Journal of Infrastructure, Policy and Development 2024, 8(8), 5938.
https://doi.org/10.24294/jipd.v8i8.5938

Article

A simple classification and clustering of poverty in rural areas using machine learning

Deddy Barnabas Lasfeto[1,*], Tuti Setyorini[2], James Josias Mauta[1], Melchior Bria[3], Obed Oktafianus Nego Nenobais[3]

[1] Department of Electrical Engineering, State Polytechnic of Kupang, Kupang City 85258, Nusa Tenggara Timur, Indonesia
[2] Department of Business Administration, State Polytechnic of Kupang, Kupang City 85258, Nusa Tenggara Timur, Indonesia
[3] Department of Civil Engineering, State Polytechnic of Kupang, Kupang City 85258, Nusa Tenggara Timur, Indonesia
*** Corresponding author:** Deddy Barnabas Lasfeto, deddylasfeto@gmail.com

CITATION

Lasfeto DB, Setyorini T, Mauta JJ, et al. (2024). A simple classification and clustering of poverty in rural areas using machine learning. Journal of Infrastructure, Policy and Development. 8(8): 5938. https://doi.org/10.24294/jipd.v8i8.5938

ARTICLE INFO

Received: 22 April 2024
Accepted: 21 May 2024
Available online: 14 August 2024

Abstract: This study aimed to determine the socio-economic poverty status of those living in rural areas using data surveys obtained from household expenditure and income. Machine learning-based classification and clustering models were proven to provide an overview of efforts to determine similarities in poverty characteristics. Efforts to address poverty classification and clustering typically involve comprehensive strategies that aim to improve socio-economic conditions in the affected areas. This research focuses on the combined application of machine learning classification and clustering techniques to analyze poverty. It aims to investigate whether the integration of classification and clustering algorithms can enhance the accuracy of poverty analysis by identifying distinct poverty classes or clusters based on multidimensional indicators. The results showed the superiority of machine learning in mapping poverty in rural areas; therefore, it can be adopted in the private sector and government domains. It is important to have access to relevant and reliable data to apply these machine learning techniques effectively. Data sources may include household surveys, census data, administrative records, satellite imagery, and other socioeconomic indicators. Machine learning classification and clustering analyses are used as a decision support tool to gain an understanding of poverty data from each village. These strategies are also used to describe the profile of poverty clusters in the community in terms of significant socio-economic indicators present in the data. Village clusters based on an analysis of existing poverty indicators are grouped into high, moderate, and low poverty levels. Machine learning can be a valuable tool for analyzing and understanding poverty by classifying individuals or households into different poverty categories and identifying patterns and clusters of poverty. These insights can inform targeted interventions, policy decisions, and resource allocation for poverty reduction programs.

Keywords: rural community; poverty; data clustering; data classification; machine learning

1. Introduction

One of the Sustainable Development Goals (SDGs) is to end all forms of poverty by 2030 (ADB, 2021). In Indonesia, poverty has been a widespread problem for over a decade and remains one of the most talked about and debated issues up to this time. The government continues to employ various ways to eradicate poverty in the country.

Poverty eradication is also a significant issue associated with development in Sabu Raijua Regency, East Nusa Tenggara Province of Indonesia. The poverty rate in Sabu Raijua Regency was slightly high, relatively 30.18% in 2020, and this figure is greater than the provincial and national averages of 20.9% and 9.78%, respectively. In addition, it took the 21st position compared to other Regencies or Cities in East Nusa

Tenggara (BPS, 2021). In the Regional Medium-Term Development Plan of Sabu Raijua Regency for 2021 to 2026, the poverty rate is targeted at 26.79% by 2026 (BPS, 2021). This certainly requires a more integrated handling pattern involving the application of various studies and development methods. One of the widely used methods to measure the poverty is a household survey that involves a comprehensive process to gather accurate data on the income, consumption, and living conditions of households to assess their poverty status (BPS, 2018).

In fact, this assumption is not entirely fulfilled, as the characteristics of the surveyed household may not necessarily have a linear relationship with the welfare level. Household characteristics tend to interact complexly with the floor material indicating poverty in one area but not another. Therefore, this study utilized the machine learning method, a branch of artificial intelligence (AI), to map poverty based on household characteristics. Social and economic data were analyzed regarding the classification and clustering of poverty in Sabu Raijua Regency using the machine learning approach. Machine learning is applied based on the idea that a system can process data, identify patterns, and make decisions with minimal manual intervention. This procedure enables computers to handle new situations through self-training, experience, analysis, and observation (Alsharkawi et al., 2021; Min et al., 2022; Omae, 2020).

Nationally, the general indicators of measuring poverty are food and non-food poverties, encompassing housing, clothing, education, and health. Moreover, to ascertain the progress of a region, it is also necessary to consider public infrastructure such as lighting, education, and transportation facilities (Ayush et al., 2020; BPS, 2018). The proposed study examined the differences in poverty analyses and developed a machine learning-based poverty mapping model using socio-economic data for addressing this issue in certain regions. This research focuses on the combined application of machine learning classification and clustering techniques to analyze poverty. It aims to investigate whether the integration of classification and clustering algorithms can enhance the accuracy of poverty analysis by identifying distinct poverty classes or clusters based on multidimensional indicators. The results complement the National Socio-Economic Survey data collected by the Central Bureau of Statistics (Kota et al., 2017). The surveys conducted focused on the classification of poverty levels concerning socio-economic data (Repollo and Robielos, 2021) and the prediction evaluation (Chitturi and Nabulsi, 2021). Additionally, the research seeks to understand how these findings can contribute to targeted poverty reduction strategies.

2. Materials and methods

The study was conducted in West Sabu District of Sabu Raijua regency which has the most extensive distribution of administrative areas and the greatest number of villages, as well as the highest population. West Sabu is a district with the widest area in Sabu Raijua Regency of Indonesia, 185.16 km^2 and with 17 villages. Similarly, West Sabu District has the largest population with 33,225 permanent residents.

The sample for this study consisted of 500 heads of households distributed in 17 villages in West Sabu District of Sabu Raijua Regency of Indonesia. The sample of this study were permanent households which are categorized as poor households.

This research was conducted by the following steps:

a) Data collection: Gather comprehensive and reliable data that includes relevant indicators of poverty, such as income, education, health, access to basic services, and other socioeconomic factors. Ensure the data is representative of the target population or geographical area of interest. Researchers have observed 500 households and used the printed form to obtaining the 14 poverty indicators of every household.
b) Data pre-processing: Clean, pre-process, and transform the data as necessary, including handling missing values, normalizing, or scaling features, and encoding categorical variables.
c) Feature engineering: Conduct feature engineering to derive new informative features or combinations of features that may better capture the multidimensional aspects of poverty.
d) Classification algorithm: conduct supervised machines learning algorithm, to classify individuals or households into different poverty classes based on the available indicators. Poverty classification was used the k-nearest neighbor (K-NN) algorithm of machine learning.
e) Clustering algorithm selection: Conduct unsupervised machine learning clustering algorithms, to identify clusters or patterns within the classified poverty classes. Poverty clustering used the K-means clustering algorithm of machine learning.
f) Model training and evaluation: Apply the selected algorithms to the classified data and evaluate the quality of the resulting clusters and classification.
g) Integration and interpretation: Analyze and interpret the identified poverty classes and clusters to gain insights into the multidimensional nature of poverty. Explore the relationships between the classified poverty classes and the clusters to understand the characteristics and patterns within each class or cluster.

The data acquired during the socio-economic survey were based on 14 national poverty indicators from the Central Bureau of Statistics. The PSE05 data update in the PPLS 2008 (Hernawati, 2017) employed a characteristic household approach with 14 qualitative poverty variables, including: 1) floor area per capita; 2) floor; 3) wall types; 4) toilet facilities; 5) drinking water; 6) lighting sources; 7) fuel; 8) purchase of meat, chicken, or milk; 9) eating frequency; 10) purchase of new clothing; 11) ability to seek medical treatment; 12) household head's business activity; 13) household head's educational qualifications; and 14) assets owned. The poverty indicators are as follows (I_1, …, I_{14}):

I_1: The area of the house floor is less than 8 m^2 per person.

I_2: The house floor is made of earth, bamboo, or cheap wood.

I_3: The house wall is made of bamboo, Rumbia, low-quality wood or unplastered wall.

I_4: No toilet facilities or shared with other houses.

I_5: The house lighting source does not use electricity.

I_6: The drinking water source includes wells, unprotected springs, rivers, or rainwater.

I_7: Cooking fuel for daily use is firewood, charcoal, or kerosene.

I_8: Only consumes meat, milk, or chicken once a week.

I_9: Only buys one new piece of clothing yearly.

I_{10}: Can only eat once or twice a day.

I_{11}: Unable to pay for medical treatment at a health center or clinic.

I_{12}: The head of the household source of income is a farmer with a land area of 500 m^2, fisherman, farmer, construction, plantation worker, or other jobs where one earns an income of less than IDR 600,000 per month.

I_{13}: The highest educational qualification of the household head: not schooled, failed to complete, or completed elementary school.

I_{14}: Does not have savings, easily sellable items worth at least IDR 500,000 such as a credit or non-credit motorbike, gold, livestock, motorboat, or other capital goods.

2.1. K-nearest neighbor (K-NN) classification method in machine learning

The K-NN is one of the simplest and most important machine learning algorithms in supervised learning. This method depends on labeled input data that can produce outputs every time new unlabeled data is given as input. K-NN is widely used because it is non-parametric, meaning it does not make basic assumptions about the data distribution. This method uses the entire dataset in its training phase. In addition, when a new prediction is made, the K-NN scans the entire training dataset to discover similar ones (Yoon, 2019). For instance, assuming an apple picture resembles that of a pear and cherry (fruits) rather than a picture of a monkey, cat, or mouse (animals), the apple is most likely a fruit.

The main objective of the K-NN algorithm is to classify a new data point by comparing it to all the previous ones. In most cases, the classification of the previous similar *K* is used to predict the current data point. In the K-NN algorithm, the prediction depends entirely on the distance measured. Therefore, this algorithm is suitable for applications with adequate domain knowledge, which helps select the appropriate size. K-NN is used for regression and prediction analyses, but this algorithm is widely used in classification problems. It is also mostly used for ease of interpretation and faster computation time. In addition, the K-NN algorithm is based on feature similarity. The classification of a specific data point is determined by the similarities between its features and the training set.

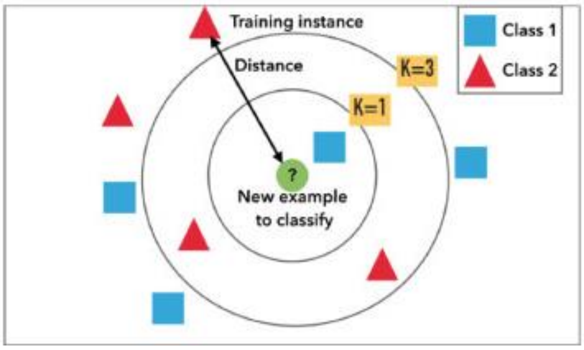

Figure 1. K-NN classification (Tran et al., 2019).

Figure 1 shows the K-NN classification model, and considering the nearest neighbors of the test sample (?), the closest one is the blue square (class 1) or $K = 1$ inside the inner circle. Assuming $K = 3$, there will be two red triangles and only one blue square under the outer circle. Therefore, the test sample is now classified as the red triangle (class 2). When $K = 5$, it is set to the first class with three squares and two triangles outside the outer circle.

The majority frequency of its neighbors can classify a case, which is then distributed to the most common class among its K-NN as measured by the distance function. The distance metric used to determine the nearest neighbors is realized using the Euclidean formula as follows:

$$D = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \tag{1}$$

where: D is the distance, n is the number of dimensions (attributes), x_i and y_i are the i-th attributes of objects data p and q. Meanwhile, the optimal best value for K can be selected by examining the data. Generally, a larger K value is more accurate as it tends to reduce noise or be invalid overall. In this study, a value of $K = 3$ was used.

2.2. K-means clustering method

The steps in the K-means clustering algorithm are stated as follows:

(1) Determine the number of K clusters: The number of K clusters is determined freely, although it should meet certain conditions, particularly the one that states its value must be lesser than the number of data;

(2) Select K points randomly as the centers (centroids) of each group (cluster);

(3) Calculate distances and allocate each data to the nearest centroid (mean);

(4) Determine the new centroid or average data in each cluster;

(5) Repeat step 3, assuming there is still data that needs to be moved to a different cluster or there is a change in the value of the centroid. Assuming there is no change, then the clustering process needs to be stopped.

These steps are shown in the following flowchart in **Figure 2**.

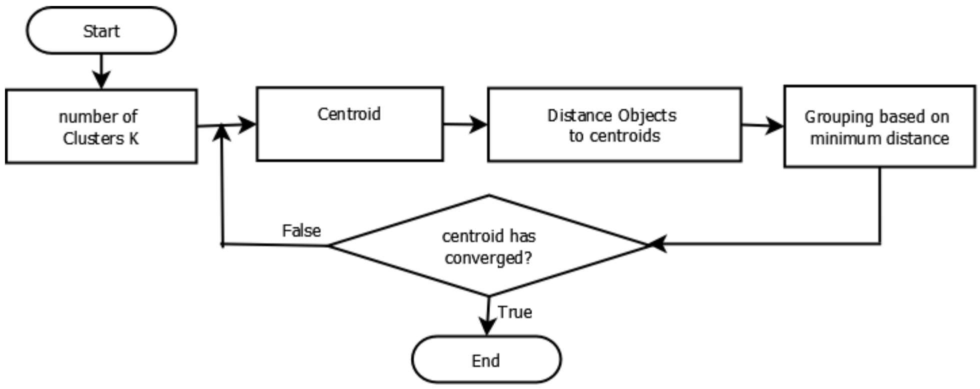

Figure 2. Flow chart of K-means clustering algorithm.

3. Results

Based on the poverty indicator that mentioned in the method section, researchers were used to survey a sample of 500 heads of households distributed in 17 villages in

West Sabu District. Datas were recapped for specific poverty indicators and calculated for each respondent (head of household) who met the criteria of poverty. **Table 1** shows the summary of the survey results.

Table 1. Summary of survey results.

Village number	I_1	I_2	I_3	I_4	I_5	I_6	I_7	I_8	I_9	I_{10}	I_{11}	I_{12}	I_{13}	I_{14}
1	26	16	27	8	2	10	28	28	27	27	27	27	6	27
2	0	21	28	2	0	0	28	28	23	20	28	25	8	12
3	5	25	26	5	0	0	27	21	16	20	27	22	15	20
4	22	22	22	6	5	5	22	19	10	20	22	20	12	10
5	23	18	23	5	18	17	23	22	15	10	23	23	13	13
6	5	28	36	9	11	5	36	36	34	23	36	35	10	33
7	9	27	29	13	0	0	28	15	14	22	29	26	14	24
8	29	29	29	9	13	28	29	28	13	29	29	29	7	28
9	3	18	26	2	0	26	26	24	19	15	26	20	3	20
10	1	21	22	0	4	21	22	22	14	17	22	20	12	22
11	28	28	28	18	17	24	28	27	28	25	28	26	20	28
12	27	28	28	7	15	28	28	28	28	21	27	28	13	28
13	14	22	31	1	3	1	35	35	26	32	35	32	11	12
14	19	25	27	7	4	12	26	31	27	22	21	28	5	11
15	1	9	27	7	1	12	22	27	8	19	28	22	12	15
16	28	28	28	4	10	26	28	28	28	5	26	28	15	28
17	23	30	31	4	5	1	30	24	28	19	30	30	10	15

The table shows over 90% of villages in West Sabu met each poverty indicator. For instance, village 1 consists of 26 heads of households which have houses with a floor area of less than 8 m^2 per capita (I_1), 16 of them have houses with floor types made of soil, bamboo, or cheap wood (I_2), 27 are only able to eat once, or twice a day (I_{10}). The detailed data of each village is shown in **Table 1**.

Table 2. Processed data records.

		N	**Percent**
Sample	Training	16	94.1%
	Holdout	1	5.9%
Valid		17	100.0%
Excluded		0	
Total		17	

Table 3. Classification with the K-NN algorithm.

Focal record	**Nearest neighbors**			**Nearest distances**		
	1	**2**	**3**	**1**	**2**	**3**
Village 7	Village 3	Village 17	Village 3	1.407	2.366	2.428

The results of the analyzed poverty classification process in each village using the machine learning technique with the k-nearest neighbor algorithm (Forero-Corba and Bennasar, 2024; Singh and Bhatia, 2007; Utmal, 2021) are stated in **Tables 2** and **3**.

Village 7 has similar poverty characteristics to the three others villages (villages 3, 17, and 13). However, the closest or almost the same characteristic to village 7 was village 3, as depicted by the value of the distance between the two data is 1.407.

The clustering method is used to group villages with similar characteristics; hence, it serves as a reference for relevant parties in employing the appropriate steps to address poverty in Sabu Raijua Regency. The results of the K-means clustering algorithm are shown in **Table 4** and **Figure 1**.

Table 4. Clustering iteration.

Iteration	Change in cluster centers		
	1	2	3
1	2.278	2.675	2.908
2	0.000	0.000	0.000

Table 4 shows the clustering process took two iterations, and afterward, the data converged did not change. Additionally, the minimum distance realized was 6437. Its convergence led to the final cluster of the K-means clustering analysis where there are no more changes in the centroid of each cluster, as shown in the following **Figure 3**.

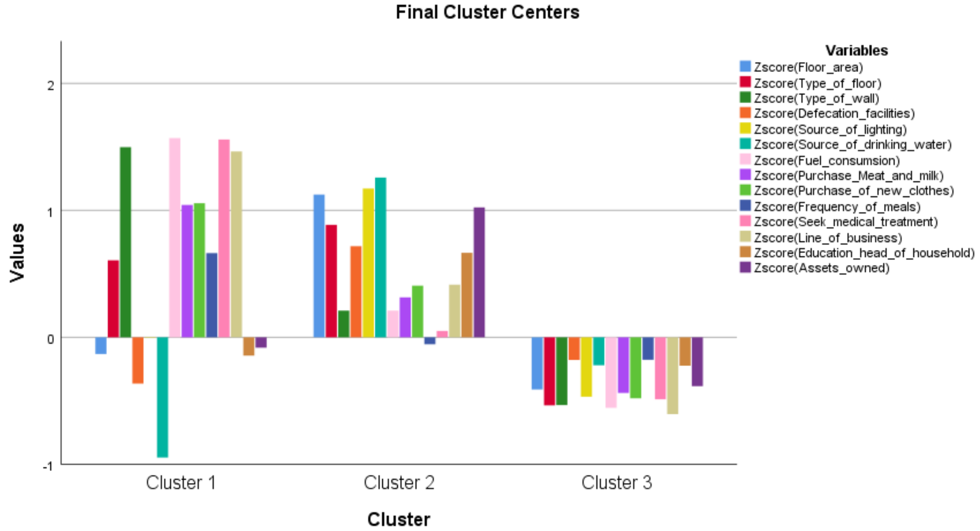

Figure 3. Clustering results.

The results of clustering all villages are displayed in the following **Table 5**.

Table 5. Village clustering for all poverty indicators.

Cluster of villages	1	2	3
Member	3	4	10
Valid	17		
Missing	0		

4. Discussion

Poverty classification refers to the categorization of individuals or households based on their level of poverty. Governments, international organizations, and researchers often use poverty classification systems to identify and target individuals or communities in need of assistance and to measure the effectiveness of poverty reduction programs (Poreddy et al., 2020; Sihombing and Arsani, 2021). The poverty indicators of 17 villages are trained, and the predictor graph of the k-nearest neighbor classification analysis for 14 indicators. In the term of poverty classification of the floor area indicator (I_1), village number 7 shows a closeness or similar characteristics with village number 17, 3, and 2. The toilet facility indicator (I_4) has similar characteristics to village number 3, 17, and 2. Overall, the distance of each poverty indicator from village 7 to its nearest neighbors is shown in **Table 3**.

This approach defines poverty based on a fixed threshold that represents the minimum income or consumption level required to meet basic human needs, such as food, shelter, clothing, and healthcare. Individuals or households falling below the established threshold are considered to be in absolute poverty (Min et al., 2022).This approach involves training a machine learning model on labeled data, where poverty status is known, to predict the poverty status of new, unlabeled data points. Features used for classification can include income, education level, household composition, employment status, and access to basic services (Castro and Álvarez, 2022). It is important to note that poverty classification systems may vary depending on the country, organization, or researcher using them. These systems are continually evolving as new research and data become available, and they serve as a tool to understand and address the challenges faced by people living in poverty.

Poverty clustering refers to the phenomenon where individuals or households experiencing poverty are geographically concentrated in specific areas or regions. It is often observed that poverty is not evenly distributed across a country or within a city, but rather tends to concentrate in certain neighborhoods, communities, or regions.

In the clustering analysis, clusters 1, 2, and 3 have relatively high, moderate, and low poverty levels, respectively (Tingzon et al., 2019; Verma and Preety, 2020). Meanwhile, from the analysis, in cluster 1, some poverty indicators that have a significant impact are: (1) The type of house wall, where most of the community used low-quality bamboo, Rumbia, wood or unplastered wall. (2) The fuel consumed, where the fuel used for daily cooking is firewood. The poverty variable needs to be discussed and reviewed in subsequent studies from the perspective of local geographic and cultural conditions. (3) The cost of medical treatment at the community health center, clinic, or hospital is unaffordable by most people in cluster 1. The government needs to intervene through the Indonesia Healthy Card program; (4) the monthly income of the household head is less than IDR 600,000, therefore, the government needs to intervene through family empowerment and economic improvement programs.

In cluster 2, some significant poverty indicators are: (1) the variable of the floor area is less than 8 m^2 per person; (5) the source of lighting does not use electricity; and (6) the sources of drinking water are from wells, unprotected springs, rivers, or rainwater. Therefore, programs prioritized for poverty alleviation in the regions grouped under cluster 2 focus on these three indicators. Cluster 3 has a relatively low poverty level, but needs to consider two variables, namely (4) not having a toilet facility with other households, and (10) only being able to eat once or twice daily. These two variables are closed to the minimum limit, and then for programs in Cluster 3, they are considered the main priority.

The analysis showed that villages 6, 13, and 17 are the three villages in cluster 1. Consequently, these three villages have similar poverty characteristics, which serve as a reference for poverty alleviation using the same pattern. Four villages are included in cluster 2, consists of villages 8, 11, 12, and 16 which have similar poverty characteristics. Ten villages are included in cluster 3, consists of villages 1, 2, 3, 4, 5, 7, 9, 10, 14, and 15. By applying k-means clustering algorithms, hierarchical clustering, or density-based clustering, patterns and groups within the data can be identified. The resulting clusters can then be analyzed to understand the characteristics and potential poverty status of each group. The application of machine learning-based classification and clustering methods is in the form of mapping the number of poor populations in Sabu Raijua Regency. The results serve as useful information for alleviating poverty in communities where high clusters (cluster 1) become a priority for the government to provide poverty alleviation intervention programs. The consequences of poverty clustering can be far-reaching. Concentrated poverty often leads to a cycle of disadvantage, as individuals and communities face limited opportunities for upward mobility (Ochoa Guaraca et al., 2021). It can result in reduced access to quality education, healthcare disparities, higher crime rates, and social isolation.

Efforts to address poverty clustering typically involve comprehensive strategies that aim to improve socio-economic conditions in the affected areas. These strategies may include targeted investments in infrastructure, job creation programs, educational initiatives, affordable housing policies, and community development projects. Additionally, promoting social inclusion, reducing discrimination, and fostering community engagement are crucial for breaking the cycle of poverty clustering and creating opportunities for individuals and communities to thrive (Alsharkawi et al., 2021).

To apply these machine learning techniques effectively, it is important to have access to relevant and reliable data. Data sources may include household surveys, census data, administrative records, satellite imagery, and other socioeconomic indicators. Feature engineering and selection are crucial steps in preparing the data for machine learning algorithms (Rozenberg and Hallegatte, 2016). Additionally, interpreting and validating the results of machine learning models is important to ensure their reliability and avoid potential biases (Ahn et al., 2020). Domain experts and policymakers should be involved in the process to provide insights, verify results, and interpret the clustering or classification outcomes in the context of poverty alleviation efforts (Isnin Hamdan et al., 2020).

The findings of machine learning classification and clustering can contribute to poverty reduction strategies. Explore how the identified poverty classes and clusters can guide policymakers in targeting interventions, allocating resources, and addressing the specific needs and challenges of different poverty groups. Future research could explore alternative machine learning algorithms, incorporating additional data sources, or investigating the dynamic nature of poverty over time.

5. Conclusion

The classification and clustering algorithms are used to map each village in West Sabu District of Sabu Raijua Regency, with similar poverty characteristics. The poverty condition can be analysed and prioritized in terms of implementing socio-economic programs in each village cluster as recommended according to 14 socio-economic indicators for poverty alleviation programs and to develop a comprehensive strategy for the community.

Machine learning classification and clustering analyses are used as a decision support tool to gain an understanding of poverty data from each village. These strategies are also used to describe the profile of poverty clusters in the community in terms of significant socio-economic indicators present in the data. Village clusters based on an analysis of existing poverty indicators are grouped into high, moderate, and low poverty levels. The analysis can be used as a basis for recommending programs addressing the most pressing poverty condition in each cluster.

Machine learning can be a valuable tool for analysing and understanding poverty by classifying individuals or households into different poverty categories and identifying patterns and clusters of poverty. These insights can inform targeted interventions, policy decisions, and resource allocation for poverty reduction programs. Policymakers, in addition to local civil society organizations and foreign organizations concerned with combating poverty, could implement the machine learning method on the ground in order to achieve the stated aims of this study and sustainability.

Author contributions: Conceptualization, DBL and TS; methodology, DBL and MB; software, DBL; validation, DBL, JJM and TS; formal analysis, DBL and OONN; investigation, DBL and JJM; resources, DBL and MB; data curation, JJM and OONN; writing—original draft preparation, DBL and TS; writing—review and editing, DBL; visualization, JJM and MB; supervision, TS and OONN; project administration, DBL; funding acquisition, DBL. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by DAPTV Kemdikbudristek RI grant number 092/SPK/D4/PPK.01.APTV/VI/2022.

Acknowledgments: This study was supported by the DAPTV of the Ministry of Education, Culture, Research, and Technology of Indonesia.

Conflict of interest: The authors declare no conflict of interest.

References

ADB. (2021). Mapping the Spatial Distribution of Poverty Using Satellite Imagery in Thailand. Asian Development Bank.

Ahn, D., Cha, M., Han, S., et al. (2020). Teaching Machines to Measure Economic Activities from Satellite Images: Challenges and Solutions. Machine Learning and Satellite Imagery.

Alsharkawi, A., Al-Fetyani, M., Dawas, M., et al. (2021). Poverty classification using machine learning: The case of Jordan. Sustainability (Switzerland), 13(3), 1–16. https://doi.org/10.3390/su13031412

Ayush, K., Uzkent, B., Tanmay, K., et al. (2020). Efficient Poverty Mapping using Deep Reinforcement Learning. Available online: http://arxiv.org/abs/2006.04224 (accessed on 22 January 2024).

BPS. (2018). SUSENAS Modul 2018. BPS.

BPS. (2021). Sabu Raijua in Figures 2021 (Indonesian). BPS.

Castro, D. A., & Álvarez, M. A. (2022). Predicting socioeconomic indicators using transfer learning on imagery data: an application in Brazil. GeoJournal, 88(1), 1081–1102. https://doi.org/10.1007/s10708-022-10618-3

Chitturi, V., & Nabulsi, Z. (2021). Predicting Poverty Level from Satellite Imagery using Deep Neural Networks. arXiv.

Forero-Corba, W., & Bennasar, F. N. (2024). Techniques and applications of Machine Learning and Artificial Intelligence in education: a systematic review. RIED-Revista Iberoamericana de Educacion a Distancia, 27(1), 209–253. https://doi.org/10.5944/ried.27.1.37491

Hernawati, I. (2017). The measurement of poverty construct in Indonesia (Indonesian). Media Informasi Penelitian Kesejahteraan Sosial, 41(3), 269–284.

Isnin Hamdan, R., Bakar, A. A., & Sani, N. S. (2020). Does Artificial Intelligence Prevail in Poverty Measurement? Journal of Physics: Conference Series, 1529(4), 042082. https://doi.org/10.1088/1742-6596/1529/4/042082

Kota, K., Iii, B., Art, B., et al. (2017). National socio-economic survey 2017 (Indonesian). BPS.

Min, P. P., Gan, Y. W., Hamzah, S. N. B., et al. (2022). Poverty Prediction Using Machine Learning Approach. Journal of Southwest Jiaotong University, 57(1), 136–146. https://doi.org/10.35741/issn.0258-2724.57.1.12

Ochoa Guaraca, M. E., Castro, R., Arias Pallaroso, A., et al. (2021). Machine learning approach for multidimensional poverty estimation. Revista Tecnológica—ESPOL, 33(2), 205–225. https://doi.org/10.37815/rte.v33n2.853

Omae, O. J. (2020). University of Nairobi Using Random Forest (RF) to Identify Key Determinants of Poverty in Kenya School of Mathematics [Marster's thesis]. School of Mathematics.

Poreddy, D., Reddy, E. V. V., Prasad, S. V., et al. (2020). Classification of Poverty Levels Using Machine Learning. Journal of Xi'an University of Architecture & Technology, 12(4), 5723–5728.

Repollo, M. P., & Robielos, R. A. C. (2021). Applying Clustering Algorithm on Poverty Analysis in a Community in the Philippines. In: Proceedings of the International Conference on Industrial Engineering and Operations Management Monterrey; 3–5 November 2021; Mexico. pp. 1511–1521.

Rozenberg, J., & Hallegatte, S. (2016). Model and Methods for Estimating the Number of People Living in Extreme Poverty Because of the Direct Impacts of Natural Disasters. World Bank, Washington, DC. https://doi.org/10.1596/1813-9450-7887

Sihombing, P. R., & Arsani, A. M. (2021). Comparison of Machine Learning Methods in Classifying Poverty in Indonesia in 2018. Jurnal Teknik Informatika (Jutif), 2(1), 51–56. https://doi.org/10.20884/1.jutif.2021.2.1.52

Singh, Y., & Bhatia, P. K. (2007). A Review of Studies on Machine Learning Techniques. International Journal of Computer Science and Security, 1(1), 70.

Tingzon, I., Orden, A., Go, K. T., et al. (2019). Mapping poverty in the Philippines using machine learning, satellite imagery, and crowd-sourced geospatial information. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-4/W19, 425–431. https://doi.org/10.5194/isprs-archives-xlii-4-w19-425-2019

Tran, H., Mining, D., Multimedia, M., & Times, L. S. (2019). A survey of machine learning and data mining. Available online: https://www.researchgate.net/publication/333457161_Survey_of_Machine_Learning_and_Data_Mining_Techniques_used_in_Multimedia_System?channel=doi&linkId=5d88ee6c299bf1996f987f9f&showFulltext=true (accessed on 14 May 2023).

Utmal, D. M. (2021). Machine Learning Its Applications, Challenges & Tools: A Review. International Journal of Computer Science and Mobile Computing, 10(3), 32–38. https://doi.org/10.47760/ijcsmc.2021.v10i03.004

Verma, D. P. K., & Preety. (2020). Application of K-Means Algorithm to Mapping Poverty Outline by Province in India. International Journal of Recent Technology and Engineering (IJRTE), 8(6), 1045–1049. https://doi.org/10.35940/ijrte.f7357.038620

Yoon, H. (2019). Artificial Intelligent Technology in Public and Private Sector: The case of Korea. In: Proceedings of the Artificial Intelligence (AI): Overview and Applications; 17 September 2019; Bangkok, Thailand.