MATHEMATICAL MODELING AND COMPUTING, Vol. 12, No. 4, pp. 1053–1064 (2025)

Enhanced Poverty Assessment through Advanced Analytical Models within the Malaysia MADANI Framework

Zulkifli F.[1], Abidin R. Z.[1], Zulkifli N. Z.[1,2]

[1] *Center of Mathematical Sciences, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Perak Branch, Tapah Campus, 35400 Tapah Road, Perak, Malaysia*
[2] *Stratos Pinnacle Sdn Bhd. Signature 2, VO6-07-02 Lingkaran SV, Sunway Velocity Cheras, 55100 Kuala Lumpur, Wilayah Persekutuan Kuala Lumpur, Malaysia*

(Received 15 December 2024; Revised 14 July 2025; Accepted 16 July 2025)

This paper presents a research framework for an innovative poverty assessment methodology aligned with the Malaysia MADANI Framework's objectives of eradicating poverty and promoting inclusive economic growth. Traditional approaches to household categorization often neglect critical demographic variables and the unequal distribution of income, leading to an incomplete understanding of poverty. To address these limitations, the framework integrates mixture ordinal regression models with machine learning algorithms, leveraging the strengths of statistical modeling and advanced predictive analytics. By conceptualizing a multi-layered analytical model, the proposed approach provides a more comprehensive and nuanced understanding of poverty dynamics within Malaysia's diverse socio-economic landscape. The model aims to deliver detailed insights essential for designing effective and targeted policy interventions. This framework is expected to overcome shortcomings of conventional methods, offering policymakers a robust and adaptable tool for poverty alleviation. Ultimately, the research seeks to advance the Malaysia MADANI Framework's vision of inclusive growth and development.

Keywords: *poverty assessment; mixture models; ordinal regression models; machine learning algorithms; income distribution; Malaysia MADANI Framework.*

2010 MSC: 62J12, 68T05, 91D10, 62P20, 62H30 **DOI:** 10.23939/mmc2025.04.1053

1. Introduction

The Malaysian MADANI Framework, anchored in principles of compassion, respect, trust, innovation, prosperity, and sustainability, aims to holistically address poverty eradication and economic growth [1]. Its vision revolves around the elimination of poverty by targeting its underlying causes. Poverty, primarily defined by income, signifies inadequate access to essential necessities for a decent standard of living [2]. Despite progress, the World Bank's Malaysia Economic Monitor report forecasts a rise in the poverty rate from 5.6% in 2019 to 8.4% in 2020, albeit with signs of sectoral improvement since late 2021 [3]. However, it is worth noting that there have been indications of sectoral improvements since late 2021, suggesting that certain sectors of the economy have started to recover and potentially alleviate some of the poverty issues.

The urban-rural division highlights the multifaceted nature of poverty. Rural regions contend with insufficient access to critical services like healthcare, education, and transportation, compounded by lower household incomes, rendering them vulnerable to economic shocks and natural disasters [2]. Conversely, urban poverty manifests in inadequate housing and basic services, exacerbating financial strain, especially among those with limited education or skills [4]. Understanding household living standards is pivotal within the Malaysia MADANI Framework to address the diverse challenges of poverty effectively.

This research was supported by the PhD Graduate MyRA Grant, referenced under RMC File No. 600-RMC/GPM LPHD 5/3 (096/2022).

Assessing household living standards entails navigating various factors such as income, education, employment, access to goods and services, housing, savings, assets, and debt, contingent on context and analytical objectives [5]. While conventional methods like the Household Income Surveys (HIS) by the Department of Statistics Malaysia (DOSM) provide insights, critiques regarding inaccuracies in family size and income distribution necessitate recalibrating categorization approaches for a more precise depiction of household economic dynamics [6–8].

Inaccurate or incomplete information on household income and living standards can lead to misguided policies, underscoring the need for a robust assessment methodology. Government initiatives, such as the 2023 budget, highlight progress in extreme poverty eradication, yet further efforts are vital to address poverty's root causes [9]. The Pangkalan Data Utama (PADU) system, integrating socio-economic data from over 200 databases, offers a holistic view beyond income, bolstering poverty reduction endeavors [10].

To enhance poverty assessment in Malaysia, employing mixture models is essential. Despite the array of statistical methods available, the scarcity of studies applying mixture models in ordinal regression poses a significant research gap [10]. Our proposed methodology seeks to bridge this gap by amalgamating finite mixture models, latent class regression, and Gaussian mixture regression. This innovative approach not only enhances prediction accuracy but also offers nuanced insights into Malaysia's complex poverty dimensions.

Incorporating machine learning algorithms into the proposed methodology forms a cornerstone for enhancing assessment accuracy and depth, particularly with large datasets. Leveraging techniques such as ensemble methods, decision trees, and neural networks facilitates analysis of complex socio-economic data patterns [11]. Machine learning algorithms bolster predictive capabilities and furnish policymakers with nuanced insights crucial for targeted interventions within the Malaysia MADANI Framework, even in the context of extensive and diverse data structures.

This study aims to advance poverty assessment within the Malaysia MADANI framework through a novel mixture of ordinal regression models integrated with machine learning algorithms. Aligned with governmental policy, national agendas, and global aspirations, this methodology promises a substantial contribution to poverty reduction at local and national levels.

2. Literature review

The concept of "living standards" encompasses a family's overall well-being, health status, and general satisfaction. In this context, poverty is commonly understood as the absence of resources and opportunities necessary to maintain a basic standard of living.

2.1. Factors influencing living standards and poverty

Economic indicators such as income, employment, education, housing, and demographic characteristics play pivotal roles in determining a family's standard of living and their susceptibility to poverty [12,13]. Income is particularly significant as it facilitates access to essential needs [13]. Despite considerable reductions in poverty rates, Malaysia continues to grapple with income inequality, evident in the substantial disparity between the wealthiest and poorest segments of society.

The stability of employment is closely linked to higher living standards, as evidenced by studies conducted by the International Labour Organisation [12] and the Malaysian Institute of Economic Research [14]. Households with secure employment tend to enjoy higher living standards compared to those with precarious job situations.

Education plays a central role in poverty alleviation and enhancing living standards, as highlighted by the United Nations Development Programme [15] and the Ministry of Education [16]. Higher levels of education are associated with increased income levels and improved employment opportunities, thereby contributing to elevated living standards [15].

The quality and affordability of housing significantly impacts living standards. Research by the Malaysian Centre for Housing Studies [17] and the Ministry of Housing and Local Government Malaysia [18] underscore the importance of access to adequate and affordable housing, emphasizing its role in poverty reduction and the enhancement of living standards.

Demographic factors such as age and gender also exert substantial influence on household well-being. Studies conducted by Adee and Lau [19], Sabri et al. [20], and others demonstrate the contribution of household head age and gender to a family's ability to escape poverty and maintain higher living standards.

2.2. Inequality in poverty in Malaysia

Despite Malaysia's concerted efforts, the unequal distribution of poverty remains a significant challenge. The New Economic Policy (NEP), implemented from 1971 to 1990, aimed to mitigate absolute poverty and rectify economic imbalances [21]. However, issues such as income disparity among ethnic groups and urban poverty persist, posing ongoing hurdles to poverty reduction efforts.

Absolute poverty inequality is particularly pronounced in certain states, notably Sabah, Kelantan, and Sarawak, according to the Economic Planning Unit [21]. Although Malaysia has witnessed an overall decrease in poverty rates, the country continues to grapple with a high Gini coefficient, indicating persistent income inequality.

Ethnic disparities are evident, with the Bumiputera community experiencing higher poverty rates compared to other ethnic groups [21]. While poverty rates have declined for some ethnic groups, urban areas typically exhibit lower poverty rates than rural areas.

2.3. Data and models on poverty and living standards

Household income data serves as a fundamental indicator of living standards, despite occasional reliability issues stemming from under- or over-reporting [22]. Alternatively, monthly expenditure is recommended over monthly income for poverty analysis. The poverty line, reflecting the minimum standard of living, is determined based on expenditure data.

Government agencies' collection of household-level data serves as the primary source for analyzing factors contributing to poverty [8, 22, 23]. However, limitations such as reliance on self-reporting, inadequate dataset comprehensiveness, and challenges with cross-sectional data persist [3].

Various models, including logit, probit, and log-linear regression, have been utilized to examine living standards and poverty [22,24,25]. Nonetheless, these models are not without limitations, including data reliability, coverage issues, and the need to consider factors beyond income and consumption [26].

Despite their limitations, these models remain invaluable tools for policymakers. They provide essential insights into household wealth and poverty, guiding the development of effective policies. Continued advancements in data methodologies and the inclusion of a broader range of socio-economic factors promise to enhance the accuracy of household wealth assessments. Systems like PADU, which integrate comprehensive socio-economic data, are particularly effective in refining these assessments. By offering detailed and nuanced information, PADU enables more targeted and effective interventions to address poverty and improve living standards in Malaysia.

2.4. Evolution of mixture models

The evolution of mixture models can be traced back to the pioneering work of Karl Pearson on crab distribution in the Bay of Naples. Pearson's introduction of a two-component normal mixture model to account for population heterogeneity marked a significant milestone in statistical modeling [27, 28]. This foundational work set the stage for the evolution of mixture models, leading to substantial advancements by subsequent researchers. Day and Wolfe, for instance, expanded upon Pearson's initial concepts, while Dempster, Laird, and Rubin introduced the Expectation–Maximization (EM) algorithm, which greatly facilitated the estimation of parameters in complex mixture models [27–29].

Despite these significant advancements, several theoretical questions about mixture models persist, posing both challenges and opportunities for future research [30]. One of the critical areas of ongoing investigation is the accuracy and reliability of existing mixture models [31]. The performance of these models in capturing the underlying distribution of data remains a key concern, particularly in applications requiring precise classification and prediction.

Mixture models have demonstrated their utility across a wide range of disciplines, including classification, machine learning, and multivariate analysis [27]. Their ability to model population heterogeneity and complex data structures has made them indispensable tools in many fields. However, the accuracy of these models can vary significantly depending on the context and the nature of the data being analyzed. For instance, the effectiveness of mixture models in handling high-dimensional data, sparse data, and data with outliers remains an active area of research [28].

Model selection is a particularly challenging aspect of mixture modeling. The choice of the number of components in a mixture model can significantly impact its accuracy and interpretability [32]. Overfitting and underfitting are common problems that arise from incorrect model selection [33]. Researchers have proposed various criteria and techniques for model selection, such as the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), but these methods are not without their limitations and continue to be refined [34].

Parameter estimation in mixture models is another area that requires careful consideration. The EM algorithm, while powerful, can sometimes converge to local optima, leading to suboptimal parameter estimates [29, 35]. Alternative estimation methods and enhancements to the EM algorithm have been proposed to address these issues, but challenges remain in ensuring robust and reliable parameter estimation.

The continued exploration and refinement of mixture models offers promising avenues for future research. Addressing existing theoretical gaps, improving parameter estimation methods, and developing robust model selection criteria are essential steps toward enhancing the accuracy and reliability of these models. By advancing the theoretical foundations and practical applications of mixture models, researchers can unlock their full potential and contribute to the development of more sophisticated and effective statistical tools. This ongoing research will not only improve the utility of mixture models in existing applications but also pave the way for new and innovative uses in emerging fields.

3. Methodology

This research introduces a novel methodology aimed at enhancing the accuracy and granularity of assessing household living standards and poverty levels in Malaysia. Leveraging the comprehensive PADU system, which encompasses a wide array of demographic data, income profiles, financial obligations, and household expenditures, this study proposes an innovative approach that integrates state-of-the-art ordinal regression modeling with advanced machine learning algorithms.

The proposed methodology is designed to fill existing gaps in current assessment frameworks by combining sophisticated statistical techniques with machine learning capabilities. By exploring the multidimensional PADU dataset, this research endeavors to provide a more comprehensive understanding of poverty dynamics in Malaysia. The research framework is structured into distinct phases, each serving a specific purpose:

Phase I: Theoretical.

This initial phase involves conducting a comprehensive theoretical overview and literature review. Its aim is to identify the necessity for the research and establish a theoretical framework based on prior studies. Through this phase, the study gains an understanding of the variables and models previously utilized by researchers. While this phase achieves the first research objective (RO1) of the study, further support through measurement analysis is necessary in subsequent phases.

Phase II: Development of a mixture of ordinal regression models and machine learning algorithms.

Building upon the insights garnered from the theoretical phase, this stage focuses on developing a unique methodology integrating mixture models with ordinal regression and machine learning algorithms. By assessing existing models for household living standards and poverty levels, this phase aims to identify their strengths and weaknesses, offering insights into areas for improvement or extension. Notably, the absence of studies employing mixture models for assessing living standards or poverty underscores the novelty of this research direction.

The proposed methodology entails the application of finite mixture models to capture the heterogeneity inherent in household data. This involves defining a discrete random variable to represent ordered responses associated with multinomial denominators. The model's formulation encompasses a set of assumptions regarding the underlying mixing process and the distribution of responses across different components. Utilizing maximum likelihood estimation techniques, model parameters can be estimated effectively.

The mixture of ordinal regression models operates under the following assumptions:

1. The unobserved mixing process can assume any one of c states, where c is finite and unknown.
2. For each observed multinomial ordered response y_i associated with a multinomial ordered denominator m_i, an unobserved random variable, Π_i, represents the component generating y_i. Additionally, the set of observations (Y_i, Π_i) is pairwise and sequentially independent.
3. Π_i follows a discrete distribution with c points of support, where $P\left(\Pi_i = j\right)$ and $\sum_{j=1}^{c} p_{ij} = 1$ for each i.
4. The probability $p_{ij} = p_j\big(\mathbf{x}_i^{(m)}, \beta\big)$ is determined by the function:
$$F\left(\alpha_i - \sum_{k=1}^{c-1} \beta_j' \mathbf{x}_i^m\right) - F\left(\alpha_{i-1} - \sum_{k=1}^{c-1} \beta_j' \mathbf{x}_i^m\right), \tag{1}$$
where $F(\cdot)$ represents the random variable error distribution of the model. The vectors $\beta_j = \big(\beta_{j1}, \beta_{j2}, \ldots, \beta_{jk_1}\big)'$ for $1 \leqslant j \leqslant c-1$ are unknown parameters. It is noteworthy that all components of β appear in each mixing probability p_{ij}.
5. The equation:
$$p_{ic} = p_c\big(\mathbf{x}_i^{(m)}, \beta\big) = 1 - \sum_{j=1}^{c-1} p_{ij} \tag{2}$$
further defines the probability distribution, where the covariates $\mathbf{x}_i^{(m)}$ and β are involved.
6. The relationship between the multinomial ordered distribution and covariates is governed by the equation:
$$\pi_{ij} = \pi_i\big(\mathbf{x}_i^{(r)}; \alpha_j\big) = \frac{\exp\left(\alpha_j' \mathbf{x}_i^{(r)} + e_{ij}\right)}{1 + \exp\left(\alpha_j' \mathbf{x}_i^{(r)} + e_{ij}\right)} \tag{3}$$
for $j = 1, 2, \ldots, c$. Here, e_{ij} is the random variable error of the model. The vectors $\alpha_j = \big(\alpha_{j1}, \ldots, \alpha_{jk_2}\big)'$ for $1 \leqslant j \leqslant c$ are unknown parameters.

As a consequence of these assumptions, the unconditional distribution of represents a finite multinomial ordered mixture. The mixing probabilities p_{ij} are linked to the covariates $\mathbf{x}_i^{(m)}$ through the cumulative logistic link function. The component distributions are multinomial ordered distributions with success probabilities π_{ij}.

Model parameters are estimated using maximum likelihood estimates derived from a combination of the EM algorithm and the quasi-Newton algorithm for a fixed number of components c.

The classification system employed for the response variable in this model aligns with that utilized by DOSM for categorizing household income as in Table 1. While income serves as a common proxy for standard of living, this research acknowledges the multifaceted nature of poverty determination, considering various factors beyond income. Subsequent phases of the study will delve deeper into these additional factors, enriching the understanding of household living standards and poverty dynamics. Here is an overview of household income categorization in Malaysia for the year 2020:

Table 1. New Malaysia household income categorization (Source: DOSM (2020)).

Group	Median	Mean	Income Level
B40			
B1	RM1,929	RM1,829	Less RM2,500
B2	RM2,786	RM2,803	RM2,501 – RM3,170
B3	RM3,556	RM3,561	RM3,171 – RM3,970
B4	RM4,387	RM4,395	RM3,971 – RM4,850
M40			
M1	RM5,336	RM5,346	RM4,851 – RM5,880
M2	RM6,421	RM6,477	RM5,881 – RM7,100
M3	RM7,828	RM7,841	RM7,101 – RM8,700
M4	RM9,695	RM9,730	RM8,701 – RM10,970
T20			
T1	RM12,586	RM12,720	RM10,971 – RM15,040
T2	RM19,781	RM24,293	RM15,041 and more

Various machine learning algorithms can be employed to construct mixture regression models. One approach is to utilize a mixture of linear regression models, where each model handles the modeling of a distinct subpopulation, also known as a mixture of the expert's model. Another strategy involves employing a clustering algorithm to identify subpopulations and then fitting separate regression models to each subgroup, termed as a cluster-wise linear regression model.

This study will focus on utilizing the clustering algorithm to classify household living standards, detailed in Phase VI. However, the algorithms employed necessitate the development of new coding programming owing to the innovative mixture model devised in this research.

The theoretical framework for an instance with the K-Means clustering algorithm is provided below. The algorithm steps are as follows:

1. Initially, K random centers are selected from the dataset to function as initial cluster centers.
2. For each data point in the dataset, compute the distances (typically Euclidean) to all centers and assign the point to the cluster with the closest center.
3. Recalculate the centroids of the newly formed clusters by computing the mean of all data points assigned to each cluster, representing the center of gravity for each cluster.
4. Iterate the steps of assigning and recalculating the centroids until convergence is attained, defined by stable centroids or a predefined number of iterations.

Mathematical representation.

Given a dataset with N households and M features, where $\mathbf{X} = \{x_1, x_2, \ldots, x_N\}$ and $C = \{c_1, c_2, \ldots, c_K\}$ denotes the centroids of K clusters:

a) Euclidean distance:

$$D(x_i, c_k) = \sqrt{\sum_{j=1}^{M} (x_{ij} - c_{kj})^2}. \tag{4}$$

b) Assignment:

$$y_i = \arg\min_k D(x_i, c_k). \tag{5}$$

c) Centroid update:

$$\mathbf{c}_k = \frac{1}{|\mathcal{S}_k|} \sum_{i \in \mathcal{S}_k} x_i, \tag{6}$$

where $\mathcal{S}_k$ represents the set of data points assigned to cluster k.

In our investigation, this algorithm is applied to socio-economic household data, where each data point denotes a household and encompasses characteristics like income, education, housing, etc. The resultant clusters signify groups of households with similar living standards.

Phase III: Simulation and testing procedures for the mixture of ordinal regression models.

During this phase, the proposed mixture regression model will undergo simulation, and its performance on simulated data will be assessed. It is expected that compared to existing models, the new method will yield more accurate predictions and enhanced interpretability. The following outlines the steps involved in simulating and testing the new mixture regression model:

1. Before simulating the new mixture regression model, it is essential to define the data generation process by establishing a set of parameters that delineate the relationship between the predictor variables and the response variable.
2. The effectiveness of the new mixture regression model will be gauged by comparing its predictions on simulated data with the actual values of the response variables. Various metrics such as mean squared error (MSE), root mean squared error (RMSE), and R-squared can be employed to evaluate the model's performance, quantifying its ability to capture the true relationship between predictors and the response variable.
3. The performance of the proposed model can be juxtaposed against existing models by simulating data using the same data generation process. Comparative evaluation of model success can be conducted using analogous criteria as in the second stage.

4. Sensitivity analyses can also be implemented to assess the robustness of the model by comparing its performance to other models under varying conditions. Evaluating the model across different settings, such as diverse noise levels and sample sizes, can provide insights into its adaptability.
5. Following the simulation and testing phase, refinement of the proposed model can be undertaken based on the findings. This may involve adjustments to the model's hyperparameters or the underlying algorithm. Modifying parameters of the mixture model or the non-linear function utilized to model the association between predictors and the response variable can enhance model performance.

Phase IV: Data collection.

To develop and validate the novel model, this study will leverage data from PADU system. PADU encompasses diverse datasets offering comprehensive insights into household income, expenditure patterns, tax information, and pertinent socio-economic factors. This system furnishes detailed information on the financial status of households across Malaysia, laying a robust foundation for a comprehensive analysis of factors influencing living standards and poverty in the nation. Datasets procured through PADU undergo meticulous pre-processing procedures focusing on data cleaning, normalization, and organization to ensure accuracy and consistency. This pivotal phase involves rectifying anomalies, addressing missing information, and structuring the data optimally for subsequent modeling and evaluation processes, employing rigorous statistical methodologies.

Phase V: Validation of real data for the mixture of ordinal regression models.

Validating a novel mixture regression model with real data involves applying the model to carefully selected research databases. This entails utilizing household income as the response variable, while other factors serve as predictor variables. The validation process for the new mixture regression model entails the following steps:

1. Partitioning the real-world dataset into training and testing sets to prevent overfitting. The model is trained on the training set while being assessed on the testing set.
2. Training the model using the proposed methodology and the training set, which involves defining the model's hyperparameters and fitting them to the training data.
3. Applying the trained model to the testing set to generate predictions. The model's predictions are then compared to the actual response variable values in the testing set using metrics such as mean squared error (MSE), root mean squared error (RMSE), and R-squared.
4. Comparing the performance of the proposed model with existing models using real data, similar to the simulation and testing phase. This comparison helps ascertain whether the new model yields superior results in real-world scenarios.
5. Conducting sensitivity analyses on the proposed model using actual data to evaluate its robustness. This involves assessing the model's performance under varying levels of noise, sample sizes, and other data variations.
6. Refining the model based on the validation results by adjusting hyperparameters, modifying the underlying algorithm, or addressing any identified issues.

Phase VI: Determining poverty levels among Malaysian households using a mixture of ordinal regression models and machine learning algorithms.

While implementing a mixture regression model and an unsupervised machine learning strategy share some similarities, there are significant distinctions between the two approaches. Key differences include:

1. Data preparation: In a mixture regression model, data must be categorized into predictor variables and response variables, whereas unsupervised machine learning focuses solely on predictor variables to identify patterns and relationships.
2. Model type: Mixture regression models utilize ordinal regression, while unsupervised machine learning may involve techniques such as K-means clustering, hierarchical clustering, or Gaussian mixture models.

3. Performance evaluation: Mixture regression model performance is assessed based on its ability to predict response variables accurately, while unsupervised machine learning evaluates the model's ability to identify meaningful clusters in the data.
4. Interpretation of results: Mixture regression model results are interpreted based on estimated coefficients and variable importance measures, whereas unsupervised machine learning focuses on cluster characteristics to uncover patterns and relationships. Both approaches may utilize visualization techniques for result interpretation, such as scatter plots or heat maps.

To ensure the accuracy and reliability of the new model for assessing poverty and living standards in Malaysia, several robustness tests can be considered:

1. Sensitivity analysis: Assessing the model's sensitivity to different input variables and parameters by varying factors such as income, education, and wealth weights.
2. Comparison with existing methods: Contrasting the results of the new model with current household classification methods like B40, M40, or T20 to gauge its accuracy and differentiation.
3. Predictive power assessment: Evaluating the model's ability to predict household living standards and poverty using separate data not used in model development.
4. Assumption sensitivity: Testing the model's robustness to changes in assumptions, such as income distribution, to understand their impact on household classification.

Interpreting the model results can provide insights into poverty drivers in urban and rural areas, aiding in policy formulation:

1. Identifying key poverty factors: Through coefficient analysis and variable importance measures, the model can help inform targeted policies and interventions.
2. Comparing urban and rural poverty levels: This comparison will discern contributing factors and tailor policies accordingly.
3. Developing accurate poverty maps: Based on the model results, poverty maps can be created to allocate resources effectively.
4. Crafting evidence-based policy recommendations: Tailored to urban and rural needs, these recommendations aim for poverty alleviation and improved quality of life.

This phase will address the second research objective (RO2), which is to determine poverty levels among urban and rural Malaysian households. The overarching research framework is depicted in Figure 1. Meanwhile, Figure 2 illustrates the detailed steps involved in the clustering algorithm. The results from the mixture of ordinal regression models combined with the clustering algorithm, using the PADU database, will address the third research objective (RO3).

4. Expected results and benefits

This study envisions yielding innovative theories, novel insights, and fresh perspectives. By integrating mixture ordinal regression modeling with machine learning algorithms, the research seeks to pioneer a ground-breaking approach to evaluating household living standards and poverty in Malaysia. It aims to uncover nuanced understandings of the multifaceted determinants of living standards and explore previously uncharted factors shaping poverty dynamics.

Through this innovative model, the study endeavors to offer a comprehensive understanding of the complex interactions among demographic variables, income disparities, and financial obligations, contributing significantly to the existing literature on poverty assessment methodologies, especially within the Malaysian context. The insights garnered from this research are poised to influence policymaking and decision-making processes by providing deeper insights into urban and rural poverty disparities. By gaining a more profound understanding of these dynamics, policymakers can formulate targeted interventions to address regional inequalities and enhance poverty alleviation efforts.

Using PADU system as a primary data source empowers stakeholders to make informed decisions backed by empirical evidence. The research outcomes are expected to provide a robust framework for government agencies, policymakers, and social institutions to devise evidence-based policies and allocate resources effectively, ultimately benefiting marginalized communities.

RO1: To investigate the influence of demographic factors such as age, gender, education, income, expenditure, and commitments in determining the standard of living of a household in Malaysia	
Phase I: Theoretical	• Conduct a literature review and develop a theoretical framework for the research issue. • Identify variables considered in previous studies and models used.

RO2: To develop a novel mixture of ordinal regression models and machine learning algorithms to accurately assess and classify the level of household living standards and overcome the limitations of existing methods.	
Phase II: Development of a Mixture of Ordinal Regression Models and Machine Learning Algorithms	• Evaluate existing models and identify their strengths and weaknesses. • Formulate a mixture of ordinal regression models. • Develop a clustering algorithm to classify household living standards. • Use *R* software for coding.
Phase III: Simulation and Testing Processes on the Mixture of Ordinal Regression Models	• Define a set of parameters that define the relationship between predictor variables and response variables. • Evaluate the accuracy of the new mixture regression model and compare it with existing models. • Use metrics such as MSE, RMSE, and R-squared. • Test the model's robustness by conducting sensitivity analyses. • Fine-tune the proposed model based on the findings.
Phase IV: Data Collection	• Collect data from PADU databases. • Use data from the last ten years. • Clean and prepare the data for analysis.
Phase V: Validation of Real Data on the Mixture of Ordinal Regression Models	• Partition the real-world dataset into a training set and a testing set. • Train the model using the proposed model and the training set. • Test the model's effectiveness on the testing set.

RO3: To investigate poverty levels among urban and rural Malaysian households while identifying contributing factors to observed variations	
Phase VI: Classification of Household Living Standards and Poverty	• Use the clustering algorithm to classify household living standards. • Apply the developed mixture of ordinal regression models to predict the level of household living standards.

Fig. 1. Proposed research framework.

Moreover, the expected results hold the potential to contribute significantly to academic scholarship and empirical research in poverty assessment methodologies. By developing a comprehensive model integrating machine learning techniques, this research could serve as a reference point for future studies, not only in Malaysia but also in similar socio-economic contexts globally.

Analyzing household income, expenditure, and demographic data in detail aims to enhance understanding of poverty dynamics over time. The anticipated outcomes can highlight trends, identify vulnerable groups, and propose strategies to combat poverty, fostering a more inclusive and equitable society. At the societal level, the proposed innovative model seeks to unravel the intricate web of factors influencing household living standards and poverty, paving the way for targeted social policy measures and more inclusive welfare programs.

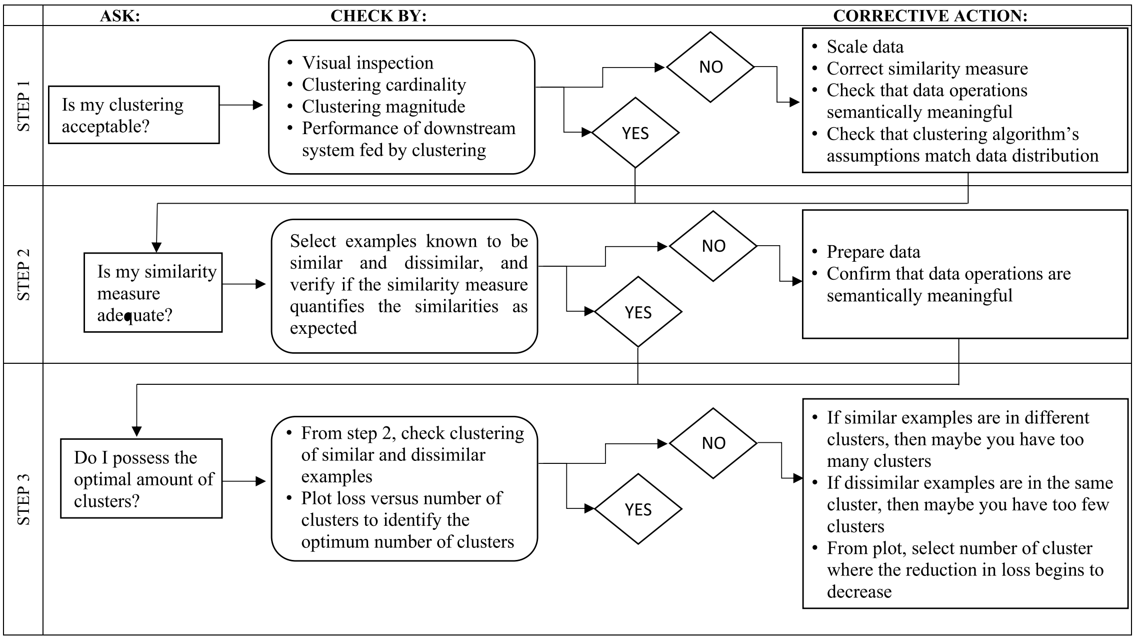

Fig. 2. Steps of clustering algorithm.

In economic terms, the research provides a robust framework for evidence-based decision-making, facilitating efficient resource allocation, strategic economic planning, and sustainable development. At the national level, the research outcomes hold the potential to redefine poverty assessment and alleviation strategies in Malaysia, aligning with national development goals. Furthermore, they contribute to Malaysia's position in the global discourse on poverty assessment methodologies, demonstrating the country's commitment to leveraging research for societal progress.

5. Conclusion

In this paper, we proposed a novel methodology integrating mixture ordinal regression modeling with machine learning algorithms to assess household living standards and poverty in Malaysia. Through a structured approach encompassing theoretical development, simulation, testing, and validation phases, our study aims to provide a comprehensive research framework for understanding and addressing poverty dynamics in the Malaysian context.

By leveraging innovative techniques such as mixture ordinal regression and machine learning algorithms, our proposed methodology offers a nuanced understanding of the complex relationships between demographic factors, income disparities, and household well-being. Through simulation and testing processes, we anticipate that our approach will yield more precise predictions and greater interpretability compared to existing models.

The validation of our proposed methodology using real-world data from PADU system is expected to provide valuable insights for policymakers and stakeholders. By examining the factors contributing to urban and rural poverty levels, our study aims to inform targeted policy interventions and resource allocation strategies to address regional inequalities effectively.

The proposed methodology holds the potential to redefine poverty assessment strategies in Malaysia and contribute to the global discourse on poverty alleviation. Through its interdisciplinary approach and rigorous methodology, our study seeks to make a significant contribution to the ongoing efforts to combat poverty and promote social and economic development in Malaysia.

In essence, this research contributes to the ongoing discussion on consumer behaviour in Malaysia and offers practical insights for stakeholders looking to navigate the complex landscape of pricing and purchasing decisions. By acknowledging the interplay of economic, social, and psychological factors, this study lays the groundwork for future studies aimed at addressing the evolving needs and preferences of Malaysian consumers in a rapidly changing economic environment.

[1] Musa M., Jemain A. A. Exploring detrending techniques in detecting long-memory of ozone time series in Malaysia by simulation. MODSIM 2011 – 19th International Congress on Modelling and Simulation – Sustaining Our Future: Understanding and Living with Uncertainty. 2197–2203 (2011).

[2] Lister R. Poverty. John Wiley & Sons (2021).

[3] World Bank. Malaysia Economic Monitor June 2022: Catching Up: Inclusive Recovery Growth for Lagging States. https://www.worldbank.org/en/country/malaysia/publication/malaysia-economic-monitor-june-2022-catching-up.

[4] Ariyanto K. Literature Review: Urban Poverty in a Sociological Perspective. Antroposen: Journal of Social Studies and Humaniora. **2** (1), 24–32 (2023).

[5] Collins M. L. Assessing Household's Living Standards and Income Resilience at the outset of the Cost-of-Living Crisis. Journal of the Statistical and Social Inquiry Society of Ireland. **52** (1), 48–59 (2023).

[6] Department of Statistics Malaysia (DOSM). Putrajaya, DOSM (2020). Household Income and Basic Amenities Survey Report 2019. https://v1.dosm.gov.my/v1/index.php?r=column/cthemeByCat&cat=120&bul_id=TU00TmRhQ1N5TUxHVWN0T2VjbXJYZz09&menu_id=amVoWU54UTl0a21NWmdhMjFMMWcyZz09.

[7] Hamid Abdul H. Pengkelasan Isi Rumah Malaysia Perlu Diperhalusi Semula. Khazanah Research Institute License: Creative Commons Attribution CC BY 30 (2020).

[8] Safari M. A. M., Masseran N., Hussain S. I. Modeling urban household income in Malaysia using parametric approach. Journal of Physics: Conference Series. **1988**, 012090 (2021).

[9] Ministry of Finance, Malaysia. Belanjawan 2023 Malaysia MADANI: Ucapan Belanjawan (2023).

[10] Ministry of Finance, Malaysia. Sistem pangkalan data PADU perkemas pemberian bantuan kerajaan, elak pertindihan. Bernama (2023).

[11] Murphy K. P. Probabilistic Machine Learning: Advanced Topics. MIT Press (2023).

[12] International Labour Office. World Employment Social Outlook: Trends 2019. Geneva: International Labour Office (2019).

[13] World Bank Group. World Bank Group Annual Reports 2020: Executive Summary. Washington, D.C. (2020).

[14] Malaysian Institute of Economic Research. Malaysian Economic Outlook 4th Quarter 2021 – Malaysian Institute of Economic Research (2021).
https://mier.org.my/malaysian-economic-outlook-4th-quarter-2021/.

[15] United Nations Development Programme. Annual Report 2020. New York (2020).

[16] Ministry of Education. Pelan Pembangunan Pendidikan Malaysia 2013–2025 (2018).

[17] Joint Center for Housing Studies. The State of the Nation's Housing 2021 (2021).

[18] Roosli R et al. Habitat Magazine, Housing for All: a Better Urban Future. Ministry of Housing and Local Government Malaysia (2020).

[19] Adee S., Lau E. A Light Purse Is a Heavy Curse: A Systematic Review. DLSU Business & Economics Review. **31** (2), 110–124 (2022).

[20] Sabri M. F., Mokhtar N., Ho C. S. F., Anthony M., Wijekoon R. Effects of Gender and Income on Malaysian's Financial Capability. Malaysian Journal of Consumer and Family Economics. **24**, 124–152 (2020).

[21] Economic Planning Unit. Mid-term Review of the Eleventh Malaysia Plan (2020).

[22] Muhammad Sapri N. A., Ibrahim K., Abu Bakar M. A., Mohd Ariff N. Ordinal Regression for Modelling the Family Well-being Among the Malaysians. Journal of Nusantara Studies. **6** (2), 424–447 (2021).

[23] Ayyash M., Sek S. K. Decomposing Inequality in Household Consumption Expenditure in Malaysia. Economies. **8** (4), 83 (2020).

[24] Garza-Rodriguez J., Ayala-Diaz G. A., Coronado-Saucedo G. G., Garza-Garza E. G., Ovando-Martinez O. Determinants of Poverty in Mexico: A Quantile Regression Analysis. Economies. **9** (2), 60 (2021).

[25] Ambaye T. K., Tsehay A. S., Hailu A. G. Application of Ordered Logit Model to Analyze Determinants of Rural Households Multidimensional Poverty in Western Ethiopia. International Journal of Development and Economic Sustainability. **9** (1), 18–62 (2021).

[26] Farah M. J., Hook C. J. Trust and the poverty trap. Proceedings of the National Academy of Sciences of the United States of America. **114** (21), 5327 (2017).

[27] Sakthivel K. M., Vidhya G. A New Finite Mixture of Probability Models With Application. Reliability: Theory & Applications. **3** (74), 135–153 (2023).

[28] Domingues T. D. Finite Mixture Models based on Scale Mixtures of Skew-Normal distributions applied to serological data (2021).

[29] De Alencar F. H. C., Matos L. A., Lachos V. H. Finite Mixture of Censored Linear Mixed Models for Irregularly Observed Longitudinal Data. Journal of Classification. **39** (3), 463–486 (2022).

[30] Fruhwirth-Schnatter S., Celeux G., Robert C. P. Handbook of Mixture Analysis. New York, Chapman and Hall/CRC (2019).

[31] Agresti A. Categorical Data Analysis. John Wiley & Sons, Inc. (2013).

[32] Cai J.-H., Song X.-Y., Lam K.-H., Ip E. H.-S. A mixture of generalized latent variable models for mixed mode and heterogeneous data. Computational Statistics & Data Analysis. **55** (11), 2889–2907 (2011).

[33] Steyerberg E. W. Overfitting and Optimism in Prediction Models. In: Clinical Prediction Models. Statistics for Biology and Health. Springer, Cham (2019).

[34] Lu P., Wang H., Tolliver D. Prediction of Bridge Component Ratings Using Ordinal Logistic Regression Model. Mathematical Problems in Engineering. **2019**, 9797584 (2019).

[35] Lyaqini S., Nachaoui M. Diabetes prediction using an improved machine learning approach. Mathematical Modeling and Computing. **8** (4), 726–735 (2021).

Дослідницький фреймворк для покращеної оцінки бідності за допомогою передових аналітичних моделей у межах Малайзійської рамкової програми MADANI

Зулькіфлі Ф.[1], Абідін Р. З.[1], Зулькіфлі Н. З.[1,2]

[1] *Центр математичних наук, Факультет комп'ютерних та математичних наук, Технологічний університет MAPA, Філія в Пераку, кампус Тапа, 35400 Tapah Road, Перак, Малайзія*
[2] *Стратос пінакле Сдн Бгд. Сігнатуре 2, VO6-07-02 Лінгкаран СВ, Санвей Велосіті Черас, 55100 Куала-Лумпур, Вілайя Персекутуан Куала-Лумпур, Малайзія*

У цій статті подано дослідницький фреймворк для інноваційної методології оцінки бідності, яка відповідає цілям Малайзійської рамкової програми MADANI щодо викорінення бідності та сприяння інклюзивному економічному зростанню. Традиційні підходи до категоризації домогосподарств часто нехтують критичними демографічними змінними та нерівномірним розподілом доходів, що призводить до неповного розуміння бідності. Щоб усунути ці обмеження, програма інтегрує моделі змішаної порядкової регресії з алгоритмами машинного навчання, використовуючи переваги статистичного моделювання та розширеної прогнозної аналітики. Концептуалізуючи багаторівневу аналітичну модель, запропонований підхід забезпечує більш повне та тонке розуміння динаміки бідності в різноманітному соціально-економічному ландшафті Малайзії. Модель спрямована на надання детальної інформації, необхідної для розробки ефективних і цілеспрямованих політичних заходів. Очікується, що ця структура подолає недоліки звичайних методів, запропонувавши політикам надійний і адаптований інструмент для подолання бідності. У кінцевому підсумку, дослідження спрямоване на просування бачення Малайзійської рамкової програми MADANI про інклюзивне зростання та розвиток.

Ключові слова: *оцінка бідності; змішані моделі; порядкові регресійні моделі; алгоритми машинного навчання; розподіл доходів; Малайзійська рамкова програма MADANI.*