International Journal of Data Science and Analysis
2025; 11(4): 114-124
http://www.sciencepublishinggroup.com/j/ijdsa
doi: 10.11648/j.ijdsa.20251104.12

Predicting Wealth Index Using an Ensemble Model of Random Forest and Multilayer Perceptron

Pinkie Akinyi Odipo*, Anthony Waititu, Herbert Imboga, Susan Mwelu

Department of Statistics and Actuarial Sciences, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya

Email address:

georgepinkie4@gmail.com (Pinkie Akinyi Odipo), awaititu@jkuat.ac.ke (Anthony Waititu), imbogaherbert@jkuat.ac.ke (Herbert Imboga), smwelu@jkuat.ac.ke (Susan Mwelu)
*Corresponding author

To cite this article:

Pinkie Akinyi Odipo, Anthony Waititu, Herbert Imboga, Susan Mwelu. (2025). Predicting Wealth Index Using an Ensemble Model of Random Forest and Multilayer Perceptron. *International Journal of Data Science and Analysis, 11*(4), 114-124.
https://doi.org/10.11648/j.ijdsa.20251104.12

Received: 30 July 2025; **Accepted:** 15 August 2025; **Published:** 12 September 2025

Abstract: Wealth inequality remains a significant challenge in Kenya, exacerbated by the limitations of traditional wealth measurement methods. This study develops and evaluates an ensemble wealth index model combining Random Forest (RF) and Multilayer Perceptron (MLP) algorithms to improve prediction accuracy using socio-economic data from the 2019 Kenya Population and Housing Census (KPHC). The models were assessed using performance metrics such as accuracy, precision, sensitivity, specificity and ROC-AUC. The Random Forest model, configured with 500 trees and a split variable count of 2, achieved 34.3% accuracy, 58.54% balanced accuracy, 65.98% out-of-bag error and performed best on the "Poorest" class with a 13.7% class error. It further recorded 41.13% precision, 34.27% recall, 83.17% specificity, and a 67.31% AUC. The MLP model, using sigmoid activation in hidden layers and softmax in the output layer, achieved 33.4% accuracy, 57.7% balanced accuracy, 30.6% precision, 32.54% recall, 82.86% specificity and AUC of 69.12%. The RF-MLP ensemble model outperformed the individual models with a 34.4% accuracy, 37.42% precision, 34.05% recall, 83.2% specificity and AUC of 68.55%. Despite modest overall accuracies, the ensemble model showed enhanced balanced accuracy and specificity, particularly in extreme wealth categories. However, classification of middle and poor wealth levels remains challenging due to feature overlap and class imbalance.

Keywords: Wealth Index, Ensemble Model, Random Forest, Multilayer Perceptron, Machine Learning

1. Introduction

Wealth inequality in Kenya continues to impede economic growth and social equity. Traditional models used for computing wealth indices are rather inadequate for capturing complex and non-linear interactions among socio-economic variables. With the growing availability of national datasets like the KPHC 2019, machine learning models present an opportunity to redefine how wealth levels are estimated and predicted.

This study introduces an ensemble model integrating Random Forest (RF) and Multi-Layer Perceptron (MLP) algorithms to predict wealth quintiles in Kenya. We focus on improving model performance and extracting meaningful patterns from census data to inform policy and social planning.

1.1. Wealth Inequality and Economic Growth

Wealth inequality is an important socio-economic issue. It has a role in shaping societal capacity and potential for unrest [19], while historical wealth inequality often leads to redistribution crises [18]. Wealth inequality also correlates with weak institutions and poor social outcomes [11, 12].

Debate exists over whether inequality hinders or promotes growth. Some argue that high-income earners invest more, boosting growth [2], while others contend that inequality restricts human capital investment and stifles economic progress [4].

Wealth inequality's deeper influence over income inequality

due to its long-term accumulation is emphasized [15, 16]. Yet, empirical studies remain limited and often overlook recent or context-specific data, like that from Kenya.

Further, recent studies emphasize the need for robust estimation techniques in national surveys, particularly in contexts with non-response or coverage errors. Researchers' work on ratio-based estimators provides insight into improving accuracy when dealing with household-level socio-economic data issues directly relevant to the construction of a reliable wealth index in this study [13].

1.2. Machine Learning for Socioeconomic Prediction

Machine learning enables data-driven insights without explicit programming. It integrates modeling, optimization and data mining to analyze patterns in large datasets [5, 7]. Its applications span education, health, business, and public policy [3].

Among learning methods, supervised learning is relevant for wealth classification. Algorithms such as decision trees, support vector machines, and neural networks are frequently applied. In this study, we focus on two key models:

Random Forest (RF): An ensemble of decision trees that uses bagging and random feature selection to improve performance and avoid overfitting [6]. RF offers robust handling of non-linear data and is suitable for identifying variable importance.

Multilayer Perceptron (MLP): A type of artificial neural network with hidden layers that capture complex non-linear relationships. It uses backpropagation for training and can approximate any continuous function [9].

The successful application of supervised machine learning models, particularly Random Forests, in classifying social behavior outcomes within Kenyan university settings has been demonstrated [10]. This supports the viability of using such models for classification tasks involving socio-economic data.

1.3. Ensemble Learning Approaches

Ensemble models combine multiple learners to achieve better predictive performance. Techniques like bagging, boosting, and stacking mitigate individual model weaknesses [17]. Ensemble models also outperform single classifiers in diverse domains [1, 8].

A similar hybrid modeling approach was employed in forecasting insurance claims, where combining classification and regression techniques led to improved model performance. This underscores the importance of ensemble strategies for complex data contexts like socio-economic profiling in Kenya [14].

This study applies a weighted ensemble of RF and MLP. RF contributes its strength in variable importance and handling limited data, while MLP captures deeper non-linear relationships. Together, they offer a more reliable framework for wealth prediction.

1.4. Research Gap

Despite progress in applying machine learning to socio-economic analysis, few studies focus on ensemble methods for wealth index prediction in Kenya. Moreover, the 2019 KPHC remains underutilized in predictive modeling. This study addresses these gaps by applying an ensemble RF-MLP model tailored to Kenya's socio-economic context. This study builds on previous Kenyan-based modeling research to tailor machine learning solutions for wealth estimation within a national socio-economic context.

2. Methodology

This section outlined the data sources, methodological approach, and analytical techniques used to develop the ensemble wealth index model. The study utilized the 2019 KPHC dataset, applying an ensemble model for improved accuracy. It included data acquisition, preprocessing, feature selection, model training, and performance evaluation. Each of these steps is discussed in detail to ensure transparency and reproducibility.

2.1. Study Data

The dataset used in this study was obtained from the KPHC 2019 data, acquired from the Kenya National Bureau of Statistics (KNBS) official data portal. The data contains a wide range of socio-economic indicators collected through national household surveys, capturing demographic and economic characteristics of Kenyan households.

Key variables used in wealth classification included household assets, dwelling type, access to utilities, education level, employment status, and geographic location. The dataset underwent cleaning, feature selection, and encoding to prepare it for machine learning applications. Prior to model training, the dataset was examined for missing values, outliers, and obvious inconsistencies in socio-economic variables. No extreme anomalies were detected that warranted removal, and minor irregularities were addressed through standard data cleaning procedures such as coding corrections and uniform formatting.

KPHC employed a two-stage stratified sampling method. In the first stage, Enumeration Areas (EAs) were stratified by region and urban/rural designation. In the second stage, households were systematically selected within each EA.

2.2. Study Variables

The study utilizes both dependent and independent variables. The dependent variable is the wealth index, categorized into five levels: Poorest, Poor, Middle, Rich, and Richest. Table 1 presents the key variables:

2.3. Random Forest Model

Random Forest is a supervised learning algorithm that belongs to the family of ensemble techniques. It creates

multiple decision trees where each tree is trained on a bootstrap sample and a random subset of features. The final prediction was made through majority voting (classification).

For n decision trees, the output was computed as:

$$Y_{RF} = \frac{1}{n}\sum_{i=1}^{n} f_i(x) \tag{1}$$

Random Forest reduced overfitting and allowed for feature importance estimation, enhancing interpretability.

2.4. Multilayer Perceptron Model

The MLP consists of multiple layers of perceptrons as shown in Figure 1. It includes input, hidden, and output layers, where each node was fully connected. The network used sigmoid activation in hidden layers and softmax in the output layer for multi-class classification.

Table 1. *Key dependent and independent variables used in the wealth index model. The table specifies the outcome variable (wealth index) and the socio-economic and demographic predictors, along with their definitions.*

Determinants	**Definition**
Dependent Variable	
Wealth Index	Poorest, Poor, Middle, Rich, Richest.
Independent Variables	
Region/County	Administrative regions.
Type of Residence (EA_Type)	Urban or rural classification
Education Level	Never Attended, Primary, Secondary, Other.
Sex of Household Head	Male or Female
Economic Activity	Employment status: Working, Unemployed
Number of Household Members	Continuous variable
Age of Household Head	Continuous variable

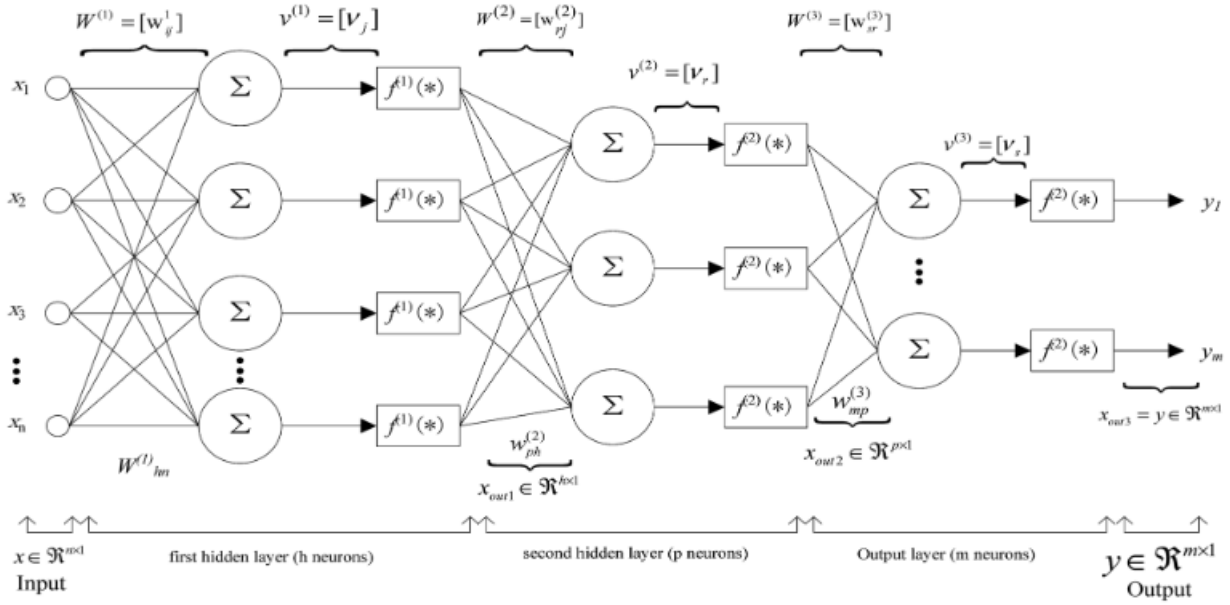

Figure 1. *Architecture of a feedforward three-layer perceptron. The diagram shows an input layer representing the model features, two fully connected hidden layers with sigmoid activation functions, and an output layer using a softmax activation function for multi-classification. Source:https://pubs.rsc.org/en/content/articlelanding/2017/ra/c7ra04147k.*

For input vector $X = (x_1, ..., x_d)$, the hidden node input and output is:

$$\upsilon_h(X;\theta) = W_{h0} + \sum_{j=1}^{d} W_{hj}x_j \tag{2}$$

The output $\phi_h(X;\theta)$ of the h^{th} hidden node in the first layer is:

$$\phi_h(X;\theta) = \psi(\upsilon_h(X;\theta)) \tag{3}$$

For hidden layers beyond the first, the input to the h^{th} hidden node in layer l is computed as:

$$\upsilon_h^{(l)}(X;\theta) = W_{h0}^{(l)} + \sum_{i=1}^{H} W_{hi}^{(l)}\,\phi_i^{(l-1)}(X;\theta) \tag{4}$$

Where:

1. $W_{h0}^{(l)}$: Bias for the h^{th} node in layer l
2. $W_{hi}^{(l)}$: Weight connecting the i^{th} node in layer $l-1$ to the h^{th} node in layer l
3. $\phi_i^{(l-1)}(X;\theta)$: Output of the i^{th} node in layer $l-1$

The output is:

$$\phi_h^{(l)}(X;\theta) = \psi(\upsilon_h^{(l)}(X;\theta)) \tag{5}$$

The net input to the output node is:

$$O_H(X;\theta) = \alpha_0 + \sum_{h=1}^{H} \alpha_h \, \phi_h^{(L)}(X;\theta) \tag{6}$$

The final output of the network is:

$$Y_{\text{MLP}} = \psi(O_H(X;\theta)) \tag{7}$$

The Activation Functions; Softmax was used in the output layer for multi-class classification while Sigmoid was used in hidden layers:

$$\psi(x) = \frac{1}{1+e^{-x}} \tag{8}$$

2.5. Ensemble Model Building

The RF and MLP models are trained separately and then combined using a weighted sum:

$$Y_{\text{RF-MLP}} = \alpha_{RF} \cdot Y_{RF} + \alpha_{MLP} \cdot Y_{MLP} \tag{9}$$

The process of weight determination involved combining subjective and objective weighting methods to enhance the performance of ensemble learning models. The weights of Random Forest (RF) and Multilayer Perceptron (MLP) are determined through a combination weighting approach that integrates both statistical methods as shown in Figure 2.

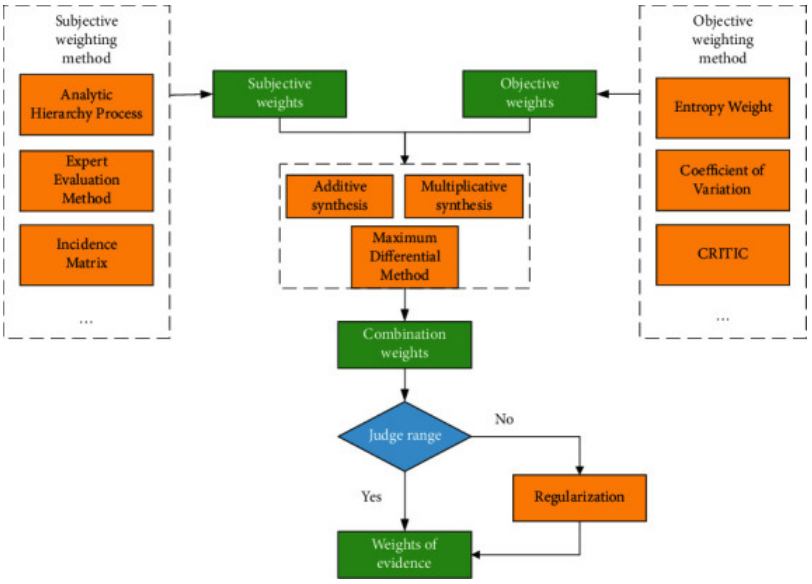

Figure 2. Process of weight of evidence assignment. The diagram illustrates how subjective weighting (e.g., Analytic Hierarchy Process) and objective weighting (e.g., entropy weight and coefficient of variation methods) are integrated to determine the final weights for Random Forest (RF) and Multilayer Perceptron (MLP) components. Source: https://onlinelibrary.wiley.com/doi/10.1155/2022/1156748

The combination weighting method integrates the subjective and objective weights by either Additive Synthesis, Multiplicative Synthesis, or Level Difference Maximization. In our study, we utilized *Additive Synthesis*, which computed the final weight of each model as follows:

$$\begin{aligned} \alpha_{\text{RF}} &= \underbrace{\omega_{\text{RF}}}_{\text{subj}} + \underbrace{\omega_{\text{RF}}}_{\text{obj}} \\ \alpha_{\text{MLP}} &= \underbrace{\omega_{\text{MLP}}}_{\text{subj}} + \underbrace{\omega_{\text{MLP}}}_{\text{obj}} \end{aligned} \tag{10}$$

The weights are then normalized to ensure they sum up to 1. For our study, we initially set:

$$\alpha_{\text{RF}} = 0.55, \quad \alpha_{\text{MLP}} = 0.45$$

based on preliminary testing indicating better performance of RF with structured data.These models are trained separately using the dataset, and their outputs are combined through a weighted averaging process determined by the calculated weights as shown in Figure 3.

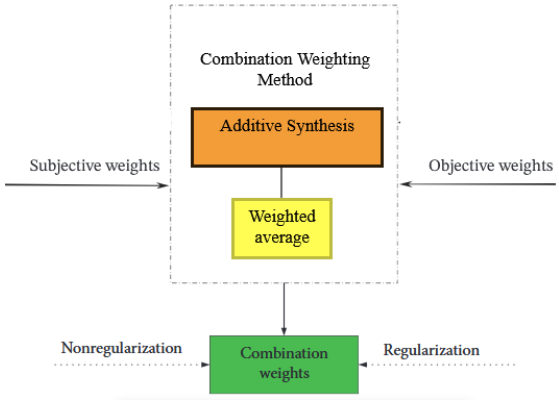

Figure 3. Process of weight combination in the RF-MLP ensemble model. The diagram shows how the Random Forest (RF) and Multilayer Perceptron (MLP) models are trained separately on the dataset, and their outputs are then combined using a weighted averaging approach based on the calculated subjective and objective weights.

The final prediction is then computed as:

$$Y_{\text{RF-MLP}} = \alpha_{\text{RF}} \times Y_{\text{RF}} + \alpha_{\text{MLP}} \times Y_{\text{MLP}} \quad (11)$$

This approach ensured that the strengths of both models are effectively utilized in making accurate predictions. Figure 4 is a visual representation of the RF-MLP Ensemble Model.

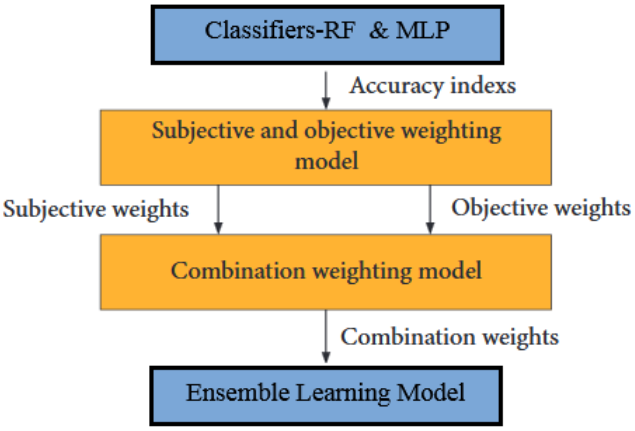

Figure 4. *RF-MLP ensemble learning model with combination weighting. The diagram illustrates the parallel training of the Random Forest (RF) and Multilayer Perceptron (MLP) models, the calculation of their subjective and objective weights, and the final combination of outputs to produce the ensemble prediction.*

2.6. Model Comparison

Model performance is evaluated using metrics in Table 2 such as accuracy, precision, recall, specificity, and ROC-AUC. Performance metrics are the main criteria for evaluating the performance of the model. The first step in performance evaluation is to select the appropriate data. The normal method is the holdout technique where the dataset is divided into training and test cases. Resampling techniques such as n-fold cross-validation is used to test the reliability of the model.

Table 2. *Performance metrics used to evaluate and compare the models. The table lists each metric, its formula, and its purpose in assessing different aspects of predictive performance, including accuracy, precision, recall, specificity, and ROC-AUC.*

Metric	Formula	Purpose
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$	Overall correctness
Precision	$\frac{TP}{TP+FP}$	Correctness of positive predictions
Recall (Sensitivity)	$\frac{TP}{TP+FN}$	Ability to detect positives
Specificity	$\frac{TN}{TN+FP}$	Ability to detect negatives
ROC-AUC	Area under ROC	Discriminative ability

ROC curve assessed the trade-off between sensitivity and specificity. The AUC value summarized model performance: AUC $\approx$ 1.0: Excellent performance and AUC = 0.5: No discriminative ability.

3. Results

3.1. Wealth Quintile Distribution

The distribution of the wealth index across the five quintiles was notably uneven. The majority of individuals belonged to the lower three quintiles, with the "Poorest" category having 3,870 individuals and the "Richest" only 1,705. This imbalance suggested a dataset skewed towards lower income groups, likely influencing the model's ability to learn uniformly across all classes.

3.2. Demographic Characteristics

Age: The mean age was 47 years, and the median was 49, indicated a right-skewed distribution with a larger number of older individuals.
Sex: The dataset was nearly gender-balanced, comprising 8,298 males and 7,220 females.
Education: Primary education was most common (9,066 individuals), followed by secondary education (4,140). Few individuals had no formal or tertiary education.

3.3. Household and Economic Features

Household Size: Household size ranged from 1 to 182 members, with a mean of 136.7 and a median of 153. The right-skewed distribution reflected a few extremely small households pulling down the mean.
Economic Activity: Economic activities were highly concentrated. The most dominant category had 9,648 individuals, and the next had 2,677. Remaining categories were sparsely populated.

3.4. Random Forest Performance

The Random Forest model, configured with 500 trees and a split variable count of 2, demonstrated modest predictive performance. It yielded an Out-of-Bag (OOB) error of 65.98%, indicating limited generalization capability. As shown in Table 3 , the model achieved an overall accuracy of 34.3%, a balanced accuracy of 58.54%, and a Cohen's Kappa coefficient of 0.162. While it performed relatively well on the "Poorest" class with a low error rate of 13.7%, it exhibited significant challenges in accurately classifying the "Poor," "Middle," and "Richest" categories.

Table 3. *Class-wise evaluation metrics for the Random Forest (RF) model. The table reports sensitivity, specificity, precision (PPV), and balanced accuracy for each wealth category, illustrating the model's performance across the different classes.*

Class	Sensitivity	Specificity	Precision	Balanced Accuracy
Poorest	87.34%	44.03%	34.16%	65.68%
Poor	0.00%	100.00%	0.00%	50.00%
Middle	0.14%	99.96%	50.00%	50.05%
Rich	66.20%	74.07%	33.07%	70.13%
Richest	15.84%	97.83%	47.37%	56.83%

Table 4. *Confusion matrix for the Random Forest (RF) model. The table compares the actual versus predicted class labels for the five wealth categories: Poorest, Poor, Middle, Rich, and Richest. Each cell represents the number of instances classified into a given category, with the last column showing the class-specific error rate.*

Actual/Predicted	Poorest	Poor	Middle	Rich	Richest	Error
Poorest	2672	0	1	398	25	13.70%
Poor	2455	0	2	535	43	100.00%
Middle	1917	0	1	916	74	99.97%
Rich	607	0	4	1290	112	35.92%
Richest	137	0	2	964	261	80.87%

Table 4 which is the condusion matrix, proved that the model performed relatively well in identifying the "Poorest" class (error rate ≈ 14%) and "Rich" class (≈ 36%). However, it failed to correctly classify any instances of the "Poor" and "Middle" classes, with nearly all samples being misclassified. The "Richest" class also had a high error rate of approximately 81%.

The variable importance in Figure 5 revealed that economic main job is the most influential predictor of wealth index, with a significantly higher importance score compared to other variables. This suggests that an individual's main source of employment plays a critical role in determining household wealth levels. Age follows as a distant second, indicating a moderate influence, while other variables contribute minimally to the model.

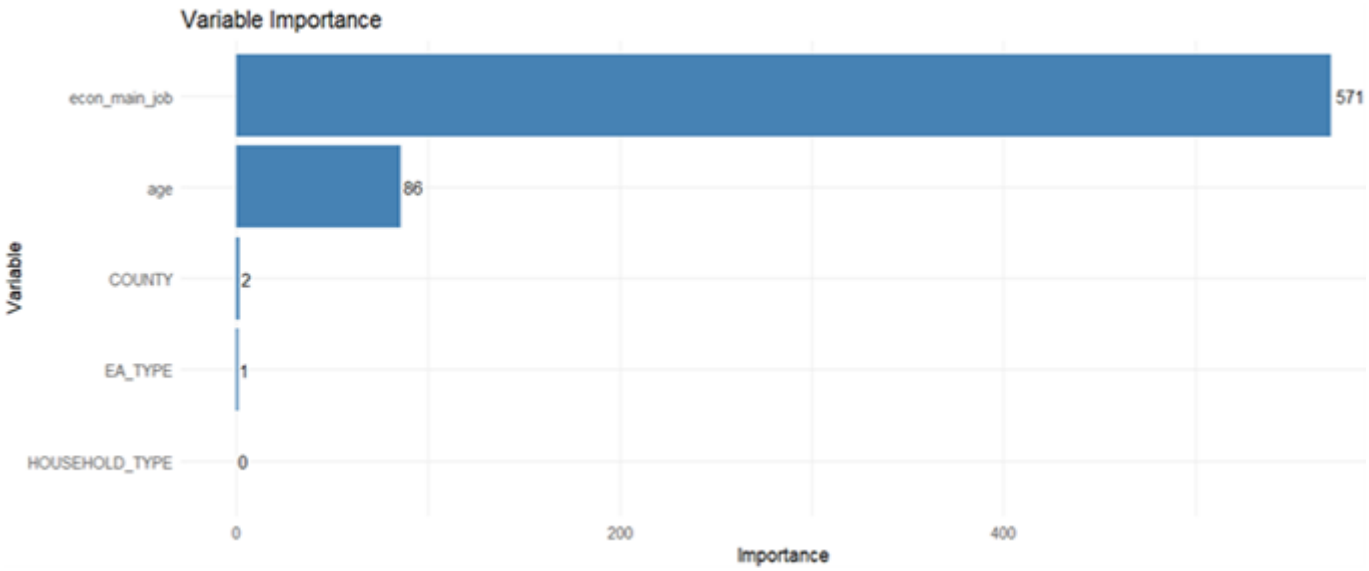

Figure 5. *Variable importance for the Random Forest (RF) model based on the Mean Decrease in Gini index. The plot ranks predictor variables according to their contribution to reducing node impurity during classification, with higher values indicating greater importance in predicting the wealth index.*

The Random Forest error plot revealed in Figure 6 shows significant disparities in classification performance across the five wealth quintiles. The model performed best on the *Poorest* class, with its error rate stabilizing around 20%, indicating relatively accurate predictions. In contrast, the *Poor* and *Middle* classes exhibit persistent error rates close to 100%, suggesting that the model fails to distinguish these classes effectively.

The *Richest* class starts with a high error but stabilized around 35–40%, showing moderate accuracy, while the *Rich* class is also poorly predicted. The overall out-of-bag (OOB) error stabilizes around 66%, highlighting the model's limited predictive power, likely due to class imbalance and overlapping feature patterns among certain quintiles.

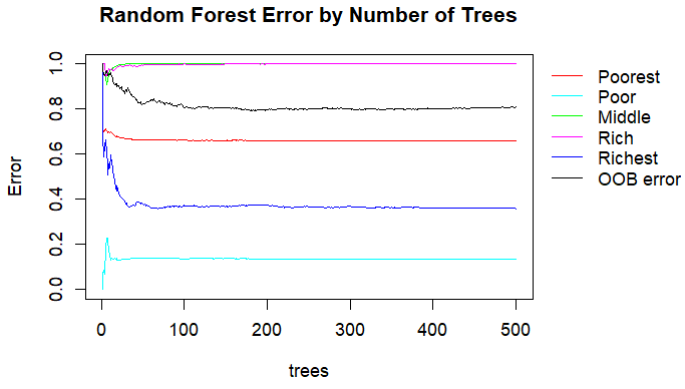

Figure 6. *Random Forest (RF) error rates by number of trees. The plot shows the out-of-bag (OOB) error for the overall model and for each wealth category as the number of trees increases. Error trends illustrate the variation in classification accuracy across the five wealth quintiles, with the Poorest class achieving the lowest error and the Poor and Middle classes showing the highest.*

3.5. Multi-layer Perceptron Performance

To assess the performance of a non-linear classifier in predicting household wealth status, a Multi-Layer Perceptron (MLP) was implemented. The model was trained using repeated 5-fold cross-validation with a grid search over two hyperparameters: the number of hidden nodes (*size*) and the weight decay parameter (*decay*). The final model selected had a size of 5 and a decay of 0.1, based on optimal classification accuracy.

The final MLP model achieved an overall accuracy of 33.4% and a Kappa value of 0.1464, indicating limited but notable agreement beyond chance. The balanced accuracy across all classes was approximately 57.7%, suggesting a moderate ability to classify the five wealth categories correctly.

Relatively high balanced accuracies were achieved as clearly shown in Table 5 for the *Poorest* (65.7%) and *Rich* (65.4%) classes. However, the model severely underperformed for the *Poor* and *Middle* classes, each exhibiting balanced accuracies near 50%, which reflects performance close to random guessing. The *Richest* class achieved the highest precision (45.1%) but still reflects limited confidence in positive predictions across most categories.

Table 5. *Class-wise evaluation metrics for the Multilayer Perceptron (MLP) model. The table reports sensitivity, specificity, precision (PPV), and balanced accuracy for each wealth category, highlighting the variation in model performance across classes.*

Class	Sensitivity	Specificity	Precision	Balanced Accuracy
Poorest	86.56%	44.76%	34.25%	65.66%
Poor	0.26%	99.66%	20.00%	49.96%
Middle	6.20%	92.76%	20.74%	49.48%
Rich	50.90%	79.95%	32.95%	65.42%
Richest	18.77%	97.18%	45.07%	57.97%

A confusion matrix was generated to examine misclassifications across the five wealth categories, clearly displayed in Figure 7 The results indicated that most misclassifications occurred between adjacent wealth classes, particularly between the *Poor* and *Middle* groups. This suggests that the model struggled to capture subtle socio-economic differences and overlapping feature distributions, which are often characteristic of middle-income strata.

Such patterns of confusion highlighted the complexity of wealth classification and the need for improved feature representation among closely related categories.

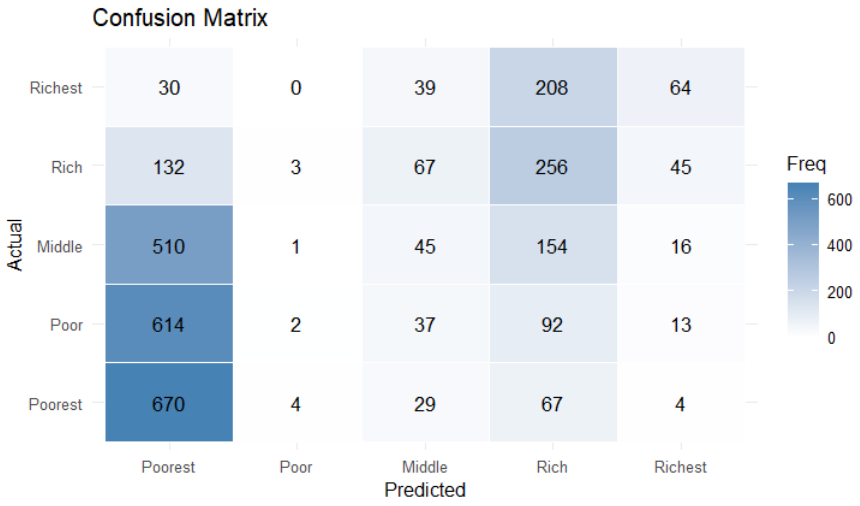

Figure 7. *Confusion matrix for the MLP model. The matrix compares actual versus predicted class labels for the wealth categories. Higher values along the diagonal indicate correct predictions, while off-diagonal values represent misclassifications, particularly between adjacent wealth categories such as Poor and Middle.*

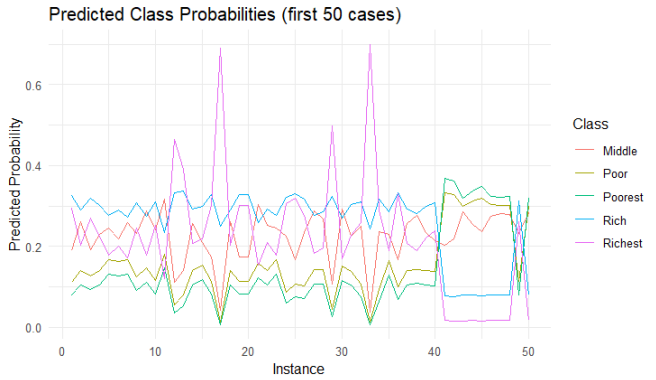

Figure 8. *Predicted class probabilities for the first 50 test instances using the MLP model. More uniform distributions indicate lower confidence, while higher peaks indicate stronger class prediction confidence.*

To further understand the model's predictive behavior, class probability distributions for the first 50 test instances were visualized in Figure 8. This plot depicted the predicted probabilities for each class, offering insight into the model's confidence levels. Many test cases exhibited low confidence, as reflected by the relatively uniform probability distribution across the five classes. This indicates considerable uncertainty in the model's predictions, further supporting the conclusion that the MLP struggled to distinguish between closely related socio-economic groups.

3.6. Ensemble Performance

The ensemble model (RF-MLP) achieved an overall accuracy of 34.4% (95% CI: 32.76% – 36.13%). The No Information Rate (NIR), representing the accuracy obtained by always predicting the most frequent class, was 24.95%, demonstrating that the ensemble significantly outperforms this baseline. The class-wise evaluation metrics are summarized in Table 6.The Kappa statistic was 0.1633, indicating fair agreement beyond chance.

McNemar's test yielded a p-value less than 2.2e-16, suggesting statistically significant differences in misclassification rates and highlighting potential bias or imbalance in the prediction errors.To evaluate the predictive performance of the ensemble model, a confusion matrix was computed using the test dataset.

Table 6. *Class-wise performance metrics for the RF-MLP ensemble model. The table reports sensitivity, specificity, precision (PPV), and balanced accuracy for each wealth category, showing how the ensemble approach performs across the different classes.*

Class	Sensitivity	Specificity	Precision	Balanced Accuracy
Poorest	87.34%	43.94%	34.12%	65.64%
Poor	0.00%	100.00%	0.00%	50.00%
Middle	1.10%	99.41%	36.36%	50.26%
Rich	64.81%	75.14%	33.54%	69.98%
Richest	17.01%	97.50%	45.67%	57.26%

Figure 9 presents the Receiver Operating Characteristic (ROC) curves for the ensemble model, illustrating its ability to discriminate among the five wealth categories. The model exhibited superior discriminative performance for the extreme wealth categories, "Richest" and "Rich," with Area Under the Curve (AUC) values of 0.837 and 0.734, respectively. In contrast, the performance for middle categories such as "Middle" and "Poor" was notably weaker, likely reflecting overlapping socio-economic features within these groups.

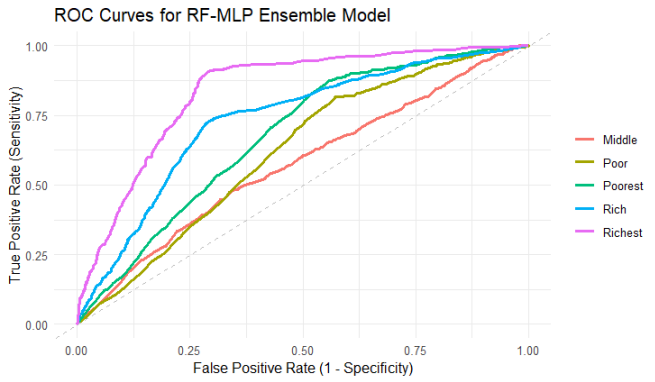

Figure 9. *ROC curves for the RF-MLP ensemble model. Each curve represents the trade-off between true positive rate and false positive rate for a given wealth category, with the AUC indicating the discriminative performance for that class. Higher AUC values correspond to better class separation.*

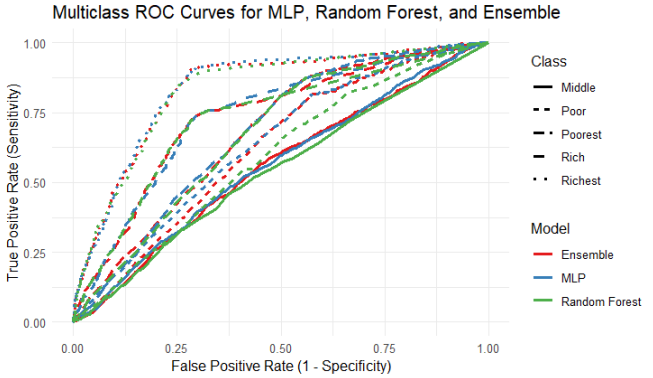

Figure 10. *Comparison of Receiver Operating Characteristic (ROC) curves for the RF, MLP, and RF-MLP ensemble models. Each curve illustrates the trade-off between true positive rate and false positive rate for the respective model, highlighting their discriminative abilities.*

3.7. Model Performance Comparison

Table 7 compares the three models across several metrics. The ensemble showed the highest overall accuracy and specificity.The MLP model achieved the highest average ROC-AUC (0.6912), indicating slightly better discriminatory power than both RF and the ensemble. While the ensemble model attained the highest accuracy and specificity, its ROC-AUC (0.6855) was marginally lower than that of the MLP.

Table 7. *Comparison of performance metrics for the RF, MLP, and RF-MLP ensemble models. The table reports the performance of each model highlighting which model performs best for each metric.*

Metric	RF	MLP	RF-MLP
Accuracy	34.27%	33.43%	34.43%
Precision	41.13%	30.60%	37.42%
Recall	34.27%	32.54%	34.05%
Specificity	83.17%	82.86%	83.20%
ROC-AUC	0.6731	0.6912	0.6855

Random Forest showed the highest precision and recall but had the lowest ROC-AUC, implying it was confident in its predictions but somewhat less reliable in distinguishing between classes. The ROC curves in Figure 10 generally show the ensemble model's curves dominating those of the individual models across most wealth classes.

The ensemble model was best at classifying extreme wealth categories (Poorest and Richest), but struggled with middle and poor classes due to overlapping features and class imbalance.

4. Discussion

This study aimed to develop and compare machine learning models for predicting household wealth levels in Kenya using socio-economic data. The models applied were Random Forest (RF), Multi-Layer Perceptron (MLP), and an ensemble model combining RF and MLP. The model performance was

evaluated using standard classification metrics: Accuracy, Precision, Recall (Sensitivity), Specificity, and ROC-AUC.

The Random Forest model achieved the highest Precision (0.4113) and Recall (0.3427), indicating strong performance in correctly identifying positive cases and minimizing false negatives, particularly in distinguishing wealthier households. The MLP model achieved the highest ROC-AUC score (0.6912), demonstrating its superior ability to discriminate across classes regardless of classification threshold. However, the MLP showed lower precision, which may reflect challenges in correctly identifying wealth categories with imbalanced class distributions.

The ensemble model (RF-MLP) was constructed by combining the outputs of both RF and MLP. This model achieved the highest Accuracy (0.3443) among the three, although the improvement over individual models was marginal. It also had a specificity(0.8320) indicating strong ability to correctly identify non-target classes and avoid false positives, particularly effective in distinguishing poorer households.It also demonstrated relatively balanced performance across the metrics showing potential in capturing broader patterns by leveraging the strengths of both base models.

While the ensemble model outperformed in terms of overall accuracy, it did not consistently surpass the individual models in all metrics. Specifically, RF remained the strongest in precision-based evaluations, while MLP dominated in threshold-independent metrics such as ROC-AUC. The ensemble model's average performance suggests that model combination may improve generalization but may also dilute the strengths of individual models if not well-tuned.

A notable limitation in this multi-class classification context is the inadequacy of accuracy as a sole performance measure. Since accuracy does not account for the distribution of correct predictions across multiple classes, it may misrepresent true model performance. As such, the confusion matrix was a critical tool in this study for assessing classification behavior for each wealth class.

While this study demonstrated the potential of machine learning models, specifically Random Forest, Multi-Layer Perceptron, and an RF-MLP ensemble for predicting household wealth categories, several limitations emerged that provide opportunities for further research.

First, the presence of class imbalance significantly affected model performance, particularly in predicting the middle and poor wealth classes. While this study did not actively implement data balancing techniques due to the research scope and dataset constraints, it explicitly acknowledges this limitation. Future research should focus on applying more advanced data balancing methods such as Synthetic Minority Over-sampling Technique (SMOTE), adaptive synthetic sampling, or cost-sensitive learning to address this challenge.

Second, the model inputs were limited to a fixed set of socio-economic indicators. Subsequent studies should consider incorporating additional features such as geospatial data, household consumption patterns, health metrics, and asset-level information. These variables could enhance the richness of the input space and improve the model's discriminatory power across wealth categories.

Third, while the RF-MLP ensemble showed improved performance in some areas, it did not consistently outperform the individual models. Future researchers could explore more sophisticated ensemble techniques such as model stacking, gradient boosting, or hybrid architectures involving convolutional or recurrent neural networks to better capture nonlinear and temporal patterns in socio-economic data.

Moreover, the study's findings were based on a single dataset (KPHC 2019), which may limit generalizability. To enhance external validity, future work should test the models across multiple datasets or in different country contexts.

Finally, interpretability of the machine learning models remains an important issue, particularly when used for policy-related decision-making. Research into explainable AI (XAI) methods could provide transparency and build trust among stakeholders using these models for social targeting or poverty alleviation programs.

5. Conclusion

The ensemble model slightly outperforms individual models in overall accuracy, but this improvement was marginal. Random Forest proves to be the most reliable in minimizing false positives, making it suitable for applications where misclassifying wealthier households as poorer can have serious policy implications.

MLP, despite lower precision, shows the best ability to separate classes across thresholds (highest AUC), suggesting better robustness in capturing complex relationships. No model achieves significantly high accuracy, highlighting the inherent difficulty of wealth classification based solely on the available features.

The multi-class classification problem is effectively captured in this study. Any intervention should be monitored to ensure that there is a significant reduction in the gap, as the ability of the model to capture the extreme classes very well tells there is an actual gap, that needs said monitoring.

Future studies should explore ensemble models across diverse datasets and regions to improve generalizability. Investigating model fairness, transparency, and explainability remains essential, particularly for government use.

ORCID

0009-0005-7143-5364 (Pinkie Akinyi Odipo)
0000-0003-0268-2968 (Anthony Waititu)
0009-0003-9963-4977 (Herbert Imboga)
0009-0005-9570-9112 (Susan Mwelu)

Abbreviations

AUC Area Under the Curve

CI	Confidence Interval
EA	Enumeration Areas
KNBS	Kenya National Bureau of Statistics
KPHC	Kenya Population and Housing Census
MLP	Multilayer Perceptron
NIR	No Information Rate
OOB	Out-of-Bag
RF	Random Forest
ROC	Receiver Operating Characteristic
SMOTE	Synthetic Minority Oversampling Technique
XAI	Explainable Artificial Intelligence

Acknowledgments

I would like first to express my sincere appreciation and gratitude to my supervisors for their invaluable guidance and for enriching this research and teaching skills over the last several years.

I am deeply impressed with their wealth of knowledge, excellence in teaching, and dedication to academic research. It has been a great honor to undertake my research project under their guidance and to learn from their immense wealth of knowledge, skill and experience.

My appreciation also goes to my fellow colleagues for their additional assistance throughout the completion of this project. Last but not least, I am especially indebted to my parents, whose love, encouragement, and unwavering support made this work possible.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Anwar, H., Qamar, U., & Muzaffar Qureshi, A. W. (2014). Global optimization ensemble model for classification methods. The Scientific World Journal. https://doi.org/10.1155/2014/245353

[2] Attanasio, O., & Binelli, C. (2003). Inequality, growth, and redistributive policies. Conference on Poverty, Inequalities and Growth: What's at Stake for Development Aid?, OECD, Paris.

[3] Azevedo, B., Amoura, Y., Rocha, A., Fernandes, F., Pacheco, M., & Pereira, A. (2022). Analyzing the MATHE platform through clustering algorithms. In Computational Science and Its Applications – ICCSA 2022 Workshops (pp. 201-218). https://doi.org/10.1007/978-3-031-10544-0_15

[4] Barro, R. J. (2000). Inequality and growth in a panel of countries. Journal of Economic Growth, *5*(1), 5-32. https://doi.org/10.1023/A:1009850119329

[5] Bishop, C. M. (2006). Pattern recognition and machine learning.

[6] Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32. https://doi.org/10.1023/A:1010933404324

[7] Gopal, M. (2018). Applied machine learning. McGraw Hill Education Private Limited.

[8] Haralabopoulos, G., Anagnostopoulos, I., & McAuley, D. (2020). Ensemble deep learning for multilabel binary classification of user-generated content. Algorithms, *13*(4), 83. https://doi.org/10.3390/a13040083

[9] Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. Neural Networks, *2*(5), 359-366. https://doi.org/10.1016/0893-6080(89)90020-8

[10] Kiprop, R. A., Wamwea, P., Imboga, H. M., & Chelule, J. K. (2023). Classification of contraceptive use among undergraduate students using a supervised machine learning technique. American Journal of Theoretical and Applied Statistics, *12*(5), 168-176. https://doi.org/10.11648/j.ajtas.20231205.12

[11] Martin, L., & Baten, J. (2022). Inequality and life expectancy in Africa and Asia, 1820–2000. Journal of Economic Behavior & Organization, *201*, 40-59. https://doi.org/10.1016/j.jebo.2022.08.011

[12] Martinangeli, A. F., & Windsteiger, L. (2024). Inequality shapes the propagation of unethical behaviours: Cheating responses to tax evasion along the income distribution. Journal of Economic Behavior & Organization, *220*, 135-181. https://doi.org/10.1016/j.jebo.2023.09.019

[13] Nderitu, J. N., Imboga, H. M., & Gathuka, M. N. (2022). On the coverage properties of the ratio-based estimator in presence of non-response error. American Journal of Theoretical and Applied Statistics, *11*(2), 49-55. https://doi.org/10.11648/j.ajtas.20221102.12

[14] Ng'elechei, W. S., & Imboga, H. M. (2020). Modeling frequency and severity of insurance claims in an insurance portfolio. American Journal of Theoretical and Applied Statistics, *9*(6), 265-270. https://doi.org/10.11648/j.ajtas.20200906.12

[15] Piketty, T. (2014). Capital in the Twenty-First Century (A. Goldhammer, Trans.). Harvard University Press. https://doi.org/10.4159/9780674369542

[16] Ravallion, M. (2012). Why don't we see poverty convergence? American Economic Review, *102*(1), 504-523. https://doi.org/10.1257/aer.102.1.504

[17] Sagi, O., & Rokach, L. (2018). Ensemble learning: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, *8*(4), e1249. https://doi.org/10.1002/widm.1249

[18] Scheidel, W. (2017). The great leveler: Violence and the history of inequality from the Stone Age to the twenty-first century. Princeton University Press. https://doi.org/10.1515/9781400884605

[19] Turchin, P. (2023). End times: Elites, counter-elites, and the path of political disintegration. Penguin.