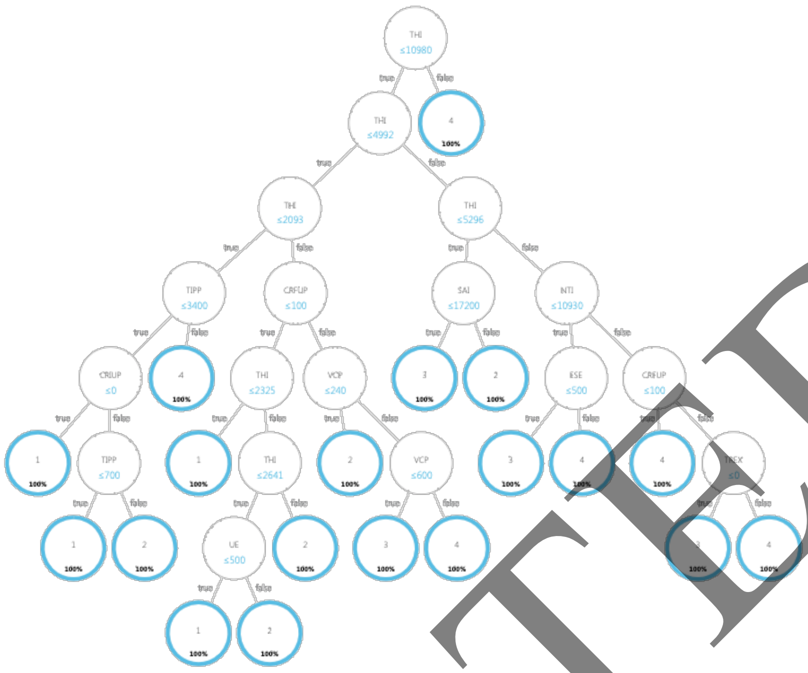

Figure 3 Classification of all 19 instances from Groups 1 to 4 with a 100% accuracy rate

1. Household Poverty 1

We delineated several specific rules that define the criteria for a household to be classified into poverty Group 1, which represents the 'extremely poor' category. These rules were derived from various socio-economic features and their respective thresholds. Table 2 explains each rule.

Table 2 Classification Rules, Syntax, and Descriptions for household poverty Group 1

Rules	**Syntaxes**	**Descriptions**
Low Total Household Income and Low Project Participation without Reduced Insecticide Costs	(THI <= 2093) = True, (TIPP <= 3400) = True, (CRIUP <= 0) = True	If total household income is less than or equal to 2093 baht/person/month, and total income from project participation is less than or equal to 3400 baht/person/month, and the household does not participate in reduced insecticide usage projects (CRIUP <= 0), then the household is classified as 'extremely poor'.
Low Total Household Income and Very Low Project Participation	(THI <= 2093) = True, (TIPP <= 3400) = True, (CRIUP <= 0) = False, (TIPP <= 700) = True	If total household income is less than or equal to 2093 baht/person/month, and total income from project participation is less than or equal to 3400 baht/person/month. The household participates in reduced insecticide usage projects, but the total income from project participation is very low, not exceeding 700 baht/person/month. The household is also classified as 'extremely poor'.
Slightly Higher Income Threshold	(THI <= 2325) = True	If total household income is less than or equal to 2325 baht/person/month, irrespective of other factors, then it falls under the 'extremely poor' category.
Low Income and Low Utility Expenditure	(THI <= 2641) = True, (UE <= 500) = True	If total household income is less than or equal to 2641 baht/person/month, and utility expenditure is less than or equal to 500 baht/person/month, this combination also places the household in the 'extremely poor' category.

2. Household Poverty 2

A series of rules were defined to categorize households into poverty Group 2. The rules for poverty Group 2 provide a detailed understanding of households in this category. They reflect a financial status that, while not as dire as the lowest poverty level, still faces significant economic limitations, highlighting the need for nuanced poverty alleviation efforts tailored to this group. Table 3 is a breakdown and explanation of each rule.

Table 3 Classification rules, syntax, and descriptions for household poverty Group 2

Rules	Syntaxes	Descriptions
Moderate Income with Specific Project Participation and Expenditure Patterns	(THI <= 2093) = True, (TIPP <= 3400) = True, (CRIUP <= 0) = False, (TIPP <= 700) = False	If total household income is less than or equal to 2093 baht/person/month and total income from project participation is also less than or equal to 3400 baht/person/month, but the household participates in reduced insecticide usage projects, and total income from project participation is more than 700 baht/person/month, these conditions classify the household in poverty group 2, indicating a slightly better but still limited economic situation.
Income Threshold Exclusion	(THI <= 2641) = False	If total household income exceeds 2641 baht/person/month, the household is not considered part of poverty group 2, suggesting a higher economic status.
Higher Income with Moderate Utility Expenditure	THI <= 2861) = True, (UE <= 500) = False	If total household income is less than or equal to 2861 baht/person/month, but utility expenditure exceeds 500 baht/person/month, this combination places the household in poverty group 2, reflecting moderate economic challenges.
Varied Income and Limited Fertilizer and Cultivation Project Participation	(THI <= 2093) = False, (CRFUP <= 100) = False, (VCP <= 240) = True	If total household income is higher than 2093 baht/person/month, participation in cost reduction in fertilizer usage project is minimal, and the household participates in the vegetable cultivation project (VCP <= 240 baht/person/month), these conditions collectively categorize the household in poverty group 2.
Higher Income with Limited Support and Assistance Income	(THI <= 5296) = True, (SAI <= 17200) = False	If total household income is less than or equal to 5296 baht/person/month, but support and assistance income (SAI) is less than 17200 baht/person/month, this suggests placement in poverty group 2, denoting a modest but constrained financial condition.

3. Household Poverty 3

Specific rules were established to identify households in poverty Group 3. This group represents those who are neither at the extreme end of poverty nor comfortably above it, but rather in a marginally poor or intermediate state. The rules for poverty Group 3 demonstrate the delicate balance of economic factors that define this category. Households in this group are not in the direct economic condition but still face significant challenges and limitations. Understanding the specific characteristics of this group is crucial for developing targeted social programs and interventions that effectively address their unique needs and help improve their socio-economic status. Table 4 represents rules, syntax, and descriptions of household poverty classification for Group 3.

Table 4 Classification rules, syntax, and descriptions for household poverty Group 3

Rules	Syntaxes	Descriptions
Moderate Income with Specific Range of Cultivation Project Participation	(THI <= 2093) = False, (CRFUP <= 100) = False, (VCP <= 240) = False, (VCP <= 600) = True	If the total household income is more than 2,093 baht/person/month, suggesting a moderate income level, and the household does not participate much in cost reduction in fertilizer usage project, with limited vegetable cultivation project participation (VCP > 240 baht/person/month), but not exceeding 600

		baht/person/month, these conditions place the household in poverty group 3, indicating a marginally poor economic status.
Higher Income with Limited Support and Assistance Income	(THI <= 5296) = True, (SAI <= 17200) = True	If total household income is less than or equal to 5,296 baht/person/month, indicating a higher but still limited income, and support and assistance income is also limited, not exceeding 17,200 baht/person/month, this combination categorizes the household in poverty group 3, reflecting some financial constraints.
Higher Income with Limited Non-Tangible Income and Essential Services Expenditure	(THI <= 5296) = False, (NTI <= 10930) = True, (ESE <= 500) = True	If total household income is less than or equal to 5,296 baht/person/month, but non-tangible income is less than or equal to 10,930 baht/person/month, and essential services expenditure is also limited (ESE <= 500 baht/person/month), these factors collectively identify the household as being in poverty group 3.
Moderate Income with Diverse Economic Indicators	(THI <= 5296) = False, (NTI <= 10930) = False, (CRFUP <= 100) = False, (TREX <= 0) = True	If total household income is less than or equal to 5,296 baht/person/month, indicating a moderate income, and non-tangible income exceeds 10,930 baht/person/month, and cost reduction in fertilizer usage project is minimal (CRFUP > 100 baht/person/month), but the household has no travel expenditures, this suggests the household belongs to poverty group 3.

4. Household Poverty 4

We established a set of rules that categorize households into povertyGroup 4, indicative of a relatively better economic status, or ‘above poor’. These rules were derived from analyzing various economic indicators and their interactions. Table 5 provides a detailed explanation of each rule for poverty Group 4.

Table 5 Classification rules, syntax, and descriptions for household poverty Group 4

Rules	Syntaxes	Descriptions
Higher Total Household Income	(THI <= 10980) = False	If total household income exceeds 10,980 baht/person/month, then the household is classified as ‘above poor’, indicating a higher economic status.
Low Income but Higher Project Participation	(THI <= 2093) = True, (TIPP <= 3400) = False	If total household income is less than or equal to 2,093 baht/person/month, but total income from project participation is greater than 3,400 baht/person/month, this suggests a significant contribution from project participation, placing the household in the ‘above poor’ category.
Moderate Income with Limited Project Participation	(THI <= 4992) = True, (CRFUP <= 100) = False, (VCP <= 600) = False	If total household income is less than or equal to 4,992 baht/person/month, and the participation in cost reduction in fertilizer usage project is minimal (> 100 baht/person/month), and the participation in vegetable cultivation project is also limited (> 600 baht/person/month), these conditions collectively define an ‘above poor’ household.
Sufficient Income with Non-Tangible Income and Moderate Essential Services Expenditure	(THI <= 5296) = False, (NTI <= 10930) = True, (ESE <= 500) = False	If total household income is less than or equal to 5,296 baht/person/month, and non-tangible income is less than or equal to 10,930 baht/person/month, and essential services expenditure is greater than 500 baht/person/month, this combination of factors indicates a status of ‘above poor’.

Adequate Income and Project Participation	(THI <= 5296) = False, (NTI <= 10930) = False, (CRFUP <= 100) = True	If total household income is less than or equal to 5,296 baht/person/month, and non-tangible income is greater than 10,930 baht/person/month. Still, the household participates in the cost reduction in fertilizer usage project (CRFUP <= 100 baht/person/month), the household is in the ‘above poor’ category.
Diverse Income Sources and Expenditures	(THI <= 5296) = False, (NTI <= 10930) = False, (CRFUP <= 100) = False, (TREX <= 0) = False	If total household income is less than or equal to 5,296 baht/person/month, and non-tangible income is higher than 10,930 baht/person/month, and cost reduction in fertilizer usage project is minimal (CRFUP > 100 baht/person/month), but travel expenditure is existent, this suggests a diverse range of income sources and expenditures, categorizing the household as ‘above poor’.

5. Guidelines for Identifying Household Poverty and Government Measures

The study utilized participatory forum approaches in Kut Bak District, engaging a diverse array of stakeholders from the Department of Provincial Administration, Ministry of Interior. These stakeholders included the district mayor, subdistrict headman, village headman, local government officials, representatives from educational institutions and agricultural offices, as well as members of the target groups. The forums were held twice, on December 14, 2023, and December 19, 2023. This multi-stakeholder engagement was meticulously designed to capture a wide range of perspectives on poverty-related issues. The research methodology was centered around horizon scanning, a strategic foresight tool employed for future scenario planning. This approach facilitated the comprehensive collection and analysis of environmental data, encompassing past, present, and anticipated near-future scenarios. The horizon scanning process was systematically structured into five key steps: (1) formulating critical questions, (2) determining timeframes and thematic frameworks, (3) data collection, (4) categorization and prioritization of issues, and (5) in-depth analysis of prioritized topics. The study concentrated on two primary research questions aimed at identifying the root causes and impacts of poverty in Kut Bak District over the past five years. The scanning framework encompassed four critical dimensions—economic, social, environmental, and governance—with data collection conducted through both online and offline channels to ensure comprehensive coverage.

Table 6 summarizes the rules defining household poverty from Groups 1 to 4, along with suggestions for government interventions. For each of these rules, tailored government interventions can significantly improve the living conditions of households classified as extremely poor, poor, rather poor, and above poor. These interventions should focus on immediate financial aid and long-term sustainability, including education, skill development trainings, infrastructure development, and community empowerment programs. By addressing the specific needs highlighted by each rule, the government can effectively target and mitigate the factors contributing to extreme poverty in these communities. These interventions should also foster economic resilience, enhance access to essential services, and promote long-term sustainable development. By focusing on education, skill development, and community engagement, the government can effectively support these households in their journey toward greater economic security and prosperity.

Table 6 Guidelines and Government Interventions for Household Poverty

Group	Guidelines	Government Interventions
1	Households with very low total income, limited project participation, and reduced or no insecticide usage are classified as ‘extremely poor’.	Increase access to financial aid and develop programs that enhance agricultural productivity and diversify income sources.
2	Households with moderate income, limited project participation, and specific expenditure patterns fall under group 2.	Enhance access to microfinance and skill development programs to boost income and diversify economic opportunities.
3	Households with higher income but limited project participation and specific expenditure patterns are in group 3.	Introduce employment support services and training programs to help these households achieve higher income stability and growth.
4	Households with higher income levels, diverse income sources, and moderate project participation are classified as ‘above poor’.	Encourage and support skill development and entrepreneurship programs to help these households enhance their income stability and growth potential.

Discussion

The primary objective of our research was to deploy the Multiclass Decision Forest model for accurately classifying households into distinct poverty levels using a range of socio-economic variables. The findings from our study demonstrated a high degree of accuracy, particularly for the extremely poor and above poor groups, which were identified with a 100% accuracy rate. This outstanding result underscores the efficacy of the Multiclass Decision Forest in handling multidimensional data, aligning well with our objective to enhance the precision of poverty classification. Our analysis showed that Total Household Income (THI) and Project Participation (TIPP) were critical in predicting poverty levels. The model's ability to differentiate effectively between different poverty statuses using these and other socio-economic indicators confirms that deploying advanced machine learning techniques can significantly improve the understanding and identification of poverty within varied contexts.

In the feature importance analysis phase, the algorithm plays a pivotal role in discerning which features are most critical in determining poverty levels. This process was integral to comprehending the primary factors driving poverty in the area. The classification of impoverished households into four distinct groups based on the poverty line involves a multifaceted analysis of economic data. The primary goals were to establish baseline data for these households, monitor income growth, and observe changes over time within the identified groups. This approach allowed for comparisons against benchmarks set by the Office of the National Economic and Social Development Council and related poverty alleviation projects. However, when employing a multiclass decision forest, the classification yielded slightly different household income levels due to the algorithm's comprehensive consideration of multiple features across three categories: household income, household expenditure, and project participation. This method offers a more precise representation of household income levels, capturing the complexity of economic conditions more effectively. This classification is achieved by evaluating a combination of influential features, enabling a nuanced understanding of the different degrees of poverty.

The application of machine learning in poverty classification has been explored in various studies with mixed results. For instance, research by Alsharkawi et al. (2021) utilized machine learning to classify poverty in Jordan, employing a similar approach but with different specific techniques, such as LightGBM, achieving a significant F1-Score of 81%. Our findings are consistent with this study regarding the high accuracy achieved through sophisticated machine learning models. However, our approach using Multiclass Decision Forests, which integrates multiple decision trees to handle complex classification tasks, provides even higher accuracy and robustness against overfitting, as indicated by our perfect classification results for specific groups. In contrast, a study by Huang & Xia (2023) focused on classifying rural relative poverty using a decision tree model based on an ID3 algorithm. While they reported good performance, the decision tree approach is often criticized for its tendency to overfit, especially with more extensive and less homogeneously distributed data. Our more complex ensemble method addresses this limitation, suggesting an advantage over traditional single decision tree approaches. The results also align with the broader findings in the field that advocate for using ensemble learning methods in socio-economic applications. For example, Rokach (2016) and Sagi & Rokach (2020) highlight the benefits of ensemble techniques, like Random Forests and Multiclass Decision Forests, for their superior performance over single-model approaches in various prediction tasks. As noted in our study and others, the consistent identification of socio-economic variables impacting poverty levels supports the notion that machine learning can provide nuanced insights into poverty dynamics, potentially informing more targeted and effective poverty alleviation strategies.

Moreover, common pool resources (CPRs) are natural or man-made resources where one person's use diminishes another person's ability to use the same resource, and where it is challenging to exclude anyone from utilizing the resource (Cooperman et al., 2022). Examples of CPRs include natural resources such as fisheries, groundwater basins, grazing lands, forests, and the atmosphere, as well as environmental systems like ecosystems, oceans, rivers, and lakes. In this study, CPRs constitute shared natural assets, with numerous features closely linked to these resources. For instance, agricultural income (AGRI) and income from land holdings (ILH) are directly tied to the use of common lands and water sources. Participation in vegetable cultivation and seedling projects (VCP, VSPP) further underscores the connection to CPRs, as these activities often depend on shared environmental assets. Additionally, involvement in community welfare, organic fertilizer, and cost reduction projects (CWPP, OFPP, CRVCP, CRFUP), along with participation in the cost reduction in insecticide usage project (CRIUP) and total income from project participation (TIPP), reflects the community's engagement with and reliance on these shared natural resources. This engagement highlights the critical role CPRs play in the livelihoods of the households studied, emphasizing the need for sustainable management and equitable access to these resources.

Suggestion

This research underscores the importance of leveraging advanced data analytics and machine learning techniques in social sciences, particularly in areas like poverty assessment where traditional methods may fall short in addressing the multi-dimensional nature of the issue. In particular, the application of the multiclass decision forest in poverty classification offers a comprehensive and nuanced understanding of poverty at the household level, enabling more effective and tailored poverty alleviation strategies. By continuing to refine these models and incorporating a broader range of data, future research can further enhance the accuracy and applicability of poverty classification, contributing to more effective poverty alleviation and common pool resource management strategies and socio-economic development. Below are examples of future research that can provide deeper insights into socioeconomic conditions and guide more effective interventions.

1) Conduct a comprehensive analysis of district educational attainment and vocational training opportunities. Such an investigation will identify gaps in knowledge and skills that, when addressed, could significantly enhance the employability and economic potential of the population.
2) Thoroughly examine healthcare services in Kut Bak District, focusing on accessibility and quality. Research into health disparities and the prevalence of chronic diseases will provide valuable insights for shaping health policies aimed at improving overall community health outcomes.
3) Given the district's dependence on agriculture, conducting an in-depth study of current farming practices, crop diversity, and agricultural productivity is imperative. This research should explore sustainable farming techniques and the role of technology in enhancing agricultural outputs.
4) Explore the potential for economic diversification within the district by assessing sectors such as local crafts, tourism, and small-scale manufacturing. These sectors could offer alternative income sources, contributing to the community's economic resilience.
5) Investigate the environmental implications of agricultural activities, focusing on the sustainability of natural resources. Key areas of study should include water resource management, soil health, and the effects of climate change on agricultural productivity.
6) Evaluate the impact and efficacy of existing government initiatives for poverty alleviation and economic support. This research will inform the refinement of current strategies and the development of new, more targeted programs to better meet the needs of residents in the Kut Bak District.

Conclusion

In conclusion, applying the Multiclass Decision Forest model in classifying household poverty levels has demonstrated considerable success, particularly in accurately identifying households at the extreme ends of the poverty spectrum. The model's remarkable 100% accuracy in classifying the 'extremely poor' (Group 1) and the 'above poor' (Group 4) indicates its effectiveness and reliability. However, the slightly lower accuracy rates for the intermediate groups - 91.7% for Group 2 (poor) and 80% for Group 3 (rather poor) - highlight areas for potential improvement, suggesting the need for more refined data analysis or the inclusion of additional socio-economic variables to better capture the complexities of these categories.

All findings, including the statistical analysis, pattern recognition, and derived rules for poverty classification, are meticulously documented. This comprehensive report is a critical resource for policymakers and other stakeholders in poverty alleviation and common pool resource management. The derived rules and model insights provide a foundation for future research and practical interventions in the field. They offer a structured approach to understanding and addressing household poverty, which can be adapted and applied in similar contexts. This study's findings are not only significant in terms of academic research but also hold substantial implications for policy-making and targeted intervention strategies. The high degree of accuracy in poverty classification enables policymakers and social workers to allocate and manage resources more efficiently and design programs tailored to different poverty groups' needs. For instance, while groups at the extreme ends may require more urgent and direct assistance, intermediate groups might benefit more from sustainable development programs, education, and skill-building initiatives.

Acknowledgement

We extend our heartfelt gratitude to the Program Management Unit on Area-Based Development (PMU A). This project could not have been completed successfully without their invaluable support and contributions.

Reference

Adnan, M. N., Ip, R. H. L., Bewong, M., & Islam, M. Z. (2021). BDF: A new decision forest algorithm. *Information Sciences*, *569*, 687-705.

Alkire, S., & Foster, J. (2011). Counting and multidimensional poverty measurement. *Journal of Public Economics, 95*(7-8), 476-487.

Alsharkawi, A., Al-Fetyani, M., Dawas, M., Saadeh, H., & Alyaman, M. (2021). Poverty classification using machine learning: The case of Jordan. *Sustainability, 13*(3), 3.

Carter, M. R., & Barrett, C. B. (2006). The economics of poverty traps and persistent poverty: An asset-based approach. *The Journal of Development Studies, 42*(2), 178-199.

Chakravarty, S. R., & Majumder, A. (2005). Measuring human poverty: A generalized index and an application using basic dimensions of life and some anthropometric indicators. *Journal of Human Development, 6*(3), 275-299.

Cooperman, A., McLarty, A. R., & Seim, B. (2022). Drivers of successful common-pool resource management: A conjoint experiment on groundwater management in Brazil. *Global Environmental Change*, *74*, 102512.

Gweshengwe, B., & Hassan, N. H. (2020). Defining the characteristics of poverty and their implications for poverty analysis. *Cogent Social Sciences, 6*(1), 1768669.

Huang, K., & Xia, F. (2023). Classification of rural relative poverty groups and measurement of the influence of land elements: A questionnaire-based analysis of 23 poor counties in China. *Land, 12*(4), Article 4.

Islam, M., Kashem, S., Momtaz, Z., & Hasan, M. M. (2023). An application of the participatory approach to develop an integrated water resources management (IWRM) system for the drought-affected region of Bangladesh. *Heliyon, 9*(3), e14260.

Kharas, H., & Dooley, M. (2022). *The evolution of global poverty*, *1990-2030*. Retrieved May 4, 2023, from https://www.brookings.edu/articles/the-evolution-of-global-poverty-1990-2030/

Lemanski, C. (2016). Poverty: Multiple perspectives and strategies. *Geography, 101*(1), 4-10.

Mah, J. C., Penwarden, J. L., Pott, H., Theou, O., & Andrew, M. K. (2023). Social vulnerability indices: A scoping review. *BMC Public Health, 23*(1), 1253.

Mansi, E., Hysa, E., Panait, M., & Voica, M. C. (2020). Poverty—A Challenge for economic development? Evidences from Western Balkan countries and the European Union. *Sustainability, 12*(18), 18.

Morris, M. H., Santos, S. C., & Neumeyer, X. (2018). *Poverty and entrepreneurship in developed economies*. Massachusetts: Edward Elgar Publishing.

Muñetón-Santa, G., Escobar-Grisales, D., López-Pabón, F. O., Pérez-Toro, P. A., & Orozco-Arroyave, J. R. (2022). Classification of poverty condition using natural language processing. *Social Indicators Research, 162*(3), 1413-1435.

Rokach, L. (2016). Decision forest: Twenty years of research. *Information Fusion*, *27*, 111-125.

Sagi, O., & Rokach, L. (2020). Explainable decision forest: Transforming a decision forest into an interpretable tree. *Information Fusion, 61*, 124-138.

Sihombing, P. R., & Arsani, A. M. (2021). Comparison of machine learning methods in classifying poverty in Indonesia in 2018. *Jurnal Teknik Informatika, 2*(1), 51-56.

Talingdan, J. A. (2019). Performance comparison of different classification algorithms for household poverty classification. In *Proceedings of the 2019 4th International Conference on Information Systems Engineering (ICISE)* (pp. 11–15). Shanghai: IEEE Computer Society.

Thorbecke, E. (2013). Multidimensional poverty: Conceptual and measurement issues. In N. Kakwani & J. Silber (Eds.), *The many dimensions of poverty*. London, UK: Palgrave Macmillan.

United Nations. (2023). *2023 Global Multidimensional Poverty Index (MPI)*. In *Human Development Reports.* Retrieved December 12, from https://hdr.undp.org/content/2023-global-multidimensional-poverty-index-mpi

Wikipedia. (2023). *Kut Bak district*. Retrieved May 4, 2023, from https://en.wikipedia.org/w/index.php?title=Kut_Bak_district&oldid=1151602823

Wong, S. Y. (2012). *Understanding poverty: Comparing basic needs approach and capability approach* (SSRN Scholarly Paper No. 2066179). Retrieved May 4, 2023, from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2066179

World Bank. (2023a). *World Bank East Asia and the Pacific economic update, October 2023: Services for development.* Washington, DC: World Bank.

World Bank. (2023b). World development report 2023: Migrants, refugees, and societies. Washington, DC: World Bank.

Article info

Received: 4 February 2024
Revised: 31 August 2024
Accepted: 14 January 2025

Author

Sudjarid

Email

p.jarupunphol@pkru.ac.th