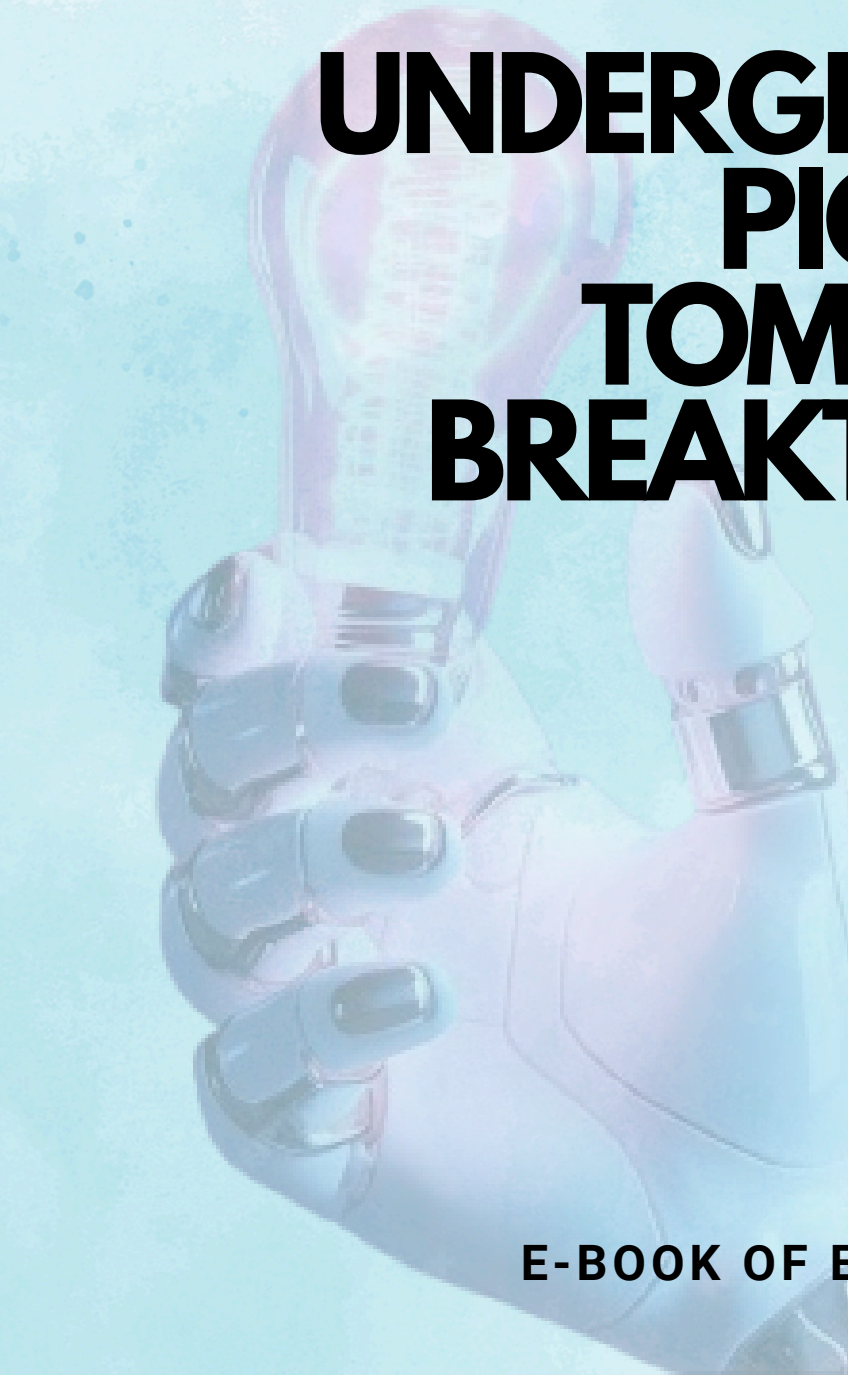
UNIVERSITI TEKNOLOGI MARA

I-URID 2025

INTERNATIONAL UNDERGRADUATE RESEARCH INNOVATION, INVENTION AND DESIGN

UNDERGRADUATES PIONEERING TOMORROW'S BREAKTHROUGH



E-BOOK OF EXTENDED ABSTRACTS

BRIDGING THE GAP: REGIONAL POVERTY PREDICTION AND POLICY INSIGHTS USING MACHINE LEARNING

Nor Azuana Ramli, Almira Damia Syahnizam*, Nurkhairul Izzati Mohd Sallehan, Vienosha A/P Selvaraju, and Nur Atieka Rafiekah Razak

Centre For Mathematical Sciences, Universiti Malaysia Pahang Al-Sultan Abdullah, Lebuh Persiaran Tun Khalil Yaakob, 26300 Kuantan, Pahang.

*almradamia@gmail.com

ABSTRACT

Poverty remains a significant socio-economic challenge in Malaysia, with regional disparities affecting the effectiveness of centralized aid programs. Many interventions lack precision, often due to the limited use of predictive tools and insufficient data-driven targeting. This project introduces a machine learning-based system for classifying poverty levels (low, moderate, high) using multi-domain socio-economic indicators, including income, employment, education, and inflation. Publicly available data from the Department of Statistics Malaysia (DOSM) from 2019 to 2022 was used to train and evaluate five classification models. Among them, the XGBoost model demonstrated the highest performance, achieving an accuracy of 77.2% and an AUC score of 0.93. Key features influencing poverty were selected based on the model that performed best. The solution is deployed as an interactive web application using Streamlit. The dashboard provides district-level insights into poverty, visualisations of poverty distribution, and a feature importance analysis. It also presents tailored policy recommendations based on dominant poverty factors in each region. This tool empowers decision-makers to implement more informed and evidence-based poverty interventions. By bridging artificial intelligence with social policy, the project contributes to more effective resource allocation and equitable development strategies across Malaysian states.

Keywords: Poverty, Machine Learning, Socioeconomic Indicators, Policy Recommendation, Streamlit Application

INTRODUCTION

Poverty relief encompasses both short-term support, including food, shelter, and emergency aid, and long-term strategies, such as education, job training, and inclusive development programs. In Malaysia, poverty is not just a personal hardship, but a systemic issue that hinders national progress, exacerbates inequality, and leads to broader social challenges. In alignment with the United Nations Sustainable Development Goals (SDGs), particularly SDG 1: No Poverty, Malaysia has committed to eradicating poverty in all its forms by 2030, as outlined in the 2030 Agenda for Sustainable Development. However, despite these commitments and numerous policy efforts, significant disparities persist across regions, particularly in underdeveloped states such as Sabah and Kelantan, where access to education, infrastructure, and economic opportunities remains limited (Md Shah et al., 2023). Programs like 1AZAM have highlighted inefficiencies in resource allocation and gaps in addressing the root causes of poverty. Additionally, demographic shifts and the rising cost of living continue to worsen socioeconomic vulnerabilities across the country (Rahman et al., 2021).

This project adopts a data-driven approach to understanding and addressing poverty in Malaysia, with a specific focus on the roles of education, household income, and employment. Publicly accessible datasets from the Department of Statistics Malaysia (DOSM) and data.gov.my are analysed across the years 2019 to 2022. Machine learning techniques, utilizing Python and SQL, are applied to model poverty levels, identify key contributing factors, and visualize patterns of regional disparity. The resulting model is integrated into an interactive Streamlit dashboard that enables users to explore poverty classification, feature importance, and tailored policy recommendations for targeted intervention.

The main objective of this study is to develop a machine learning-based model that accurately classifies poverty levels across Malaysian districts and identifies the underlying socio-economic factors contributing to poverty. By training and evaluating multiple machine learning models, such as decision trees, XGBoost, artificial neural networks (ANNs), k-nearest neighbors (k-NN), and Naive Bayes, the best model is selected based on its performance metrics. In addition to developing and deploying the most effective model for predicting poverty, the study also aims to propose actionable, data-driven solutions by mapping model-derived insights to practical policy recommendations.

METHODOLOGY

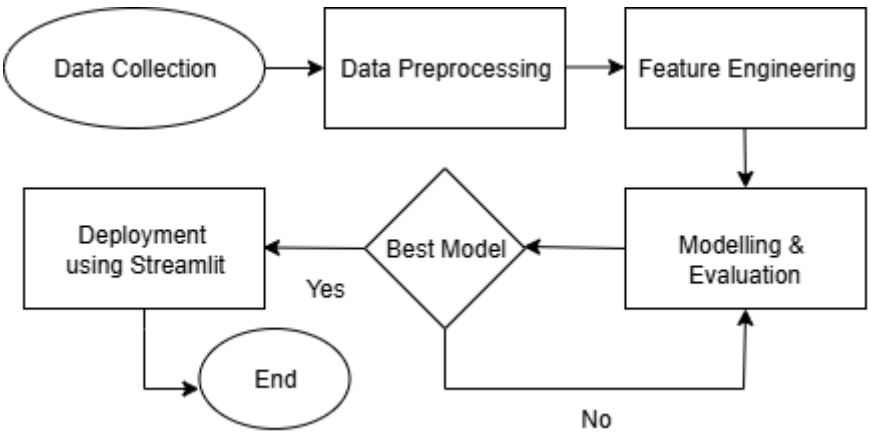


Figure 1: Flowchart of Methodology

The methodology involved a structured pipeline comprising data collection, preprocessing, feature engineering, modelling, and evaluation, followed by deployment using Streamlit, as shown in Figure 1. Data from six distinct sources covering poverty metrics, household income, labor force statistics, educational enrolment, completion rates, and inflation were cleaned, integrated using SQL, and preprocessed to handle missing values, outliers, and categorical encoding. The dataset was then split into training (70%) and testing (30%) sets, scaled using StandardScaler, and dimensionality was reduced via Principal Component Analysis (PCA) to improve model efficiency.

Five machine learning algorithms were implemented, such as decision tree, Naive Bayes, k-NN, XGBoost, and ANNs. Each model was evaluated based on accuracy, precision, recall, F1-score, confusion matrices, and ROC-AUC curves.

RESULTS AND DISCUSSION

XGBoost emerged as the top-performing model with the highest macro-average AUC score (0.93), as shown in Figure 2, demonstrating strong ability in classifying poverty categories and identifying influential factor features. Key factors affecting poverty included mean and median income, labor force participation rate, and regional disparities, underscoring the importance of economic stability and employment opportunities. The analysis of error metrics further reinforced the model's performance, with KNN recording the lowest Mean Absolute Error (MAE) at 0.3113 and Root Mean Squared Error (RMSE) at 0.6951, indicating high predictive accuracy. XGBoost achieves comparable MAE (0.3174) and RMSE (0.7046). However, XGBoost was selected for its superior interpretability and feature importance analysis, which is crucial for deriving actionable policy insights.

Model	Macro-Average AUC	AUC for Class 0	AUC for Class 1	AUC for Class 2
XGBoost	0.932858	0.962344	0.936109	0.900094
KNN	0.931568	0.962698	0.931874	0.900132
ANN	0.906288	0.949940	0.906541	0.862321
Decision Tree	0.883276	0.917135	0.887277	0.845405
Naive Bayes	0.718115	0.777946	0.742854	0.633466

Figure 2: Macro-Average Area Under the ROC Curve (AUC) of Each Model

The XGBoost predictive model was utilized to identify the key factors that influence poverty. By analyzing the feature importance scores derived from the model, we were able to determine the most significant variables contributing to the poverty levels. Figure 3 illustrates the top features, providing valuable insights into the main factors of poverty and laying the foundation for proposing actionable solutions. The top features influencing poverty are primarily centered around income levels, labor force dynamics, and regional disparities. Notably, the mean and median income emerged as the most critical features, with feature scores of 0.3137 and 0.3129, respectively, reflecting the central role of financial stability in poverty alleviation. Higher income levels directly correlate with reduced poverty, highlighting the need for targeted interventions aimed at increasing household income. The results clearly indicate that interventions should prioritise boosting household income, creating stable employment opportunities, and addressing regional inequities, particularly in states like Sabah and Kelantan. Furthermore, while education metrics were less dominant in the short-term predictive model, they remain vital for long-term poverty reduction strategies, highlighting the need for sustained investment in educational infrastructure and digital access.

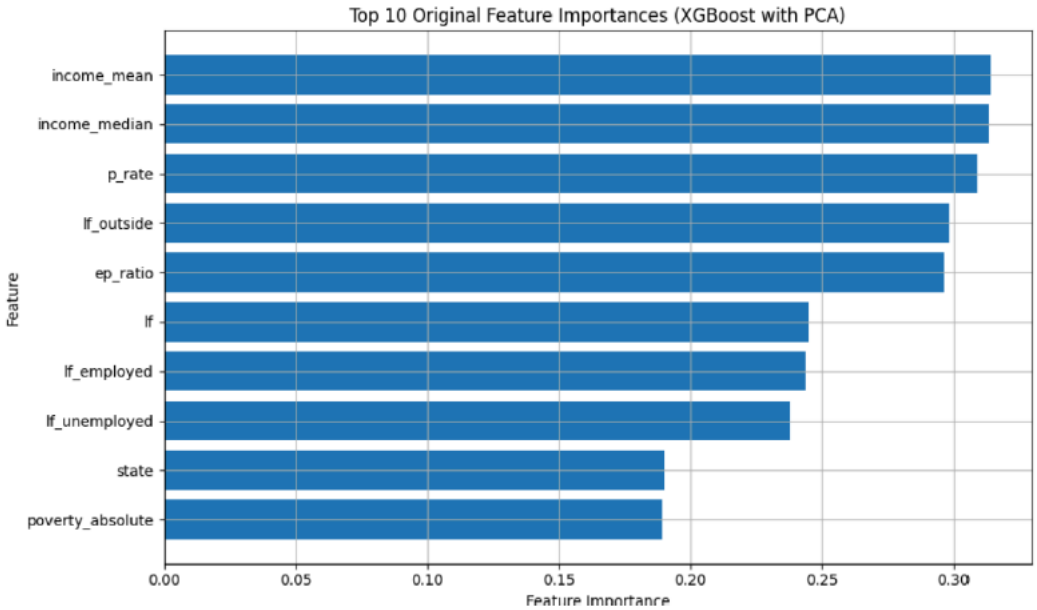


Figure 3: Top 10 Feature Importances from XGBoost with PCA

This comprehensive, data-driven approach successfully translates complex machine learning outputs into practical, targeted solutions, embodying the principles of evidence-based policymaking that are increasingly central to modern governance (Amazing Hope Ekeh et al., 2025). It provides a replicable and scalable framework for leveraging advanced analytics in social policy, ensuring that public and NGO resources are allocated efficiently and equitably to the regions and demographics where they are most

needed (Maidin et al., 2024). By moving beyond one-size-fits-all solutions to offer targeted, region-specific interventions such as tailored employment programs in Sabah and Kelantan or digital education infrastructure projects in rural areas, this methodology directly addresses the critical issue of regional disparities. The integration of diverse socioeconomic data sources creates a holistic view of poverty's multifaceted nature, a strategy that aligns with other data-centric policy frameworks developed in Malaysia (Abas et al., 2025). Ultimately, by enabling precise and proactive interventions, this research makes a significant contribution to fostering more sustainable and inclusive socioeconomic development across Malaysia, demonstrating the transformative power of machine learning in addressing entrenched societal challenges.

CONCLUSION

This study successfully demonstrated the application of machine learning in classifying poverty levels across Malaysian states and districts by integrating diverse socio-economic indicators. The XGBoost model, identified as the best-performing algorithm with an AUC of 0.93, enabled the extraction of key features such as household income, labour force participation, and education, which are central to understanding regional poverty disparities. The findings underscore the importance of data-driven approaches in supporting policymakers in designing targeted interventions, particularly in underdeveloped regions such as Sabah and Kelantan, where structural inequities persist.

The novelty of this work lies in the deployment of the model within an interactive Streamlit web application that not only visualises poverty insights but also provides state-specific recommendations on how to address the dominant factors of poverty. By linking feature importance analysis directly to actionable policy suggestions, the system moves beyond static predictions to offer practical guidance for filling poverty gaps in each state. This integration of explainable AI with policy-focused recommendations represents a unique and value-added contribution, as it transforms complex machine learning outputs into user-friendly, decision-support tools. Such an approach ensures that predictive analytics are not only accurate but also impactful in shaping equitable poverty alleviation strategies aligned with Malaysia's development goals.

REFERENCES

Abas, A., Ibrahim, M. M. Bin, & Azmi, N. H. (2025). Data-Driven Framework for Optimizing Internship Efficiency and Addressing Skills Mismatch in Malaysian Higher Education. *22nd International Learning and Technology Conference: Human-Machine Dynamics Fueling a Sustainable Future, L and T 2025*, 280–285. https://doi.org/10.1109/LT64002.2025.10940990

Amazing Hope Ekeh, Charles Elachi Apeh, Chinekwu Somtochukwu Odionu, & Blessing Austin-Gabriel. (2025). Advancing public safety and housing solutions: A comprehensive framework for machine learning and predictive analytics in urban policy optimization. *International Journal of Frontiers in Science and Technology Research*, *8*(1), 024–043. https://doi.org/10.53294/ijfstr.2025.8.1.0024

Maidin, M. R., Hussin, M. F., Zakaria, N. A. Z., Ab.Rahim, S. A. E., & Samsudin, K. (2024). A Comprehensive Framework for Crime Risk Assessment in Malaysia: Integrating Data-Driven Insights for Enhanced Public Safety. *International Journal of Academic Research in Business and Social Sciences*, *14*(11). https://doi.org/10.6007/ijarbss/v14-i11/23982

Md Shah, J., Hussin, R., & Idris, A. (2023). POVERTY ERADICATION PROJECT IN SABAH, MALAYSIA: NEW INITIATIVE, NEW CHALLENGES? In *Journal of the Malaysian Institute of Planners VOLUME* (Vol. 21).

Mohd Nor Radieah, & Khelghat-Doost Hamoon. (2019). 1AZAM Programme: The Challenges and Prospects of Poverty Eradication in Malaysia. *International Journal of Academic Research in Business and Social Sciences*, *9*(1), 345–356. https://doi.org/10.6007/ijarbss/v9-i1/5420