Multi-Algorithm Approach to Enhancing Social Assistance Efficiency Through Accurate Poverty Classification

Christofer Satria, Peter Wijaya Sugijanto, Anthony Anggrawan, I Nyoman Yoga Sumadewa, Aprilia Dwi Dayani, Rini Anggriani
Universitas Bumigora, Mataram, Indonesia

Article Info

Article history:

Received July 19, 2024
Revised November 03, 2024
Accepted November 13, 2024

Keywords:

Classification
Multi-Algorithm
Poverty
Social Assistance Efficiency

ABSTRACT

Determining poverty status in Lombok Utara district depends on criteria such as income, access to health and education services, and housing conditions. These factors are crucial for assessing the level of community welfare and guiding the allocation of social assistance by the district government. The purpose of this study is to address the gap by utilizing advanced data mining techniques to improve the accuracy of poverty status classification in North Lombok, thereby informing more effective social assistance policies. The method used in this research is the Random Forest (RF), K-Nearest Neighbor (KNN), and Naïve Bayes with split data, 80% data training, and 20% data testing. The finding indicated that the machine learning model of the RF algorithm, which achieved an accuracy rate of 82.56%, proved to play an important role in this process by effectively distinguishing between different categories of poverty based on these criteria. In comparison, the KNN algorithm achieved an accuracy of 70.94%, and the Naïve Bayes model achieved an accuracy of 53.47%. It means that the RF algorithm's machine learning model is more accurate than the KNN and Naïve Bayes algorithm in predicting or recommending Recipients of Social Assistance from the District Government. The implication is that RF machine learning can help social service officers predict the community's economic status. The high accuracy of the RF algorithm enhances its role in informing targeted policy decisions and optimizing the effectiveness of social assistance programs. Nonetheless, continuous improvement is essential to refine the model's predictive capabilities and ensure the accuracy and reliability of poverty assessments. These continuous improvements are essential to effectively alleviate poverty and break the cycle of socio-economic disparities in the region.

Corresponding Author:

Christofer Satria, +62 8873452345
Faculty of Art & Design and Study Program of Visual Communication Design,
Universitas Bumigora, Mataram, Indonesia,
Email: chris87@universitasbumigora.ac.id.

How to Cite:
C. Satria, P. Sugijanto, A. Anggrawan, I. N. Sumadewa, A. Dayani, and R. Anggriani, "Multi-Algorithm Approach to Enhancing Social Assistance Efficiency Through Accurate Poverty Classification", *MATRIK: Jurnal Manajemen, Teknik Informatika, dan Rekayasa Komputer*, Vol. 24, No. 1, pp. 167-178, November, 2024.

1. INTRODUCTION

North Lombok Regency is one of the regencies located in West Nusa Tenggara (NTB) province and is the youngest regency in NTB. Based on data from the Central Statistics Agency (BPS), in 2023, the poverty rate in North Lombok Regency reached 25.80 percent. Although this figure shows a decrease of 0.13 percent from the previous year, this high poverty rate remains a serious challenge for the local government. This minimal decline shows the need for more effective and sustainable interventions in poverty alleviation efforts in North Lombok Regency. The determination of poverty status in North Lombok Regency is based on the criteria used to measure the level of community welfare. These criteria include income, access to health and education services, and housing conditions. The North Lombok district government uses this data to determine who is entitled to receive social assistance. This assistance is expected to ease the burden on the poor and help them escape the poverty cycle. However, the effectiveness of this assistance is highly dependent on the accuracy of the data and the criteria used.

This study seeks to fill the gap in previous research by further exploring the factors influencing poverty in the North Lombok Regency. Despite various studies on poverty and welfare classifications, there remains a lack of localized analysis that considers the specific socio-economic conditions of North Lombok. Previous research has highlighted the importance of accurate data and targeted interventions in poverty alleviation. For instance, [1] emphasized the need for community-specific data in formulating effective poverty reduction strategies. Similarly, [2] pointed out that precise identification of poverty determinants is crucial for policy effectiveness. In the context of Indonesia, studies by [3] and [4] have demonstrated the impact of accurate poverty classification on the success of social assistance programs. Additionally, [5] has stressed the need to improve poverty measurement techniques to enhance policy outcomes continuously. This study aims to address these gaps by utilizing advanced data mining techniques to improve the classification accuracy of poverty status in North Lombok, thereby informing more effective social assistance policies. By employing methods such as Random Forest (RF), K-Nearest Neighbor (KNN), and Naïve Bayes, this research intends to identify and analyze the key factors that contribute to poverty in the region. The findings are expected to provide a more precise understanding of poverty dynamics in North Lombok, leading to more targeted and effective interventions. This approach not only aims to improve the accuracy of welfare classifications but also to ensure that the benefits of social assistance programs reach those who need them most, ultimately contributing to more significant poverty alleviation in the region.

Previous research shows that poverty reduction policies and programs must be based on accurate data and include in-depth analysis of the factors that cause poverty. The study conducted by [6] aims to classify poverty status using the C4.5 classification method to determine the factors that affect poverty. Another study by [7] The binary logistic regression method aims to identify the characteristics and factors affecting extreme poverty in poor households. In contrast to research by [8, 9], which aims to see the relationship between village funds and poverty in North Lombok Regency; the method used is geographically weighted regression (GWR), showing that village funds can reduce poverty in every village in North Lombok Regency but have no significant effect because other variables outside the model still influence the resulting model. Research conducted by [10] aims to rank community social assistance recipients using the weighted product method based on five criteria: employment status, house status, domicile, number of dependents, and total monthly income. A study by [11] uses qualitative methods to describe the process of implementing integrated social protection, showing that social protection efforts for the poor through an integrated social protection system are implemented by incorporating five aspects of services, such as one-stop service, partnerships, service mechanism programs, case management, and information and intervention systems. Another study by [12] created a decision support system to help make it easier to select and determine poverty levels by applying the Multi-Objective Optimization based on Ratio Analysis (MOORA) method, entering criteria and alternative data into a system that automatically calculates and produces data with the lowest to highest values based on 25 alternative data and 5 criteria.

In recent years, machine learning techniques have gained traction in poverty classification and prediction. [13] utilized the RF algorithm to classify poverty status and found it to be highly effective in capturing complex patterns in socio-economic data. Similarly, a study by [14] compared the performance of K-NN and Support Vector Machine (SVM) algorithms in poverty classification, highlighting the strengths and limitations of each method in different contexts. Research [15] applied Gradient Boosting Machines (GBM) to identify key poverty determinants, demonstrating the algorithm's superior predictive accuracy. The study conducted by [16] aims to classify poverty status using the C4.5 classification method to determine the factors that affect poverty. Another study also conducted by [17] aims to identify the characteristics and factors affecting extreme poverty in poor households using the binary logistic regression method. Research conducted by [18] aims to rank recipients of social assistance for the community using the weighted product method based on five criteria: employment status, house status, domicile, number of dependents, and total monthly income. Research conducted by [19] uses qualitative methods to describe the process of implementing integrated social protection, showing that social protection efforts for the poor through an integrated social protection system are implemented by implementing five aspects of services, such as one-stop service, partnerships, service mechanism programs, case management, and information and intervention systems. Unlike previous research that often focused on a single classification method, this study incorporates multiple algorithms RF, K-NN, and Naïve Bayes to provide a more comprehensive analysis of poverty classification. By comparing the performance of these algorithms, this research aims to identify the most effective method for accurately classifying poverty status and guiding social assistance programs. The addition of these methods allows for a more robust and nuanced understanding of the factors contributing to poverty in Lombok Utara district, ultimately enhancing the targeting and effectiveness of poverty alleviation efforts.

2. RESEARCH METHOD

Finding information hidden in a pile of records in a database requires its own approach. One of the approaches used in this research is Knowledge Discovery in Database (KDD) [20]. KDD is the process of extracting information from a large database. KDD is a method for obtaining knowledge from existing databases. In the database, some tables are interconnected/connected. The results of the knowledge obtained in this process can be used as a basis for knowledge for decision-making purposes, as shown in Figure 1.

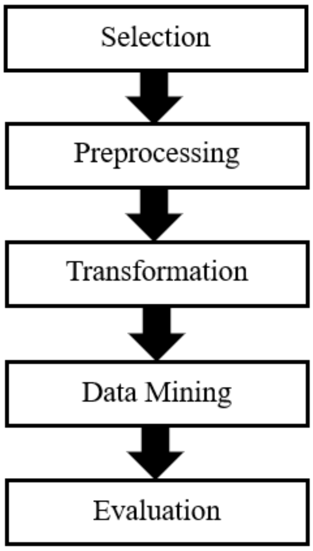

Figure 1. KDD Method

Data selection is the first stage in the KDD process, which aims to select a subset of relevant and useful data for further analysis. The data selected should reflect the context of the problem to be solved and should be relevant to the research objectives so that the analysis is more accurate and focused. The preprocessing stage includes, among others, cleaning the data from noise, such as data duplication, checking for inconsistent data, providing missing values, and correcting errors in the data, such as printing errors. Data must be transformed before processing using data mining. This aims to adjust the data processing based on the algorithms and software used in the data processing. This process is to find patterns and interesting information in selected data using specific techniques and methods. Data mining techniques, methods, and algorithms vary widely. The appropriate method or algorithm selection depends on the overall KDD objectives and process. It is the process of evaluating the results of the data mining information pattern process output. Researchers tried to compare the KNN, Naive Bayes, and RFt methods in Factors Affecting Poverty in North Lombok [6].

2.1. Model Testing

2.1.1. Random Forest

RF is an ensemble method of decision trees. Random forest combines decision simplicity. Flexible trees produce huge improvements in accuracy. Random Forest uses the Bootstrap aggregating (Bagging) technique by creating a bootstrapped dataset to reduce variance in noisy data sets and aggregating by voting on the most results. The bootstrapped dataset will train each decision tree model in parallel [9].

2.1.2. KNN Method

KNN is a method using a supervised learning algorithm in which several new attributes whose class is unknown can be searched for by looking for similarities based on the majority of KNN as a reference in determining the class of an attribute. The KNN method can be calculated using the Euclidean distance to determine the distance between datasets [7]. The Euclidean distance formula can be seen in Equation 1, where $d(p, q)$ represents the Euclidean distance between two points p and q. p_i and q_i represent the i-th coordinates of points p and q in n-dimensional space. $(pi - qi)^2$ represents the squared difference between the i-th components of points p and q.

$$d(p,q) = \sqrt{\sum_i (p_i - q_i)^2} \tag{1}$$

2.1.3. Naïve Bayes

Naïve Bayes is a simple classifying probabilistic method based on Bayes' theorem where classification is carried out efficiently through training sets of many data. Naïve Bayes assumes that the value of an input attribute in a given class does not depend on the values of other attributes. Bayes' theorem itself was put forward by British scientist Thomas Bayes, namely predicting future opportunities based on previous experience, so it is known as Bayes' theorem [7]. Where the Bayes theory Equation 2. Where, P(C|X) is the Probability of hypothesis C based on condition X (posterior probability). P(C) is the Probability of hypothesis C (prior probability). P(X|C) is Probability of X based on conditions in hypothesis C, and $P(X)$ is the Probability of X.

$$P(C \mid X) = \frac{P(X \mid C)P(C)}{P(X)} \tag{2}$$

2.2. Performance analysis

We employed four evaluation parameters in this study: precision, recall, F1-score, and accuracy. These metrics were calculated using model performance estimates [9]. The performance metric used for our model evaluation was calculated using the procedure stated in Equations 2, 3, 4, and 5. Accuracy indicates the number of samples accurately corrected for a given total number of samples. Precision indicates the number of samples accurately corrected for a given total number of samples. Recall indicates the prediction of positive values to the actual positive value. Where, TP is True Positive, TN is True Negative, FP is False Positive, and FN is False Negative.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

$$precision = \frac{TP}{TP + FP} \tag{4}$$

$$recall = \frac{TP}{TP + Fn} \tag{5}$$

$$\text{F1 score} = \frac{2 * Recall + Precision}{Recall + Precision} \tag{6}$$

3. RESULT AND ANALYSIS

The poverty status in North Lombok Regency was classified using RapidMiner. This approach involved data mining techniques to analyze various socio-economic factors affecting the population systematically. The results of this study are **in line** with or supported by Previous research, showing that poverty reduction policies and programs must be based on accurate data and include in-depth analysis of the factors that cause poverty. The study conducted by [6] aims to classify poverty status using the C4.5 classification method to determine the factors that affect poverty. Another study by [7] aims to identify the characteristics and factors affecting extreme poverty in poor households using the binary logistic regression method. **The finding** indicated that the machine learning model of the RF algorithm, which achieved an accuracy rate of 82.56%, proved to play an important role in this process by effectively distinguishing between different categories of poverty based on these criteria. In comparison, the KNN algorithm achieved an accuracy of 70.94%, and the Naïve Bayes model achieved an accuracy of 53.47%. RapidMiner aimed to identify patterns and categorize individuals or households based on their poverty status, thereby providing a more precise and data-driven foundation for policy-making and targeted social assistance programs, as shown in Figure 2. Based on Figure 1, the classification process uses several operators contained in RapidMiner. The operators used have their respective functions and purposes at different stages. The stages in the classification process are as follows:

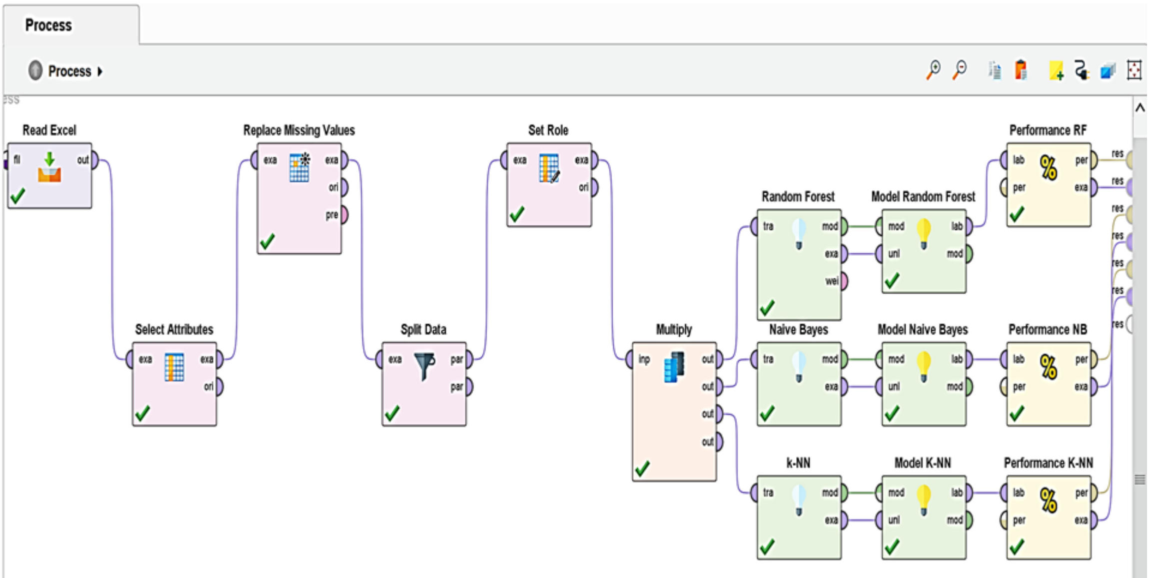

Figure 2. Classification Process Using RapidMiner

3.1. Selection

In the data selection phase, the 'Read Excel' operator is pivotal for initiating comprehensive data retrieval directly from Excel files. This operator seamlessly extracts structured datasets from Excel spreadsheets, ensuring that diverse and valuable information is readily accessible for analysis. Its functionality extends beyond mere data importation by providing essential capabilities for handling various data formats and structures inherent in Excel files. Analysts can efficiently manage data integration tasks by utilizing the 'Read Excel' operator, preparing the groundwork for subsequent exploration and detailed analysis. This initial step not only streamlines the workflow but also enhances the accuracy and reliability of data-driven insights derived from Excel-based datasets [21]. During the data import stage, the system usually previews the data before it is entered. This preview serves the important purpose of verifying the accuracy and completeness of the imported information. The preview allows analysts and users to visually inspect the data set, ensuring that the data's format, structure, and content match expectations before proceeding further. This precautionary step helps to reduce potential errors or discrepancies that may arise during data transfer, thereby facilitating smoother and more reliable data integration into the analysis pipeline [22].

After completing the data import stage, the next crucial step in the data preparation process is the application of the 'Select Attributes' operator. This operator is fundamental in refining the dataset by determining which attributes will be utilized for subsequent analysis and which ones will be excluded. With an initial set of 32 attributes imported, the 'Select Attributes' operator carefully evaluates each attribute based on its relevance to the specific analytical goals and modeling requirements. By retaining 19 of the 32 attributes, the operator ensures that only the most pertinent variables are included in further processing. This selective approach streamlines the dataset and enhances the accuracy and efficiency of subsequent data mining tasks. The attributes selected typically have the most significant impact on the outcomes of interest, ensuring that the analytical models derived from the data are robust and effective. Moreover, the 'Select Attributes' operator enables analysts to focus on meaningful variables directly contributing to achieving actionable insights and informed decision-making. This step is essential for optimizing computational resources and improving the interpretability of the analysis results, ultimately leading to more reliable and impactful outcomes in data-driven projects.

3.2. Preprocessing

To ensure the integrity and accuracy of data used in the mining process, thorough data cleaning is essential. This process involves systematically removing irrelevant data, including empty, duplicate, and missing entries, which can skew the results of data mining analyses. A critical tool for this task is the 'Replace Missing Value' operator, which methodically handles missing data points to improve the reliability of subsequent data processing steps [23] This meticulous approach not only enhances the dataset's quality but also ensures that the analytical models built upon it are robust and capable of producing accurate insights. Analysts can mitigate potential biases and errors by employing effective data cleaning practices, thereby increasing the validity and usefulness of their data-driven findings for decision-making and strategic planning, as shown in Figure 3.

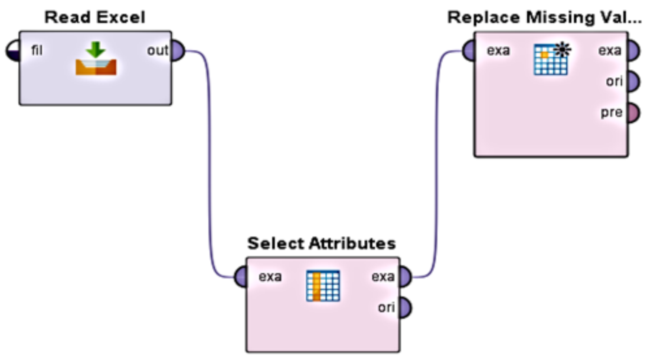

Figure 3. Use of Missing Value Operator

3.3. Transformation

This transformation involves converting the selected data into a format suitable for mining applications. In the context of this final project, two distinct data transformation processes are undertaken. The first process focuses on standardizing and normalizing the data, ensuring that all variables are on a consistent scale to facilitate accurate comparisons and computations during analysis. This step is crucial for improving the reliability and interpretability of the results derived from the data mining models. The second data transformation process involves feature engineering, where new features or variables are created from existing data to enhance predictive accuracy and model performance [24]. Techniques such as dimensionality reduction, categorical variable encoding, and creating interaction variables are applied to enrich the dataset with additional insights that can uncover hidden patterns and relationships. The project aims to optimize the dataset for effective data mining exploration by implementing these data transformation processes. This approach prepares the data for sophisticated analytical techniques and maximizes the potential for uncovering actionable insights that can drive informed decision-making and strategic initiatives. The systematic transformation of data ensures that the final analysis is robust, reliable, and capable of effectively generating valuable outcomes that meet the project's objectives.

3.3.1. Split Data

Using the Split Data operator, a classification method is applied to effectively divide the dataset into two subsets: training data and testing data. This division aims to facilitate a thorough evaluation and comparison of model performance. The training data subset is used to train the model on patterns and relationships in the data, while the testing data subset serves as an independent data set to assess how well the model generalizes to new, unprecedented data [25]. By systematically comparing the model's predictions against known outcomes in the test data, analysts can gauge the accuracy and reliability of the model before applying it in practical applications. This approach ensures that the model's predictive capabilities are rigorously validated, thus providing confidence in its effectiveness for making informed decisions based on data-driven insights, as shown in Table 1. The dataset containing 3400 records for determining eligibility status among the poor is divided into training and testing subsets. This division allocates 80% of the data for training purposes, where models are developed, and parameters are tuned to learn patterns and relationships within the dataset. The remaining 20% is reserved for testing the trained models, allowing for an independent evaluation of their performance on unseen data. This approach ensures that the models generalize well to new data and provides a reliable assessment of their predictive capabilities before deployment in practical applications, as shown in Table 2.

Table 1. Split Data Process

No	Wealth Decile	Gender	Job	Education	House Ownership	Having Savings	Roof Type	Wall Type	Floor Type
1	3	Male	Entrepreneur	High school graduate/equivalent	Self-owned	No	Asbestos/zinc	Wood/Board	Wood/Board
2	3	Male	Freelancer	Elementary school graduate/equivalent	Self-owned	No	Asbestos/zinc	Wood/Board	Wood/Board
3	3	Male	Merchant	Elementary school graduate/equivalent	Self-owned	No	Asbestos/zinc	Wall	Cement
4	3	Male	Entrepreneur	Elementary school graduate/equivalent	Self-owned	No	Asbestos/zinc	Other	Cement
5	2	Male	Private Employee	Junior high school graduate/equivalent	Self-owned	Yes	Asbestos/zinc	Zinc	Cement
6	2	Male	Entrepreneur	Elementary school graduate/equivalent	Self-owned	No	Asbestos/zinc	Wood/Board	Cement
...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...
3439	1	Male	Farmer	Elementary school graduate/equivalent	Self-owned	No	Asbestos/zinc	Wood/Board	Cement
3440	1	Male	Farmer	Not/not yet in school	Self-owned	No	Asbestos/zinc	Wall	Cement
3441	3	Male	Farmer	Junior high school graduate/equivalent	Self-owned	Yes	Asbestos/zinc	Wall	Ceramic
3442	1	Male	Fisherman	Elementary school graduate/equivalent	Self-owned	yes	Asbestos/zinc	Wall	Cement

Table 2. Split data parameters

Data Training	Data Testing
80%	20%

3.4. Data Mining

RF algorithm is a powerful ensemble learning method used for classification tasks [26]. It builds multiple decision trees for classification tasks during training and outputting class modes. When using RapidMiner, a comprehensive data mining platform, you can implement the RF, Naïve Bayes, or KNN algorithm through a series of systematic steps. Here is a detailed guide on implementing the RapidMiner tool's RF algorithm.

Performance

In RapidMiner, the performance operator is an important tool used to evaluate the effectiveness of predictive models by measuring accuracy, precision, recall, F1-score, and other performance metrics [27]. This operator comprehensively assesses how well the model predicts outcomes on training and testing datasets. By integrating the Performance operator into the workflow, users can obtain a detailed evaluation of the model's strengths and weaknesses, enabling a better understanding of its predictive capabilities. This operator generates performance metrics that facilitate comparison between different models or configurations, thus aiding in selecting the best model for a given task. In addition, the tool supports different types of performance evaluation, such as classification, regression, and clustering, making it versatile for various machine-learning tasks. Insights gained from using the Performance operator enable data scientists to make informed decisions in model customization, improving model accuracy and reliability, as shown in Figure 4.

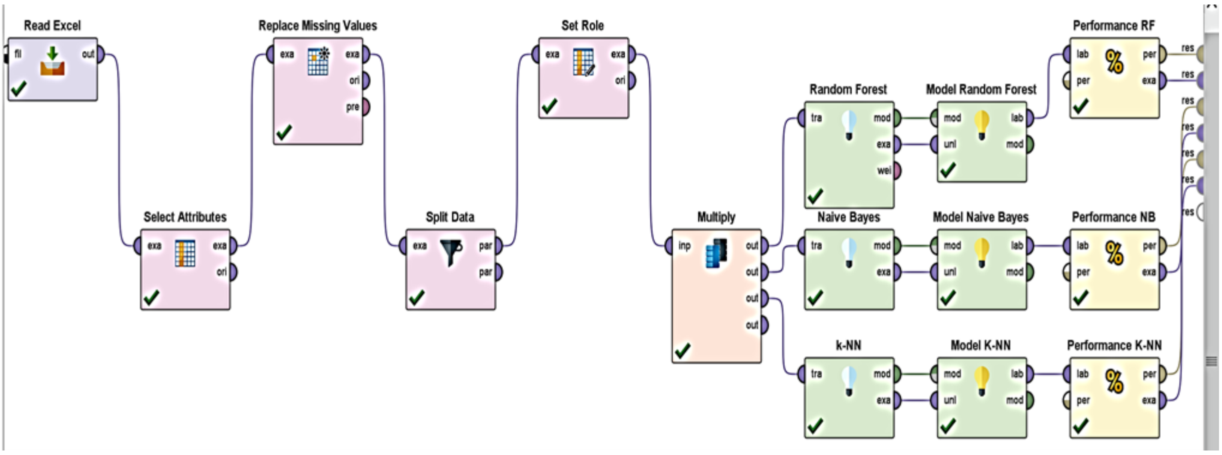

Figure 4. Use of Performance Classification Operators

3.5. Evaluation

The evaluation stage in data mining and machine learning is critical for determining the quality and effectiveness of predictive models [28]. This stage involves systematically assessing the model's performance using various metrics and techniques to ensure the model meets the desired objectives and can generalize well to unseen data. Evaluation provides insight into the model's predictive power, helps identify areas for improvement, and validates its applicability in real-world scenarios resulting from the RF, KNN, and Naïve Bayes algorithms, as seen in Figure 5, Figure 6, and Figure 7. The following is a summary table of the results of measuring accuracy, precision, and recall; as seen in Table 3, the RF method or algorithm has the highest accuracy of the four methods.

accuracy: 82.56%

	true 3	true 2	true 1	class precision
pred. 3	790	111	83	80.28%
pred. 2	11	436	17	93.97%
pred. 1	113	145	1047	80.23%
class recall	86.43%	63.01%	91.28%	

Figure 5. RF Performance Model

accuracy: 70.94%

	true 3	true 2	true 1	class precision
pred. 3	703	174	185	66.20%
pred. 2	97	392	104	66.10%
pred. 1	114	126	858	78.14%
class recall	76.91%	56.65%	74.80%	

Figure 6. KNN Performance model

accuracy: 53.47%

	true 3	true 2	true 1	class precision
pred. 3	553	294	235	51.11%
pred. 2	51	88	81	40.00%
pred. 1	310	310	831	57.27%
class recall	60.50%	12.72%	72.45%	

Figure 7. Naïve Bayes Performance Model

Table 3. Comparison of Method Performance Results

Method	Accuracy	Precision	Recall
RF	82.56 %	80.28 %	86.43%
KNN	70.94 %	66.20%	76.91%
Naïve Bayes	53.47 %	51.11%	60.50%

4. CONCLUSION

The use of the RF algorithm in classifying poverty status in North Lombok District demonstrates a high level of effectiveness, with an accuracy rate of 82.56%, compared to KNN with 70.94% and the Naïve Bayes model with 53.47%. This indicates that the RF model can distinguish various poverty status categories based on the features included. The high accuracy rate suggests that the model can capture the underlying patterns in the data, making it a valuable tool to inform policy decisions and target social assistance programs more effectively. However, there is room for improvement to enhance the model's performance and ensure more accurate and reliable predictions. Further tuning the hyperparameters of the Random Forest algorithm and experimenting with other advanced algorithms, such as GBM or XGBoost, could potentially achieve better performance.

5. ACKNOWLEDGEMENTS

We would like to express our deepest gratitude to all parties involved in this research for their valuable contributions and support. We are especially grateful to the reviewers for their insightful comments and suggestions, which have greatly improved the quality and presentation of this manuscript.

6. DECLARATIONS

AUTHOR CONTIBUTION

All authors contributed to the writing of this article.

FUNDING STATEMENT

-

COMPETING INTEREST

The author declares no conflict of interest.

REFERENCES

[1] M. M. Jamadar and K. S. Sridhar, "Urban poverty beyond 'slums': Mapping its multidimensionality," pp. 1–19, https://doi.org/10.1080/23792949.2024.2352550.

[2] S. E. Widodo and R. Wulandari, "Poverty and Public Policy: Local Government Efforts to Reduce Extreme Poverty," vol. 5, no. 1, pp. 93–104, https://doi.org/10.35326/jsip.v5i1.5067.

[3] A. Tenri and M. Hakim, "Impact of the Social Assistance Program on Poverty Levels in Urban Areas," vol. 6, no. 1, pp. 789–797.

[4] K. R. Lekobane and G. Ton, "Does social protection reach those left behind: Empirical evidence from Botswana using multidimensional poverty approaches," pp. 1–15, https://doi.org/10.1080/19439342.2024.2355635.

[5] T. O. Jejeniwa, N. Z. Mhlongo, and T. O. Jejeniwa, "AI Solutions for Developmental Economics: Opportunities and Challenges in Financial Inclusion and Poverty Alleviation," vol. 6, no. 4, pp. 108–123, https://doi.org/10.51594/ijae.v6i4.1073.

[6] F. N. Umma, B. Warsito, and D. A. I. Maruddani, "Klasifikasi Status Kemiskinan Rumah Tangga Dengan Algoritma C5.0 di Kabupaten Pemalang," vol. 10, no. 2, pp. 221–229, https://doi.org/10.14710/j.gauss.v10i2.29934.

[7] L. O. Faujan and N. Agustina, "Analisis Faktor-Faktor yang Memengaruhi Status Kemiskinan Ekstrem Rumah Tangga di Provinsi Maluku Tahun 2021," vol. 2023, no. 1, pp. 343–352, https://doi.org/10.34123/semnasoffstat.v2023i1.1639.

[8] A. Hariyanto, B. Juanda, E. Rustiadi, and S. Mulatsih, "The Effectiveness of Village Funds in Alleviating Rural Poverty: A Case Study of Belitung Regency," vol. 39, no. 1, pp. 197–208, https://doi.org/10.29313/mimbar.v39i1.2309.

[9] A. Artino, B. Juanda, and S. Mulatsih, "Keterkaitan Dana Desa terhadap Kemiskinan di Kabupaten Lombok Utara," vol. 21, no. 3, p. 381, https://doi.org/10.14710/tataloka.21.3.381-389.

[10] M. Sari and D. Susianto, "Decision Support System for Determining Indigent Public Health Insurance Participants With Weighted Product Method in Pringsewu," vol. 7, no. 1, p. 70, https://doi.org/10.56327/ijiscs.v7i1.1556.

[11] D. Kardeti, R. E. Agiati, P. Pribowo, A. Alfrojems, D. A. G. Pratiwi, and E. Susanto, "The Integrated Social Protection for the Poor in an Autonomous Regency of West Java Indonesia," vol. 4, no. 12, pp. 3640–3646, https://doi.org/10.47191/ijsshr/v4-i12-25.

[12] A. S. Rizalitaher and C. Bisri, "Penentuan Tingkat Kemiskinan Masyarakat Menggunakan Metode MOORA," vol. 3, no. 1, pp. 34–46, https://doi.org/10.55537/jibm.v3i1.680.

[13] F. Ramayanti, D. Vionanda, D. Permana, and Z. Zilrahmi, "Application of Random Forest to Identify for Poor Households in West Sumatera Province," vol. 1, no. 2, pp. 97–104, https://doi.org/10.24036/ujsds/vol1-iss2/31.

[14] F. Fauziah, M. A. Tiro, and F. Ruliana, "Comparison of k-Nearest Neighbor (k-NN) and Support Vector Machine (SVM) Methods for Classification of Poverty Data in Papua," vol. 2, no. 2, pp. 83–91, https://doi.org/10.35877/mathscience741.

[15] A. Alsharkawi, M. Al-Fetyani, M. Dawas, H. Saadeh, and M. Alyaman, "Poverty Classification Using Machine Learning: The Case of Jordan," vol. 13, no. 3, p. 1412, https://doi.org/10.3390/su13031412.

[16] R. Riliandhita, I. Maulana, and P. Purwanto, "Klasifikasi Penentuan Status Kemiskinan Penduduk Kelurahan Karangpawitan Karawang Menggunakan Metode C4.5," vol. 8, no. 2, pp. 1791–1796, https://doi.org/10.36040/jati.v8i2.9219.

[17] A. Fatikhurrizqi and B. D. Kurniawan, "Peran Bantuan Sosial dalam Pengentasan Kemiskinan Ekstrem di Jawa Timur Tahun 2020," vol. 2022, no. 1, pp. 1027–1036, https://doi.org/10.34123/semnasoffstat.v2022i1.1322.

[18] D. P. Sari, "Penerapan Metode Weighted Product untuk Penentuan Penerima Bansos kepada Masyarakat Terdampak COVID-19," vol. 9, no. 01, pp. 5–10, https://doi.org/10.33884/jif.v9i01.2714.

[19] A. M. Asrandi T, S. A. Wati, A. Wahab, and A. Alfian, "Efektivitas Program Sistem Informasi Kesejahteraan Sosial (SIKS-Ng) dalam Mendukung Program SLRT dan Puskesos Dinas Sosial Provinsi Sulawesi Selatan," vol. 3, no. 09, pp. 1294–1305, https://doi.org/10.36418/jiss.v3i09.695.

[20] F. Alghifari and D. Juardi, "Penerapan Data Mining Pada Penjualan Makanan Dan Minuman Menggunakan Metode Algoritma Naïve Bayes: Studi Kasus : Makan Barbeque Sepuasnya," vol. 9, no. 2, pp. 75–81, https://doi.org/10.33884/jif.v9i02.3755.

[21] M. F. Julianto, S. W. Hadi, S. Setiaji, W. Gata, and R. Pebrianto, "Clustering Pencapaian Target Penjualan Rumah Para Karyawan Marketing Menggunakan Rapid Miner dan Algoritma K-Means," vol. 8, no. 2, pp. 79–85, https://doi.org/10.31294/bi.v8i2.8189.

[22] R. S. P. Lubis, "Visualization Evaluation With the Rapid Miner Application Using the C4.5 Algorithm," vol. 11, no. 03, pp. 173–179.

[23] D. Rifaldi, Abdul Fadlil, and Herman, "Teknik Preprocessing Pada Text Mining Menggunakan Data Tweet "Mental Health"," vol. 3, no. 2, pp. 161–171, https://doi.org/10.51454/decode.v3i2.131.

[24] M. A. Jassim and S. N. Abdulwahid, "Data Mining preparation: Process, Techniques and Major Issues in Data Analysis," vol. 1090, no. 1, p. 012053, https://doi.org/10.1088/1757-899X/1090/1/012053.

[25] R. Oktafiani, A. Hermawan, and D. Avianto, "Pengaruh Komposisi Split data Terhadap Performa Klasifikasi Penyakit Kanker Payudara Menggunakan Algoritma Machine Learning," vol. 9, no. 1, pp. 19–28, https://doi.org/10.34128/jsi.v9i1.622.

[26] B. Alhajahmad and M. Ataş, "Boosting Predictive Power: Random Forest and Gradient Boosted Trees in Ensemble Learning," in *Proceeding Book of 2nd International Conference on Contemporary Academic Research ICCAR 2023*. All Sciences Academy, https://doi.org/10.59287/as-proceedings.133.

[27] M. F. Mustapha, A. N. I. Zulkifli, O. Kairan, N. N. S. M. Zizi, N. N. Yahya, and N. M. Mohamad, "The prediction of student's academic performance using RapidMiner," vol. 32, no. 1, pp. 363–371, https://doi.org/10.11591/ijeecs.v32.i1.pp363-371.

[28] A. Soni, C. Arora, R. Kaushik, and V. Upadhyay, "Evaluating the Impact of Data Quality on Machine Learning Model Performance," vol. 14, no. 1, pp. 13–18, https://doi.org/10.36893/JNAO.2023.V14I1.0013-0018.

[This page intentionally left blank.]