ISSN 2669-2481 / eISSN 2669-249X

BUSINESS, MANAGEMENT AND ECONOMICS ENGINEERING

2025
Volume 23
Issue 1
Pages 92–107

https://doi.org/10.3846/bmee.2025.20808

AN EXPLORATORY STUDY OF HIGH-EDUCATED POVERTY THROUGH MACHINE LEARNING APPROACH: A CASE STUDY OF EAST JAVA, INDONESIA

Dias SATRIA✉, Risti PERMANI, Kodrad WINARNO, David KALUGE, Citra Rahayu INDRASWARI, Radityo Putro HANDRITO

Department of Economics, Faculty of Economics and Business, Brawijaya University, Malang, Indonesia

Article History:
- received 17 January 2024
- accepted 10 December 2024

Abstract. *Purpose* – the purpose of the article is to compare the impotance factors for high-educated poverty modeling in the four cultural regions of East Java.

Research methodology – using data from the Indonesian National Survey, this study employs Random Forest (RF), Extreme Gradient Boosting (XGBoost), Artificial Neural Network (ANN), and K-Nearest Neighbor (KNN) algorithms as classification methods in machine learning.

Findings – the analysis results show that all agorithms has high accuracy with XGboost as the best performing model. The number of household members and the use of internet technology are among the most important variables affecting high-educated poverty in each cultural region in East Java.

Research limitations – the research was exclusively carried out in East Java, Indonesia, potentially restricting its applicability to other geographical areas or nations. Additionally, it relied on cross-sectional data, implying that a causal relationship between the independent and dependent variables cannot be inferred.

Practical implications – based on the results of this analysis, it is recommended that the government and related parties focus more on improving access to education and the internet in cultural areas in East Java.

Originality/Value – this study addresses a significant knowledge gap by exploring regional variations in the factors affecting high-educated poverty, emphasizing the importance of tailored strategies for each cultural region in East Java.

Keywords: higher education, machine learning, poverty.

JEL Classification: C69, I23, I32.

✉Corresponding author. E-mail: *dias.satria@ub.ac.id*

1. Introduction

Poverty remains a significant issue in human development (Nili, 2018). Various strategies have been implemented by governments to eradicate poverty, yet challenges persist. The COVID-19 pandemic has exacerbated global poverty levels, reversing over four years of progress and pushing an estimated 93 million people into extreme poverty in 2020 (United Nations, 2022). In 2021, the pandemic increased the global poverty rate from 7.8% to 9.1%, with about 97 million more people living on less than US$1.90 per day, disproportionately affecting the world's poorest (World Bank, 2023). Previous studies have investigated factors affecting poverty. Mora-Rivera and García-Mora (2021) found positive impacts on housing, food, health, and education for residents with internet access in Mexico. According to Human

Capital Theory by Gary Becker (Flores & Morejón, 2022), education serves as an investment in human capital, yielding substantial returns through higher income. Education influences mindset, attitude, and behavior, impacting economic decisions and providing a pathway out of poverty (Chankseliani & McCowan, 2021). Olopade et al. (2019) concluded that education and health are key determinants of economic growth. Through this few study it can be concluded that education is a critical factor in poverty alleviation.

Another major provision that capable in support poverty alleviation is technology, particularly for low- and lower-middle-income countries as highlighted by Lechman and Popowska (2022), even though the relationship remains uncertain for less developed countries (Galperin & Viecens, 2017). The increased use of technology post-pandemic necessitates updated research on its impact on poverty. Research on high-educated poverty is scarce. Datzberger (2018) studied educated Ugandans still classified as poor, attributing the issue to limited employment opportunities and misaligned school curricula with local economic needs.

Indonesia provides a relevant case study for this research. The national poverty rate declined from 24.2% in 1999 to 9.78% in 2020, but recent trends indicate stagnation (World Bank, 2023). Compared to other Southeast Asian countries like Malaysia and Thailand, with poverty rates of 0.4% and 6.1% respectively, Indonesia's rate remains high. Despite substantial economic growth, poverty eradication remains a critical policy focus. Each region has unique problem characteristics, especially the problem of poverty which is linked to social welfare (Zhou & Liu, 2022). This research focuses on East Java, a province with significant economic potential yet high poverty rates. As the second-largest province after Papua, has diverse economic sectors including industry, agriculture, fisheries, and tourism. However, the persistence of high-educated poverty in East Java suggests that education alone is insufficient for poverty eradication, indicating the need for a deeper exploration of contributing factors. The evidence is provided by Indonesia Statistics (BPS-Statistics Indonesia, 2021), East Java had a poverty rate of 11.40% in March 2021, higher than West Java's 8.4%. Notably, a unique feature in East Java is "high-educated poverty," where individuals with secondary education live below the poverty line. In 2020, 19.5% of the poor in East Java had completed 12 years of compulsory education (BPS-Statistics Indonesia, 2021).

This study aims to compare the importance of factors influencing high-educated poverty across the four cultural regions of East Java in 2021. Specifically, it addresses the research question: "Do different cultural regions show differences in the factors influencing high-educated poverty in East Java?" To answer this question, the study utilizes data from the Indonesian National Survey and employs machine learning algorithms – Random Forest (RF), Extreme Gradient Boosting (XGBoost), Artificial Neural Network (ANN), and K-Nearest Neighbor (KNN) – as classification methods. Machine learning facilitates the analysis of large, complex data without requiring specific assumptions. Previous studies have used these techniques for various topics such as GDP growth forecasting (Yoon, 2021), credit card fraud detection (Itoo et al., 2021), and financial risk evaluation (Guo, 2023). This research contributes to the literature by highlighting high-educated poverty in Indonesia, employing machine learning to model influencing factors, and providing insights for policymakers to craft targeted interventions. Additionally, this study addresses the knowledge gap by exploring regional variations in these factors, emphasizing the need for tailored strategies in East Java.

The remainder of this article is organized into five sections. Section 1 describes the high-educated poverty in East Java. Section 2 reviews the literature on theoretical foundations and previous research on high-educated poverty and digitalization in Indonesia and globally. Section 3 explains the methodology used in this study. Section 4 discusses the results of the analysis and explains the types of digitalization that most importance to the model of high-educated poverty in East Java. The final section presents the concluding remarks.

2. Literature review

In terms of poverty, Hofmarcher (2021) found that the education sector plays an important role in poverty alleviation programs in Europe. Education plays a pivotal role in poverty alleviation by serving as both a direct and indirect pathway to improved economic opportunities and quality of life. Higher educational attainment often leads to better job prospects, higher wages, and increased economic stability. Education emerges as a crucial factor in poverty alleviation through multiple dimensions. The research highlights that enhanced levels of education, particularly compulsory and primary education, have a substantial impact on reducing poverty (Spada et al., 2023). Similarly, the study focusing on rural China underscores that different levels of education, especially primary and junior secondary education, play distinct roles in alleviating rural poverty (Liu et al., 2023). The spatial dynamics revealed that regional educational improvements not only benefit the immediate area but also influence neighbouring regions, emphasizing the broader impact of education on poverty reduction. These findings collectively affirm that investment in education fosters economic opportunities, supports long-term poverty alleviation strategies, and contributes to a more equitable distribution of resources and opportunities.

In opposite of Hofmarcher (2021), Spada et al. (2023), and Liu et al. (2023) study, Meo et al. (2020) work gave a fresh idea about the non-linear relationship between unemployment, governance, and poverty in Pakistan. Long-term unemployment will create too much immoral crime in countries such as frustration, homelessness, family tension, loss of confidence, social isolation, self-esteem, and poverty (Siddiqa, 2021). The phonemenon of high-educated poverty presents a paradox where individuals with advanced educational qualifications remain trapped in poverty despite their higher levels of education. This situation often arises due to a mismatch between educational attainment and job market demands, where highly educated individuals face unemployment or underemployment in roles that do not utilize their skills or provide adequate compensation. Factors contributing to high-educated poverty include economic downturns, structural changes in the labour market, and over-saturation of certain professions, which can dilute the value of higher education degrees. In addition, systemic issues such as inadequate social safety nets and disparities in access to quality education and professional opportunities exacerbate this problem. Consequently, while education is a powerful tool for poverty alleviation, its effectiveness can be undermined by economic and structural barriers that fail to ensure meaningful employment and fair wages for all educated individuals.

From globalization perspective, several previous studies have examined poverty, showing that digital technology and internet access are among the things that affect poverty (Barbero

& Rodríguez-Crespo, 2022; Gautam et al., 2022; Lechman & Popowska, 2022). In the current era, technology is developing so rapidly that it affects daily human activities (Burov et al., 2020). The emergence of gadgets and internet access has changed the way people interact, access information, shop, seek entertainment, and even manage aspects of daily life such as banking and payments. Access to these technologies not only changes lifestyles but also provides new opportunities, especially in terms of economics and employment. Gadget ownership and internet access has become more than just a trend, but also a tool that can help poorer sections of society to access economic and educational opportunities. Based on this with the use of meta data across various countries, research by Si et al. (2020) has analyzed the relationship between digital technology and poverty. Digital technology is considered to form a new solution in poverty alleviation efforts because business actors both companies and entrepreneurs utilize digital technology in economic activities. Zahra et al. (2023) in their research stated that digital technology and the active role currently played by new businesses illustrate a complex evolutionary process.

The implication of technology (digitalization) plays a crucial role in reducing poverty by improving access to information, services, and economic opportunities. As technology advances, it provides new ways for individuals to participate in the economy, such as through online education, remote work, and e-commerce, which can create income opportunities and enhance livelihoods. Furthermore, digital tools can improve efficiency in various sectors, including agriculture and healthcare, by providing better market information, increasing productivity, and facilitating access to essential services. Studies have shown that countries with higher levels of digital development tend to experience lower poverty rates, as technology can bridge gaps in traditional infrastructure and empower marginalized communities (Lechman & Popowska, 2022). Thus, fostering digital inclusion and investing in technological infrastructure can be effective strategies for alleviating poverty and promoting sustainable development.

In Indonesia, there already studies have investigated the determinants and nature of poverty as it relates to poverty reduction, unemployment, and national economic growth supported as literature review on poverty. For example, Erlando et al. (2020) found a relationship between economic growth, inequality and poverty. Investment has a direct effect on poverty, while economic growth has no direct effect on poverty. Poverty in Indonesia is also caused by many other factors, including populatin growth, investment, education, health, market structure, and government regulation (Jacobus et al., 2018).

Various perspectives and data analysis methods are applied to build poverty models with the hope of describing poverty as well as possible. Analytical methods that can be used in poverty modeling include correlation analysis (Jiang et al., 2020), panel data regression analysis (Omar & Inaba, 2020) and logistic regression analysis (Dogan et al., 2022). Currently, there are many developments of analysis methods that are more flexible on complex data. One of them is machine learning which is considered more effective than the use of statistical learning methods in handling large and complex data cases. In addition, the advantage of machine learning is the freedom from assumptions that sometimes accompany statistical modeling, so that the model produced by machine learning has a high level of accuracy (Hu et al., 2019, 2022). The use of machine learning in poverty modeling has also been applied by Yao et al.

(2023) with deep learning and Alsharkawi et al. (2021) who analyzed multidimensional poverty with LightGBM and Bagged Decision Tree.

Based on previous research, this research use several machine learning algorithms in modeling high-educated poverty in East Java, namely RF, XGBoost, ANN, and KNN. Previous research by Uyen and Thu (2023) using several two MCDM (Multi-Criteria Decision Making) the FUCA (Faire Un Choix Adéquat) method and the CURLI (Collaborative Unbiased Rank List Integration) is capable to provide a best decision of education supporting equipment for teachere in classrooms. The choosen machine learning algorithms will bring the best insight of methode that capable accessing high-educated poverty insight in Indonesia

3. Data source and methodology

The data used in study is secondary Indonesian National Survey done by Indonesia statistics (*BPS / Badan Pusat Statistik*) in 2021. As describe in data document, respondent of this data was obtained from a survey in East Java with the provision of having received at least a high school education with a total of 30,719 respondents with the age of 18–60 years old which is the working age in Indonesia. East Java is chosen in this study due two reasons: 1) East Java is known as of a tenacious area in Indonesia with agriculture, industry, and trade being the main economic pillars of the province; 2) East Java population density in Indonesia is second after West Java, it makes the area more vulnerable to economic problems, including poverty.

According to Arzaqi and Astuti (2019), East Java is divided into four cultural regions that highlighted in this research, namely the Arek cultural region, Pandalungan cultural region, Madura cultural region, and Mataraman cultural region as shown in Figure 1. The people of the Arek region are identical to urban areas so they are known to be more open to change, adaptable, determined, and have high solidarity. The people of the Madura region are characterized by their love of entrepreneurship/trading and have a high spirit of mobility in an effort to increase their income. In other words, the people of the Madura region have a high migratory spirit and have a high spirit. The Pandalungan community is known as a fusion

Figure 1. The distribution of cultural areas in East Java
(author's own conception, based on Tableau software)

between Madurese and Javanese cultures. Pandalungan society is often referred to as a hybrid society. Their characteristics are hardworking, aggressive, and expansive. Meanwhile, the community is known as a community that prefers to settle in safe conditions, so they prefer to become civil servants and lack entrepreneurial spirit.

This research has categorized people who are educated, namely someone with a high school, vocational school, and university education. The variable used in this research is poverty status as the dependent variable. Poverty is a condition when a person or group of people has limitations in fulfilling their basic rights to develop and maintain a dignified (Adji et al., 2020). The example of basic needs is food, clothing, housing, education, and health services. BPS measures poverty with reference to the Poverty Line or Minimum Poverty Line (MPL), which is the threshold value of income needed to fulfill these basic needs. The poverty line in Indonesia is calculated by adding together the minimum value of food expenditure and the minimum value of non-food expenditure. The data is categorized into 2 categories, namely 0 for non-poor where a person is above the poverty line and 1 for poor where a person is below the poverty line. For independent variable is demographical characteristics (i.e age, gender, marital status, number of household members, and disability), regional aspects (i.e life region and migrant status), and digitalization (i.e internet access and using technology) High-educated poverty is a condition where a person has completed secondary education but is still classified as poor. Table 1 summarizing the variables to be observed in this study.

The method used in this research is quantitative through machine learning approach. Machine learning is a branch of artificial intelligence (AI) that focuses on developing algorithms and computer models that can learn from data and experience without having to be explicitly programmed (Satria, 2023). This approach allows machines to identify patterns,

Table 1. Characteristic of variables (author's documentation)

Variable	Category
Age	18–60 years old
Gender	0: Male 1: Female
Marrital status	0: Others 1: Marriage
Disablility	0: No 1: Yes
Migran status	0: No 1: Yes
Internet access	0: No 1: Yes
Using technology	0: No 1: Yes
Life region	0: Urban areas 1: Rural areas
Poverty status	0: Affluent 1: Poor

make predictions, and make decisions based on given data. In general, there are three types of machine learning: supervised learning, unsupervised learning, and reinforcement learning (Astutik et al., 2021). In supervised learning, the model learns from pre-labeled data, which means that each data instance has a label that states the desired outcome. Supervised learning is the type of machine learning used in this research because it has labeled the desired decision, which is the poor status of educated households in East Java.

The data is processed using R and vizualize using Tableau Software. There are four machine learning models to determine high-educated poverty status. First is Random Forest (RF), this algorithm is effective for classification, regression, and other tasks. It is based on an "ensemble learning" technique that combines many decision trees to produce more accurate and stable predictions. Each tree in RF is randomly constructed with a randomly selected subset of training data and uses only a random subset of features (attributes) for decision-making at each node (Yoon, 2021). By utilizing this variation, RF overcomes the problem of overfitting and reduces bias, thereby improving the generalization performance of the model. This RF machine learning algorithm is very popular for classification and regression purposes. In this study, we have used it for classification purposes. There are three phases in classification with (Nayeem et al., 2021). In the first phase, a forest of decision trees is generated from a large number of trees. In the second phase, the trees used to create the forest predict a class name. In the third phase, the correct class name is assigned to the test data based on a majority vote.

Second is Extreme Gradient Boosting (XGBoost), which is known as machine learning algorithm with the category of supervised learning tasks that have good performance and can be used for regression and classification (Ibrahem Ahmed Osman et al., 2021). XGBoost uses an ensemble of decision trees whose goal is to minimize an objective function using a gradient-based optimization method. The objective function in Xgboost is a combination of loss and regularization functions. The predicted value at the i-th node is shown in the following Equation (1).

$$\hat{y}_i = \sum_{m-1}^{M} f_m\left(x_i\right). \tag{1}$$

The prediction of the i - th value is symbolised by $\hat{y}_i$, M is the number of iterations, and $f_m\left(x_i\right)$ is the prediction of the i-th data. Furthermore, the objective function to be minimized is shown in Equation (2).

$$Obj = \sum_{n-1}^{n} l\left(y_i, \hat{y}_i\right) + \sum_{m-1}^{M} \Omega f_m, \tag{2}$$

n is the number of training data, y_i is the actual target value for the i-th data, $\hat{y}_i$ is the prediction at the last node for the i-th data, l is the loss function, Ω is the regularization function, and f_m is the m-th model in the ensemble. The regularization function can be written through the following Equation (3).

$$\Omega\left(f\right) = \gamma T + \frac{1}{2}\lambda \sum_{j-1}^{T} w_j^2 + \alpha \sum_{j-1}^{T} \left|w_j\right|. \tag{3}$$

Where T is the number of nodes in the decision tree, w_j is the weight at the j-th node, γ, λ, α are hyperparameters that control the degree of regularization.

Third algorithm is Artificial Neural Network (ANN), that is inspired by human neural networks. This model consists of several interconnected components called neurons. The neurons form the input layer, hidden layer, and output layer (Djurović et al., 2024). The training process is carried out with a backpropagation or resilient propagation algorithm (Astutik et al., 2021).The network makes predictions on the training data and calculates the prediction error. The error is then passed back through the network to update the weights and bias, so that the network can learn to recognize patterns in the data (Patil et al., 2022). An activation function is applied to the output of each neuron to introduce non-linearity and allow the network to understand more complex patterns. With repeated iterations, neural networks can learn patterns and perform tasks such as classification, regression, and more complex tasks such as deep learning when dealing with deep neural networks with many hidden layers.

Last is K-Nearest Neighbors (KNN), the simple yet effective machine learning algorithms in handling classification and regression problems. In the context of classification, KNN finds the k-nearest neighbors of the new data to be predicted based on euclidean distance or other metrics. This distance measures how similar the new data is to the existing training data in the feature space. KNN then chooses the majority of labels from the k-nearest neighbors as the prediction for the new data. Basically, KNN does not perform a learning process like other machine learning algorithms. Instead, the KNN model stores the entire training dataset in memory and performs distance comparisons for each new incoming data. The number of neighbors (k) selected can be adjusted as needed, and choosing the right k value can have a significant impact on the model's performance.

To reach the conclusion of research data result, each machine learning algorithm is evaluated with several methods to determine the performance of the model obtained. The performance of the machine learning algorithm shows how good the algorithm is in classifying the status of the educated poor in East Java. A confusion matrix is a table that is often used to describe the performance of a classification algorithm. It summarizes the results of a classification task by showing the counts of true positive (TP), false positive (FP), true negative (TN), and false negative (FN) predictions. Each cell in the matrix provides insight into how many instances were correctly or incorrectly classified for each class. Some of the performance evaluation methods in this research are accuracy (4), recall (5), precision (6) and F1-score (7).

$$Accuracy\left(Acc\right)=\frac{TP+TN}{TP+TN+FP+FN}; \tag{4}$$

$$Recall\left(Re\right)=\frac{TP}{TP+TN}; \tag{5}$$

$$Precision\left(Pre\right)=\frac{TP}{TP+FP}; \tag{6}$$

$$F1-Score=2\frac{\left(Pre\right)\left(Re\right)}{Pre+Re}. \tag{7}$$

Where accuracy (4) is a metric that measures how many model predictions are correct compared to the overall data. Recall (5) formula is to measure the proportion of actual positive cases that were correctly identified. Precision (6) measures how many of the cases predicted as positive were actually positive. Lastly F1-score (7) provides a single score that balances recall and precision metrics.

4. Results and discussion

This research discusses high-educated poverty based on cultural areas in East Java Province using several machine learning algorithms. Figure 2 presents an overview of high-educated poverty in East Java, highlighting that the causes of poverty extend beyond low education to include other factors.

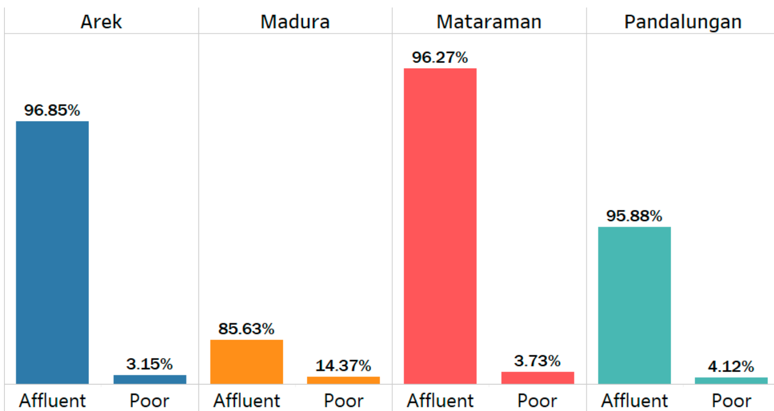

Figure 2. High-educated poverty in East Java
(author's own conception, based on Tableau software)

Figure 3 shows that the highest percentage of high-educated poverty is in the Madura Cultural Region at 14.37%, while the Arek region has the lowest at 3.15%. These variations suggest that cultural characteristics influence high-educated poverty, necessitating a regional analysis to identify the most critical factors for poverty eradication in each cultural area.

High-educated poverty in the Arek, Pandalungan, Mataraman, and Madura cultural regions was classified using RF, XGBoost, ANN, and KNN algorithms, as described in Table 2. The result indicates that each cultural region's data yields is totally different from each other. But, the result is obvious with XGBoost algorithm that consistently outperformed the other models, achieving over 90% accuracy except for the Madura region, where it was 84%. On the contrary RF, KNN, and ANN model performance more inferior, particulary ANN methode is revealed the lowest performance across all regions. It also indicates that each machine learning model performance calculates different values in each cultural region in East Java, and enlighten the possibility that different data will give different performances even processed with the same algorithm.

Table 2. Model performance (author's own conception, based on R software)

	MACHINE LEARNING MODEL			
	Xgboost	RF	KNN	NN
Culture Region of Madura:				
Accuracy	0.841	0.825	0.803	0.786
Precision	0.932	0.886	0.889	0.942
Recall	0.948	0.923	0.902	0.851
F1 Score	0.911	0.9	0.888	0.871
Culture Region of Arek				
Accuracy	0.961	0.955	0.947	0.882
Precision	0.987	0.971	0.973	0.991
Recall	0.99	0.985	0.975	0.9
F1 Score	0.98	0.977	0.973	0.936
Culture Region of Mataraman				
Accuracy	0.9556	0.941	0.937	0.872
F1 Score	0.981	0.97	0.968	0.93
precision	0.977	0.965	0.966	0.988
recall	0.992	0.976	0.97	0.896
Culture Region of Pandalungan				
Accuracy	0.954	0.95	0.93	0.899
F1 Score	0.976	0.974	0.963	0.946
precision	0.973	0.963	0.966	0.986
recall	0.991	0.987	0.963	0.926

Another finding is the result indicates the differences in model performance in each cultural region may affected by important variables that shaping poverty in each cultural region are different. Figure 3 is provided as evidence, that sho the order of the most important variables that contribute to the occurrence of high-educated poverty in the cultural regions

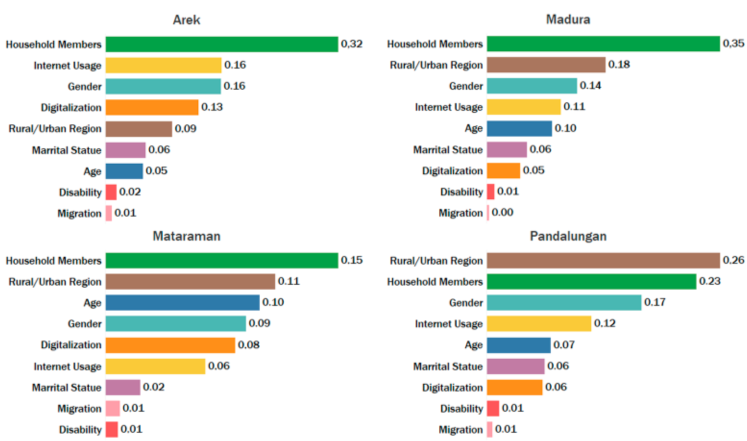

Figure 3. Variable importance based on impurity decrease (author's own conception, based on Tableau software)

of East Java is different in each cultural region. These results indicate the complexity of the interaction of variables that play a role in determining the level of high-educated poverty in each cultural region. The most important variable affecting high-educated poverty in the Arek, Madura, and Mataraman regions is the number of household members. Meanwhile, in the Pandalungan cultural region, the variable of many household members is in second place after the variable of region of residence. This shows how strong the variable of many household members is in influencing high-educated poverty in East Java.

The varying model performances suggest that different variables shape poverty in each cultural region. Figure 3 illustrates these variables, revealing the complexity of high-educated poverty determinants. The number of household members is the most important variable in the Arek, Madura, and Mataraman regions, while in Pandalungan, rural/urban region is the most influence variable. The result is inline with Weldearegay et al. (2021), which identified family size as a critical factor in farmer poverty in Ethiopia, affecting the region economic resources. The prominence of the region of residence in Pandalungan underscores disparities in access to education and economic opportunities. This variable's significance points to the need for targeted interventions to address regional inequalities.

From a policymaker's perspective, these findings indicate the need for region-specific interventions by local goverment. In regions where family size is dominant, policies should focus on family planning and economic support for large households. In areas where residence is more critical, improving infrastructure and access to services should be prioritized. For society, addressing high-educated poverty requires a multifaceted approach, considering individual and community needs. Employers and employees could benefit from policies enhancing digital skills and technology access, thereby improving employment opportunities.

The order of the most important variables within each cultural region can be explained through the perspective of several aspects. First, the social, economic and cultural factors underlying each cultural region have different influences in shaping the pattern of high-educated poverty. This can be caused by differences in population characteristics, accessibility to education services, levels of social mobility, and other socioeconomic aspects unique to each cultural region.

In the Arek cultural region, the results of the analysis show that there are four most important variables for high-educated poverty: number of household members, internet use, gender, and use of digital technology. The number of household members ranked first in influencing high-educated poverty can reflect the level of limited economic resources available. The more household members, the more difficult it is for families to allocate funds for education, which in turn can hinder access to quality education. Next is the variable of internet and digital technology use plays an important role in the context of current technological and information developments. The use of the internet and digital technology can open up opportunities for access to online learning resources and information in developing soft skills. Especially in the Arek culture area, which is known for its dynamic and adaptive spirit, the utilization of this technology can be a bridge to improve individual competencies in various aspects of life, including interpersonal skills, problem solving, and creativity. As a community that tends to be open to change and innovation, Arek culture residents can benefit from access to information and online learning platforms, which in turn can support the strengthening of their soft skills in the face of evolving global demands. Moreover, the use of the internet and technology has become a widespread lifestyle in various walks of life today (Marques et al., 2019). Even the significance of the pinnacle of scientific innovation lies in the incorporation of the conceptual, technological, and contextual framework of the internet as well as the utilization of internet technology (Roblek et al., 2020).

The gender factor that emerges as the third most important variable also indicates the existence of gender disparities in high-educated poverty in Arek culture areas. There may be differences in educational opportunities between men and women that could affect the level of high-educated poverty. These reasons may include social, cultural and economic factors that can affect educational accessibility and participation between the sexes. In the context of Arek culture, known for its spirit of inclusiveness and equality, the roles of women and men in daily life are often seen as equal. However, in social reality, aspects of tradition and norms that have not been fully resolved may still have an influence on access to education between the sexes. Therefore, an in-depth understanding of the cultural dynamics and social changes in the Arek region is crucial in designing efforts that can reduce gender disparities in high-educated poverty and support equitable access to education for all citizens.

In the Madura cultural region, the analysis shows that there are four variables that have a significant influence on high-educated poverty, namely the number of household members, region of residence (village/city), gender, and internet usage. The role of the region of residence variable shows that there are differences in access to educational facilities and opportunities between urban and rural areas in the Madura cultural region. In urban areas, there may be better access to schools, learning resource centers and additional educational opportunities. While in rural areas, geographical and infrastructural challenges may limit accessibility to education and employment. In addition, gender also emerges as an important variable in describing the social aspects related to high-educated poverty in Madurese cultural areas. Madurese culture, which generally emphasizes equality and the active role of women in various fields of life, may still face challenges in providing equal educational opportunities for men and women. This may be due to traditional factors and social norms that can influence educational choices and career opportunities between the sexes. Another most important variable is internet that shows the role of information technology and access to learning resources in overcoming geographical and resource limitations within Madura. Although some areas may have infrastructure limitations, the characteristics of Madurese people who are known for their adaptive and innovating spirit may support the adoption of digital technology to support education.

In the Mataraman cultural region, the variable area of residence plays an important role in terms of accessibility to education facilities and services. Differences between urban and rural areas within the Mataraman cultural region may affect the availability and quality of education available. Geographical and infrastructural factors may play a role in limiting access to education in rural areas. The age and gender factors that emerged as important variables also reflect social and demographic aspects that are relevant in the context of high-educated poverty in the Mataraman cultural area. It affect educational and economic opportunities, with education perhaps favored at a young age and differences in opportunities based on gender. Regarding the absence of internet and digitalization as the most important variables, this may be explained by the characteristics of the Mataraman people who may still tend to rely on traditional access to daily activities, that justify the absence of this variable in the main factors affecting poverty. Limited technological infrastructure and low internet accessibility may also play a role in explaining why internet and digitalization are not included in the four main variables.

Lastly, in the Pandalungan cultural region, the importance of the region of residence variable in influencing poverty in the Pandalungan cultural region can be seen through the perspective of community characteristics. The areas is posses a few different types of settlements, namely rural and urban, may have significant differences in terms of access to education services, employment, and other economic resources. In general, rural communities may face challenges in terms of accessibility to educational facilities and economic opportunities, while urban communities may have better access to educational infrastructure and services as well as diverse employment opportunities. In the context of the community characteristics of the Pandalungan region, the existence of rural communities that may rely on the agricultural sector and have limited access to infrastructure may affect the level of high-educated poverty. On the other hand, in urban areas, more job opportunities and access to formal education may help reduce poverty rates. Therefore, an in-depth understanding of the differences between rural and urban settlements in the Pandalungan region is crucial in designing intervention strategies to address high-educated poverty. The dominating variable of region of residence in poverty factors in the Pandalungan cultural region may also illustrate the other factors such as age, gender, and education may have a more equal influence across this region. In this case, the emphasis on geographical factors as the most important variable indicates the significant role of the social and economic environment in shaping the level of high-educated poverty in the Pandalungan cultural region.

5. Conclusions

The modeling of high-educated poverty in this study is done with several machine learning algorithms. This reserach discover Xgboost is the best performing models of high-educated poverty in East Java. Analysis results show that the use of RF, Xgboost, ANN, and KNN has high accuracy with Xgboost as the best performing model. The result enlighted the most affected variable towards high-educated poverty in East Java is the number of household members and the use of internet technology. The number of household members can affect the poverty rate because the limited resources that must be fulfilled for each family member. If there are many household members, resources such as education costs, health care, and other living needs must be divided and reduce the ability to fulfill adequate education needs. Aside of number of household members, the use of internet has a major impacts of high-educated poverty in current lifestyle. Internet use capable to provides opportunities and makes it easier to find information, broaden horizons, and develop skills, which can increase competitiveness and employment opportunities for graduates of secondary education and above. However, it is important to remember that internet access must be evenly distributed across all cultural areas in East Java so that all residents have the same opportunity to utilize this internet. Another important discovery is gender and urban-rural dispartities to grab educational opportunities itself are significant, but not oblivious in every region.

Based on the results of this analysis, it is recommended that the government and related parties focus more on improving access to education and the internet in cultural areas in East Java. Additionally, digital skills training programs and equitable provision of technology infrastructure is also needed to help reduce disparities and provide more equitable opportunities for graduates of secondary education and above to overcome poverty. In the near future, it is hoped that a more inclusive and competitive society can be created, that impacted on reduction in poverty among people who have completed secondary education in the cultural areas of East Java.

There are several limitations that need to be acknowledged when interpreting the results of this study. The research was exclusively carried out in East Java, Indonesia, potentially restricting its applicability to other geographical areas or nations. Additionally, it relied on cross-sectional data, implying that a causal relationship between the independent and dependent variables cannot be inferred. As a result, these limitations need to be considered when interpreting the findings of this study. Future research is suggested to apply the research in different provinces and do a comparative study to broaden the research result that incapable discovered by this study.

References

Adji, A., Hidayat, T., Tuhiman, H., Kurniawat, S., & Maula, A. (2020). *Measurement of poverty line in Indonesia: Theoretical review and proposed improvements* (Working Paper No. 48-e). TNP2K. https://tnp2k.go.id/download/88787WP%2048_Measurement%20of%20Poverty%20Line%20in%20Indonesia-Theoretical%20Review%20and%20Proposed%20Improvements.pdf

Alsharkawi, A., Al-Fetyani, M., Dawas, M., Saadeh, H., & Alyaman, M. (2021). Poverty classification using machine learning: The case of Jordan. *Sustainability*, *13*(3), Article 1412. https://doi.org/10.3390/su13031412

Arzaqi, R. S., & Astuti, E. T. (2019). Study of income inequality in east Java in 2010–2017. *National Seminar on Official Statistics*, *2019*(1), 514–523. https://doi.org/10.34123/semnasoffstat.v2019i1.195

Astutik, S., Pramoedyo, H., Rahmi, N. S., Irsandy, D., & Damayanti, R. H. P. Y. (2021). Rainfall data modeling with artificial neural networks approach. *Journal of Physics: Conference Series*, *2123*, Article 012029. IOP Publishing. https://doi.org/10.1088/1742-6596/2123/1/012029

Barbero, J., & Rodríguez-Crespo, E. (2022). Technological, institutional, and geographical peripheries: Regional development and risk of poverty in the European regions. *The Annals of Regional Science*, *69*(2), 311–332. https://doi.org/10.1007/s00168-022-01127-9

BPS-Statistics Indonesia. (2021). *The percentage of poor people in East Java in March 2021 reached 11.40 percent*. https://jatim.bps.go.id/en/pressrelease/2021/07/15/1233/the-percentage-of-poor-people-in-east-java-in-march-2021-reached-11-40-percent.html

Burov, O., Bykov, V., & Lytvynova, S. (2020). ICT evolution: From single computational tasks to modeling of life. *CEUR Workshop Proceedings*, *2732*, 583–590.

Chankseliani, M., & McCowan, T. (2021). Higher education and the sustainable development goals. *Higher Education*, *81*, 1–8. https://doi.org/10.1007/s10734-020-00652-w

Datzberger, S. (2018). Why education is not helping the poor. Findings from Uganda. *World Development*, *110*, 124–139. https://doi.org/10.1016/j.worlddev.2018.05.022

Djurović, S., Lazarević, D., Ćirković, B., Mišić, M., Ivković, M., Stojčetović, B., Petković, M., & Ašonja, A. (2024). Modeling and prediction of surface roughness in hybrid manufacturing–milling after FDM using artificial neural networks. *Applied Sciences*, *14*(14), Article 5980. https://doi.org/10.3390/app14145980

Dogan, E., Madaleno, M., Inglesi-Lotz, R., & Taskin, D. (2022). Race and energy poverty: Evidence from African-American households. *Energy Economics*, *108*, Article 105908. https://doi.org/10.1016/j.eneco.2022.105908

Erlando, A., Riyanto, F. D., & Masakazu, S. (2020). Financial inclusion, economic growth, and poverty alleviation: Evidence from eastern Indonesia. *Heliyon*, *6*(10), Article e05235. https://doi.org/10.1016/j.heliyon.2020.e05235

Flores, K. M. G., & Morejón, V. M. M. (2022). Fundamentals of human capital and education elements of the entrepreneurship ecosystem. *CrossCultural Management Journal*, *2022*(2), 149–157.

Galperin, H., & Viecens, M. F. (2017). Connected for development? Theory and evidence about the impact of Internet technologies on poverty alleviation. *Development Policy Review*, *35*(3), 315–336. https://doi.org/10.1111/dpr.12210

Gautam, R. S., Rastogi, S., Rawal, A., Bhimavarapu, V. M., Kanoujiya, J., & Rastogi, S. (2022). Financial technology and its impact on digital literacy in India: Using poverty as a moderating variable. *Journal of Risk and Financial Management*, *15*(7), Article 311. https://doi.org/10.3390/jrfm15070311

Guo, Y. (2023). Research on the application of ANN in enterprise financial risk evaluation information system with digital empowerment. *Proceedings of the 3rd International Conference on Economic Development and Business Culture* (*ICEDBC 2023*), 672–680. https://doi.org/10.2991/978-94-6463-246-0_81

Hofmarcher, T. (2021). The effect of education on poverty: A European perspective. *Economics of Education Review*, *83*, Article 102124. https://doi.org/10.1016/j.econedurev.2021.102124

Hu, L., He, S., Han, Z., Xiao, H., Su, S., Weng, M., & Cai, Z. (2019). Monitoring housing rental prices based on social media: An integrated approach of machine-learning algorithms and hedonic modeling to inform equitable housing policies. *Land Use Policy*, *82*(129), 657–673. https://doi.org/10.1016/j.landusepol.2018.12.030

Hu, S., Ge, Y., Liu, M., Ren, Z., & Zhang, X. (2022). Village-level poverty identification using machine learning, high-resolution images, and geospatial data. *International Journal of Applied Earth Observation and Geoinformation*, *107*, Article 102694. https://doi.org/10.1016/j.jag.2022.102694

Ibrahem Ahmed Osman, A., Ahmed, A. N., Chow, M. F., Huang, Y. F., & El-Shafie, A. (2021). Extreme gradient boosting (Xgboost) model to predict the groundwater levels in Selangor Malaysia. *Ain Shams Engineering Journal*, *12*(2), 1545–1556. https://doi.org/10.1016/j.asej.2020.11.011

Itoo, F., Meenakshi, & Singh, S. (2021). Comparison and analysis of logistic regression, Naïve Bayes and KNN machine learning algorithms for credit card fraud detection. *International Journal of Information Technology*, *13*(4), 1503–1511. https://doi.org/10.1007/s41870-020-00430-y

Jacobus, E. H., Kindangen, P., & Walewangko, E. N. (2018). Analisis faktor-faktor yang mempengaruhi kemiskinan rumah tangga di Sulawesi Utara [Analysis of factors affecting household poverty in North Sulawesi]. *Journal of Regional Economic and Finance Development*, *19*(3). https://doi.org/10.35794/jpekd.19900.19.7.2018

Jiang, L., Wang, H., Tong, A., Hu, Z., Duan, H., Zhang, X., & Wang, Y. (2020). The measurement of green finance development index and its poverty reduction effect: Dynamic panel analysis based on improved entropy method. *Discrete Dynamics in Nature and Society*, *2020*, Article 8851684. https://doi.org/10.1155/2020/8851684

Lechman, E., & Popowska, M. (2022a). Harnessing digital technologies for poverty reduction. Evidence for low-income and lower-middle income countries. *Telecommunications Policy*, *46*(6), Article 102313. https://doi.org/10.1016/j.telpol.2022.102313

Liu, W., Li, J., & Zhao, R. (2023). The effects of rural education on poverty in China: A spatial econometric perspective. *Journal of the Asia Pacific Economy*, *28*(1), 176–198. https://doi.org/10.1080/13547860.2021.1877240

Nili, S. (2018). Global poverty, global sacrifices, and natural resource reforms. *International Theory*, *11*, 48–80. https://doi.org/10.1017/S1752971918000209

Marques, G., Pitarma, R., Garcia, N. M., & Pombo, N. (2019). Internet of things architectures, technologies, applications, challenges, and future directions for enhanced living environments and healthcare systems: A review. *Electronics*, *8*(10), Article 1081. https://doi.org/10.3390/electronics8101081

Meo, M. S., Kumar, B., Chughtai, S., Khan, V. J., Dost, M. K. B., & Nisar, Q. A. (2020). Impact of unemployment and governance on poverty in Pakistan: A fresh insight from non-linear ARDL co-integration approach. *Global Business Review*, *24*(5), 1007–1024. https://doi.org/10.1177/0972150920920440

Mora-Rivera, J., & García-Mora, F. (2021). Internet access and poverty reduction: Evidence from rural and urban Mexico. *Telecommunications Policy*, *45*(2), Article 102076. https://doi.org/10.1016/j.telpol.2020.102076

Nayeem, M. J., Rana, S., Alam, F., & Rahman, M. A. (2021, February 27–28). Prediction of hepatitis disease using K-nearest neighbors, naive bayes, support vector machine, multi-layer perceptron and random forest. In *Proceedings of the 2021 International Conference on Information and Communication Technology for Sustainable Development* (*ICICT4SD*) (pp. 280–284). Dhaka, Bangladesh. IEEE. https://doi.org/10.1109/ICICT4SD50815.2021.9397013

Olopade, B. C., Okodua, H., Oladosun, M., & Asaleye, A. J. (2019). Human capital and poverty reduction in OPEC member-countries. *Heliyon*, *5*(8), Article e02279. https://doi.org/10.1016/j.heliyon.2019.e02279

Omar, M. A., & Inaba, K. (2020). Does financial inclusion reduce poverty and income inequality in developing countries? A panel data analysis. *Journal of Economic Structures*, *9*, Article 37. https://doi.org/10.1186/s40008-020-00214-4

Patil, A. A., Desai, S. S., Patil, L. N., & Patil, S. A. (2022). Adopting artificial neural network for wear investigation of ball bearing materials under pure sliding condition. *Applied Engineering Letters*, *7*(2), 81–88. https://doi.org/10.18485/aeletters.2022.7.2.5

Roblek, V., Meško, M., Bach, M. P., Thorpe, O., & Šprajc, P. (2020). The interaction between internet, sustainable development, and emergence of society 5.0. *Data*, *5*(3), Article 80. https://doi.org/10.3390/data5030080

Satria, D. (2023). Predicting banking stock prices using RNN, LSTM, and GRU approach. *Applied Computer Science*, *19*(1), 82–94. https://doi.org/10.35784/acs-2023-06

Si, S., Ahlstrom, D., Wei, J., & Cullen, J. (2020). Business, entrepreneurship and innovation toward poverty reduction. *Entrepreneurship and Regional Development*, *32*(1–2), 1–20. https://doi.org/10.1080/08985626.2019.1640485

Siddiqa, A. (2021). Determinants of unemployment in selected developing countries: A panel data analysis. *Journal of Economic Impact*, *3*(1), 19–26. https://doi.org/10.52223/jei3012103

Spada, A., Fiore, M., & Galati, A. (2023). The impact of education and culture on poverty reduction: Evidence from panel data of European countries. *Social Indicators Research*, *175*, 927–940. https://doi.org/10.1007/s11205-023-03155-0

United Nations. (2022). *The sustainable development goals report*. https://unstats.un.org/sdgs/report/2022/The-Sustainable-Development-Goals-Report-2022.pdf

Uyen, V. T. N., & Thu, P. X. (2023). The multi-criteria decision-making method: Selection of support equipment for classroom instructors. *Applied Engineering Letters*, *8*(4), 148–157. https://doi.org/10.18485/aeletters.2023.8.4.2

Weldearegay, S. K., Tefera, M. M., & Feleke, S. T. (2021). Impact of urban expansion to peri-urban smallholder farmers' poverty in Tigray, North Ethiopia. *Heliyon*, 7(6), Article e07303. https://doi.org/10.1016/j.heliyon.2021.e07303

World Bank. (2023). *The World Bank in Indonesia*. https://www.worldbank.org/en/country/indonesia/overview

Yao, Y., Zhou, J., Sun, Z., Guan, Q., Guo, Z., Xu, Y., Zhang, J., Hong, Y., Cai, Y., & Wang, R. (2023). Estimating China's poverty reduction efficiency by integrating multi-source geospatial data and deep learning techniques. *Geo-Spatial Information Science*, *27*(4), 1000–1016. https://doi.org/10.1080/10095020.2023.2165975

Yoon, J. (2021). Forecasting of real GDP growth using machine learning models: Gradient boosting and random forest approach. *Computational Economics*, *57*, 247–265. https://doi.org/10.1007/s10614-020-10054-w

Zahra, S. A., Liu, W., & Si, S. (2023). How digital technology promotes entrepreneurship in ecosystems. *Technovation*, *119*, Article 102457. https://doi.org/10.1016/j.technovation.2022.102457

Zhou, Y., & Liu, Y. (2022). The geography of poverty: Review and research prospects. *Journal of Rural Studies*, *93*, 408–416. https://doi.org/10.1016/j.jrurstud.2019.01.008