DOI: https://doi.org/10.34123/jurnalasks.v17i2.866

POLITEKNIK STATISTIKA STIS

Jurnal Aplikasi Statistika & Komputasi Statistik
Vol. 17 No. 2, December 2025
e- ISSN: 2615-1367, p-ISSN:2086-4132

Available online at https://jurnal.stis.ac.id/

Integrating Multi-Criteria Decision Analysis and Machine Learning for Fine-Scale Mapping of Safe Drinking Water Access in Bengkulu Province, Indonesia

Andrew Maruli Tua Tampubolon[1*], Bony Parulian Josaphat[2], Asriadi Sakka[3], Yohanes Wahyu Trio Pramono[4]

[1,2]Politeknik Statistika STIS, Jakarta, Indonesia, [3]Korea National University of Science and Technology (UST), Daejeon, Republic of Korea, [4]School of Information Management, Victoria University of Wellington, New Zealand
**Corresponding Author: E-mail address: 222111894@stis.ac.id*

ARTICLE INFO

Article history:
Received 12 December, 2025
Revised 21 December, 2025
Accepted 23 December, 2025
Published 31 December, 2025

Keywords:
Big Data; Machine Learning; Multi-Criteria Decision Analysis; Safe Drinking Water Access; Satellite Imagery

Abstract

Introduction/Main Objectives: This study aims to develop a 1 km × 1 km level estimation model of safe drinking water access using multisource satellite imagery, point of interest (POI), and aquifer productivity maps. **Background Problems:** There is a lack of alternative data sources for estimating safe drinking water access that are cost-, time-, and labor-efficient while maintaining high accuracy and frequent updates. **Novelty:** This study integrates Multi-Criteria Decision Analysis (MCDA) and machine learning methods to estimate and map safe drinking water access at a 1 km × 1 km resolution. **Research Methods:** Multisource geospatial data were used to construct the model. Within the MCDA approach, the Weighted Product Model (WPM) was employed to develop the Safe Drinking Water Access Index (SDWAI). Meanwhile, the machine learning regression algorithms Adaptive Boosting Regression (ABR) and Gradient Boosting Regression (GBR) were applied to estimate safe drinking water access at a fine spatial scale. The study was conducted in Bengkulu Province, Indonesia. **Finding/Results:** WPM yielded the best MCDA performance (R^2 = 0.3699, RMSE = 10.6566, MAE = 9.5427, MAPE = 0.1405), while ABR showed the best machine learning performance (R^2 = 0.4361, RMSE = 10.0813, MAE = 8.3750, MAPE = 0.1333).

1. Introduction

Water constitutes a vital element for human well-being and the advancement of development [1], [2]. Clean water refers to water of sufficient quality suitable for consumption and daily use, including sanitation [3]. The growing population has increased demand for clean water, prompting the Indonesian government to enhance access to safe water across various regions [4]. Adequate drinking water availability is vital for supporting healthy and sustainable living [5]. However, approximately 25% of the global population still lacks access to safe water [6]. In Indonesia, drinking water quality standards are regulated by Minister of Health Regulation No. 492 of 2010, mandating that water must be physically, chemically, and microbiologically safe [6] – [8], with providers responsible for ensuring safety throughout distribution [7].

According to WHO/UNICEF, access to safe drinking water includes sources capable of providing safe water such as piped supplies, bore wells, protected springs, and rainwater harvesting [9] – [11]. The Sustainable Development Goals (SDGs) emphasize universal access to safe and affordable drinking water [12]. Despite this, challenges to equitable access in Indonesia include limited funding, insufficient infrastructure, capacity constraints, and natural disaster risks [13]. The Ministry of Public Works and Housing (PUPR) targets increased household connections to piped water networks, requiring accurate and detailed spatial data at the household level to prioritize interventions [14], [15].

Detailed data availability is also essential for village governments to manage safe drinking water and plan village fund allocations per recent regulations [16], [17]. Currently, access data primarily derives from the National Socio-Economic Survey (Susenas), which is costly and limited to district/city levels, complicating priority area identification and program evaluation [18], [19]. As an alternative, satellite imagery, points of interest (POI), and other spatial data offer advantages in speed, cost-efficiency, ease of collection, and spatial detail [20]. Aquifer productivity maps further support assessment of regional groundwater capacity for clean water supply [21]. To support comprehensive spatial decision-making, Multi-Criteria Decision Analysis (MCDA) integrates relevant indicators into composite indices for policy guidance [22]. However, MCDA faces limitations with high-dimensional and multicollinear data [23]; thus, machine learning techniques complement MCDA by effectively extracting patterns from complex datasets [21], [24].

Recent studies have identified several geospatial indicators as proxies for understanding safe drinking water access. Elevation serves as a proxy for piped infrastructure accessibility, as high-altitude areas often face pipeline construction challenges [25]. Slope indicates groundwater potential [26], while the Topographic Wetness Index (TWI) relates to surface water utilization, a major water source for household piped systems in Indonesia [27]. Night-time Light (NTL) proxies electricity consumption, indicating potential access to electric water pumps, with proven efficacy in developing countries [28]. The Human Settlement Index (HSI) reflects urban areas with high socio-economic activities [29]. Carbon monoxide (CO) pollution correlates with economic growth and poverty, with poorer areas often lacking private water infrastructure such as pipes, pumps, or protected wells [25], [30]. Land Surface Temperature (LST), linked to urban heat island effects, distinguishes urban and rural areas, correlating with disparities in safe water access—urban residents commonly use piped water, while rural populations rely on groundwater or unprotected sources [31].

Additionally, points of interest (POI) related to economic facilities serve as proxies because communities without piped water often depend on accessible economic outlets such as bottled water vendors [32]. POI distance water bodies and POI distance IPA are used in modeling access to safe drinking water [25], [33], while aquifer productivity maps remain critical for assessing groundwater availability [21]. To complement official data on safe drinking water access, this study develops an estimation mapping approach using multisource remote sensing, POI, and aquifer productivity maps. This method aggregates data at a fine 1 x 1 km scale to produce detailed maps that can be updated more efficiently. Consequently, policy decisions aimed at improving equitable access to safe drinking water are expected to be better informed and more effective.

2. Material and Methods

2.1. Study Area

Bengkulu Province is located to the west of the Bukit Barisan mountain range. It comprises 10 regencies/municipalities and 129 districts. In Bengkulu, protected wells are the most widely used primary source of drinking water for households, accounting for 29.72 percent. According to data from BPS-Statistics Indonesia in 2022, Bengkulu ranks as the second lowest province in Indonesia in terms of the percentage of access to safe drinking water, following Papua.

2.2. Data Source

In this study, satellite imagery, points of interest (POI), and other geospatial data were used to develop an index and estimation model for access to safe drinking water. The datasets are described in detail in Table 1.

Table 1. Data summary

Variable	Spatial Resolution	Temporal Resolution	Data Source	References
Elevation	30 m	-	NASA SRTM	[25], [33], [34]
Slope	30 m	-	NASA SRTM	[25], [33], [34]
Topographic Wetness Index (TWI)	30 m	-	NASA SRTM	[35]
Human Settlement Index (HSI)	463,83 m	monthly	Sentinel-2, Suomi-NPP VIIRS	[34]
Night-time Light (NTL)	463,83 m	monthly	Suomi-NPP VIIRS	[15], [33]
Carbon Monoxide (CO)	1113,2 m	< 1 day	Sentinel-5P	[30]
Land Surface Temperature (LST)	1000 m	1 day	MODIS	[36], [37]
POI Distance Water Bodies	Point	dynamically	Sentinel-2, OpenStreetMap	[25], [33], [38]
POI Distance & Density Ekonomic	Point	dynamically	OpenStreetMap	[33]
POI Distance IPA	Point	dynamically	Ministry of Public Works and Housing	[25]
Aquifer Productivity	100 m	-	Ministry of Energy and Mineral Resources	[21]
Population Grid	100 m	-	Worldpop	[22], [25], [33]
Percentage of households with access to safe drinking water	Regency, district	2022	BPS-Statistics Indonesia	

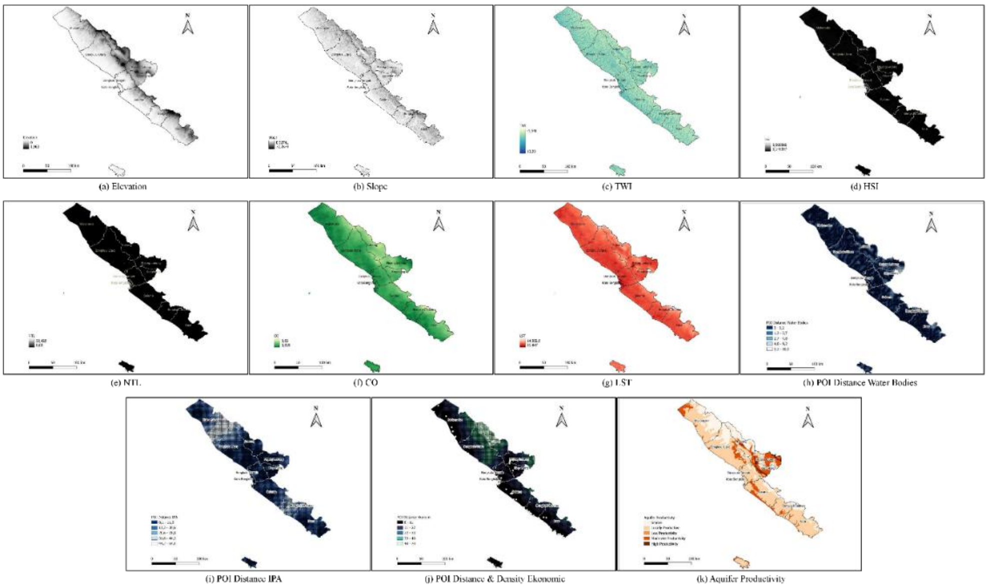

Figure 1. Data visualization

This research utilizes remote sensing data, specifically multisource satellite imagery, combined with spatial POI data obtained from OpenStreetMap (OSM). The satellite imagery sources used in this study comprise Elevation, Slope, and the Topographic Wetness Index (TWI) derived from NASA's SRTM; Nighttime Light (NTL) and Human Settlement Index (HSI) obtained from Suomi-NPP; Carbon Monoxide (CO) data from Sentinel-5P; and Land Surface Temperature (LST) data from MODIS. POI data were collected from both OpenStreetMap (OSM) and the Ministry of Public Works. Additional geospatial datasets included the Aquifer Productivity Map sourced from the Ministry of Energy and Mineral Resources and population distribution data from WorldPop. The data visualizations used in this study are presented in Figure 1.

2.3. Research Framework

This research generated a map of safe drinking water access at a 1 km × 1 km spatial resolution to offer a more efficient and update-ready alternative to household survey data, while maintaining accuracy and detailed spatial representation. The methodological steps involved data acquisition, preprocessing, integration and transformation, correlation analysis and variable selection, model construction, and subsequent validation and interpretation. Analysis and visualization were carried out using QGIS 3.26.2 and Python 3.12. The expected outcome is a spatially detailed 1 km × 1 km map of access to safe drinking water, accompanied by its validation results, as illustrated in Figure 2.

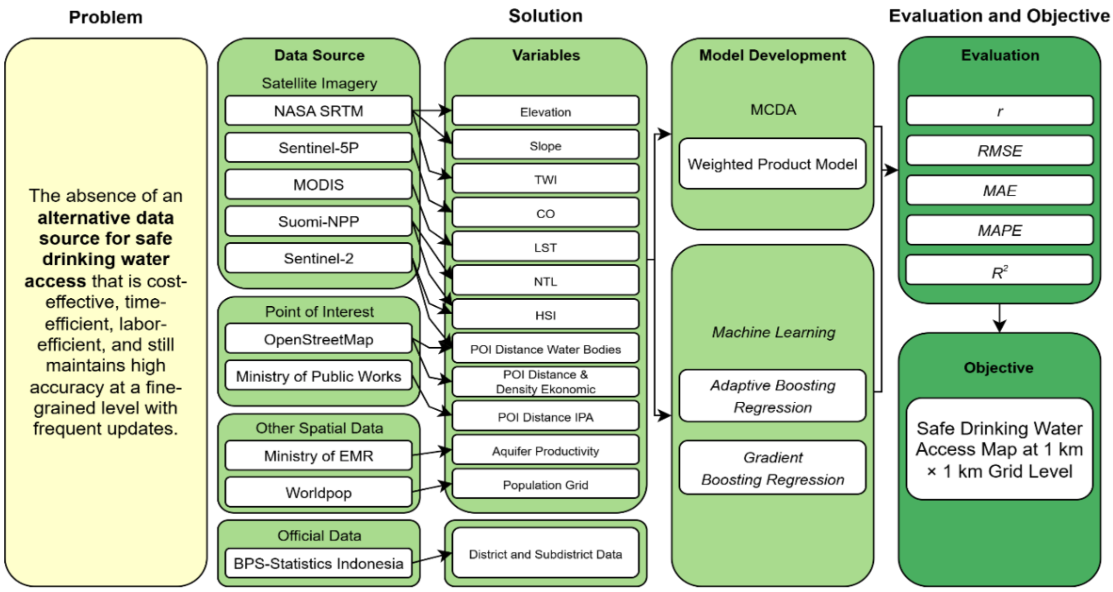

Figure 2. The research framework

2.4. Data Transformation

Feature extraction from the dataset produced median values for each 1 km × 1 km geographic zone across various variables. Subsequently, the values were normalized using the Yeo–Johnson power transformation, a method capable of processing variables with either positive or negative ranges [39]. This transformation aims to approximate a normal (Gaussian) distribution, thereby enhancing analytical and modeling performance [40], as formulated in Equation (1) [39].

$$y_\lambda(x) = \begin{cases} \frac{(1+x)^{\lambda}-1}{\lambda}, \lambda\neq 0, x\geq 0 \\ \log(1+x), \lambda=0, x\geq 0 \\ -\frac{(1+x)^{2-\lambda}}{2-\lambda}, \lambda\neq 2, x<0 \\ -\log(1-x), \lambda=2, x<0 \end{cases} \quad (1)$$

In this transformation, x denotes the variable value or input data, and λ is a parameter estimated using the Maximum Likelihood method under the assumption of normally distributed variables. When $\lambda = 1$, the transformation is linear; for $\lambda < 1$, it compresses the right tail and stretches the left tail, making a right-skewed distribution more symmetric; and for $\lambda > 1$, it adjusts a left-skewed distribution toward symmetry.

2.5. Pearson Correlation

The Pearson correlation coefficient (r) has a value range from 0 to 1, with the sign indicating whether the relationship is positive or negative. The interpretation of the strength of the Pearson correlation is shown in Table 2 [41].

Table 2. Interpretation of correlation coefficient

Range of Values	Level of Intensity
$0.00 \leq \|\boldsymbol{r}\| \leq 0.199$	Very Weak
$0.20 \leq \|\boldsymbol{r}\| \leq 0.399$	Weak
$0.40 \leq \|\boldsymbol{r}\| \leq 0.599$	Strong Enough
$0.60 \leq \|\boldsymbol{r}\| \leq 0.799$	Strong
$0.80 \leq \|\boldsymbol{r}\| \leq 1.000$	Very Strong

2.6. Multi-Criteria Decision Analysis (MCDA)

Safe Drinking Water Access Index (SDWAI) was constructed using a Multi-Criteria Decision Analysis (MCDA) approach with the Weighted Product Model (WPM). In WPM, the calculation involves a multiplicative process, where each alternative's value is raised to the power of its corresponding criterion weight and then multiplied across all criteria to obtain the final score for each alternative [42]. The weight of each indicator was determined based on its Pearson correlation coefficient. The model is formulated as follows:

$$WPM_j = \prod_{i=1}^{n} |X_{ij}|^{w_i} \tag{2}$$

where WPM_j represents the WPM index value for the j-th grid, X_{ij} denotes the i-th indicator at the j-th grid level, and w_i refers to the weight of the i-th indicator.

2.7. Machine Learning

The estimation model was developed using a machine learning approach, which is widely applied in estimation tasks due to its focus on achieving high accuracy—an essential requirement in predictive modeling [43]. In this approach, the algorithms employed include Adaptive Boosting Regression (ABR) and Gradient Boosting Regression (GBR). Optimal parameters and hyperparameters for each model were determined through a grid search combined with 5-fold cross validation.

Model development was conducted at the sub-district level using data from two different time periods. Data from 2021 were used for model training, while data from 2022 were reserved exclusively for model testing. This temporal holdout (time-based split) strategy was adopted to prevent data leakage and to provide a more realistic assessment of model performance when applied to future data [35]. Such a separation between training and testing datasets is commonly recommended in predictive modeling to ensure the generalizability of the developed models [44].

2.8. Model Evaluation

The evaluation metrics used include RMSE, MAE, MAPE, Pearson's correlation coefficient, and the coefficient of determination (R^2). RMSE, MAE, and MAPE are used to quantify the magnitude of error between predicted and observed values, where lower values indicate higher prediction accuracy [45]. Pearson's correlation coefficient is employed to assess the strength of the linear relationship between the estimated values and official statistics, while R^2 measures the proportion of variance in the official data explained by the model. Together, these metrics provide a comprehensive evaluation of model performance by capturing both prediction accuracy and the consistency of statistical and spatial patterns relative to official data. The formulations of RMSE, MAE, MAPE, and R^2 are presented in Equations (3) to (6).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{3}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{\hat{y}_i}\right| \tag{4}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{5}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{6}$$

3. Results and Discussion

3.1. Variable Selection

Before constructing the Safe Drinking Water Access Index (SDWAI), a Pearson correlation analysis was conducted to examine the relationship between spatially derived variables and the percentage of access to safe drinking water. All variables were aggregated to the regency/municipality level and transformed using the Yeo–Johnson method to improve their distributional properties prior to analysis. Pearson's correlation coefficient was employed to quantify the strength and direction of linear relationships between the transformed variables and official safe drinking water access data. Variables exhibiting statistically significant linear correlations with the percentage of access to safe drinking water were considered suitable for inclusion in the SDWAI. The analysis was conducted using official statistics available for only 10 regencies/municipalities in Bengkulu Province.

Statistical significance was evaluated using a significance level of $\alpha = 0.15$ to reduce the risk of excluding potentially relevant variables, given the limited number of regency/municipality-level observations. Variables with p-values less than 0.15 were considered statistically significant. The significance test was used to identify variables exhibiting sufficiently strong empirical relationships for inclusion in the SDWAI. As shown in Table 3, most variables exhibited statistically significant correlations at the 15% level. However, the Topographic Wetness Index (TWI), POI Distance to Economic Facilities, and POI Distance to IPA did not meet this criterion and were therefore excluded from the SDWAI construction. The correlation coefficients of the retained variables were subsequently used as weights in the SDWAI formulation.

Table 3. Results of the Pearson correlation

Variable	Correlation	p-value	Significant
Elevation	-0.570	0.085	Yes
Slope	-0.508	0.134	Yes
Topographic Wetness Index (TWI)	0.486	0.154	No
Human Settlement Index (HSI)	0.514	0.129	Yes
Night-time light (NTL)	0.497	0.144	Yes
Carbon Monoxide (CO)	0.580	0.079	Yes
Land Surface Temperature (LST)	0.503	0.139	Yes
POI Distance Water Bodies	0.825	0.003	Yes
POI Distance Ekonomic	-0.189	0.601	No
POI Density Ekonomic	0.620	0.056	Yes
POI Distance IPA	0.284	0.427	No
Aquifer Productivity	-0.512	0.131	Yes

3.2. The SDWAI Evaluation and Development

The evaluation to determine the best SDWAI was conducted using correlation (r), R^2, RMSE, MAE, and MAPE, with the SDWAI aggregated at the regencies/municipalities level. Table 4 presents the evaluation results at this level. The correlation coefficient (r) of 0.6491 suggests that the estimated values and the official data are strongly positively correlated. The coefficient of determination (R^2) of 0.3699 suggests that approximately 36.99% of the variation in official safe drinking water access data can be explained by the WPM-based estimates.

Table 4. Results of the SDWAI evaluation at the regency/municipality level

Method	r	R^2	RMSE	MAE	MAPE
Weight Product Model (WPM)	0.6491	0.3699	10.6566	9.5427	0.1405

Access to safe drinking water is closely related to population distribution. The Safe Drinking Water Access Index (SDWAI) was generated by overlaying it with grid-level population data obtained from WorldPop. This approach allows the index to better reflect spatial variations in demand for drinking water services and the concentration of populations potentially affected by limited access. The following section displays the SDWAI map generated at a 1 km × 1 km spatial level, as shown in Figure 3.

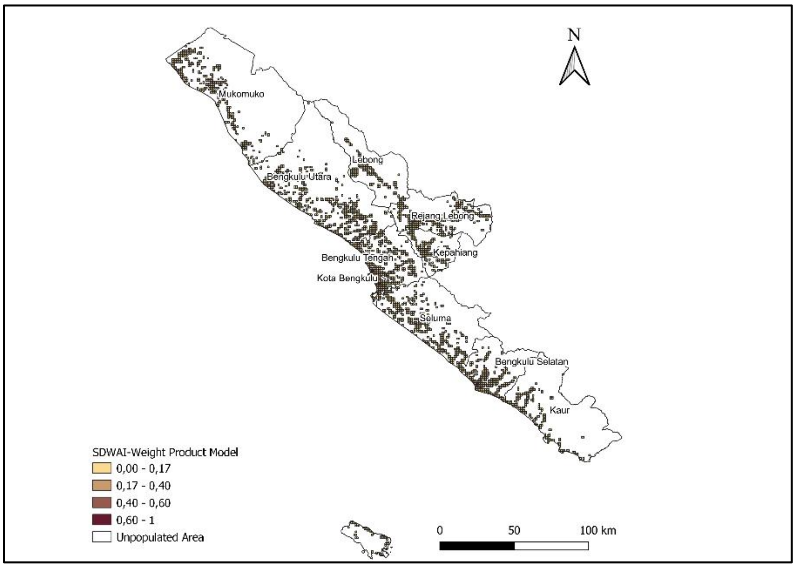

Figure 3. The obtained scaled SDWAI-WPM

The WPM-based mapping results show that areas with index values of 0.00–0.17 are generally located far from major activity centers, such as most of Kaur Regency and the western part of North Bengkulu Regency. Values of 0.17–0.40 are more widely distributed, covering large parts of Mukomuko, North Bengkulu, South Bengkulu, and Rejang Lebong Regencies. Values of 0.40–0.60 are concentrated around the center of South Bengkulu Regency and several locations in Bengkulu Municipality. Meanwhile, values of 0.60–1 are found only in limited areas in the city center of Bengkulu Municipality.

3.3. The Best Parameter/Hyperparameter Selection

The machine learning model was developed by selecting the optimal parameters and hyperparameters through a grid search process. Table 5 presents the best parameter and hyperparameter configurations derived from the 5-fold cross-validation performed with a random state of 42. This procedure ensures that the selected models achieve a balance between predictive accuracy and model generalizability while minimizing the risk of overfitting. The use of cross-validation further enhances the robustness of the parameter selection by evaluating model performance across multiple data partitions.

Each model was trained using grid search and 5-fold cross validation to determine the optimal hyperparameters. The model performance was evaluated through two validation approaches, which included 5-fold cross-validation and testing on a hold-out dataset, using RMSE, MAE, and MAPE as the evaluation metrics. The evaluation results are presented in Table 6.

Table 5. Results of hyperparameter selection

Model	Parameter/ Hyperparameter	Options	Selected
Adaptive Boosting Regression (ABR)	n_estimators	100, 500, 1000	100
	learning_rate	0.01, 0.1, 1.0	0.01
Gradient Boosting Regression (GBR)	n_estimators	500, 1000	500
	learning_rate	0.1, 0.01, 0.001	0.001
	max_depth	1, 3, 9, 12	1
	min_samples_split	8, 10, 12	8

Table 6. Results of the evaluation with 5-fold cross-validation and data testing

Model	RMSE		MAE		MAPE	
	5-*fold* CV	*Test*	5-*fold* CV	*Test*	5-*fold* CV	*Test*
ABR	0.9279	0.9709	0.7412	0.7896	0.0189	0.0103
GBR	0.9699	1.0228	0.8036	0.8574	0.0123	0.0104

Based on the RMSE, MAE, and MAPE values from the test data, the ABR model achieved the best performance, with an RMSE of 0.9709, MAE of 0.7896, and MAPE of 0.0103, the lowest among all models. This indicates that ABR was able to produce predictions that were closest to the actual values. The superior performance of ABR suggests its effectiveness in capturing complex, non-linear relationships between multisource geospatial predictors and safe drinking water access. Moreover, the ensemble-based nature of ABR enhances model stability by iteratively reducing prediction errors from weak learners.

Table 7. Results of the ABR evaluation at the regency/municipality level

Method	r	R^2	RMSE	MAE	MAPE
Adaptive Boosting Regression (ABR)	0.7706	0.4361	10.0813	8.3750	0.1333

As shown in Table 7, the correlation coefficient (r) of 0.7706 indicates a strong positive relationship between the estimated values and the official data. The coefficient of determination (R^2) of 0.4361 suggests that approximately 43.61% of the variation in official safe drinking water access data can be explained by the ABR model estimates. This level of explanatory power is notable given the inherent spatial heterogeneity and the limited availability of fine-scale ground truth data. The results indicate that the ABR model provides a reliable approximation of official statistics while enabling spatial disaggregation at a much finer resolution.

3.4. The Safe Drinking Water Access Mapping

The machine learning models were trained using data on the percentage of households with access to safe drinking water, enabling estimation at the 1 km × 1 km spatial grid. Since SDWAI values range from 0 to 1, the first step to enable comparison with the mapped estimates was to align their data ranges. The evaluation employed simple linear regression, assigning SDWAI as the independent variable and the official district-level estimates as the dependent variable. The resulting regression model was then applied to generate SDWAI estimates at the spatial grid. Figure 4 presents the estimated distribution of safe drinking water access derived from the SDWAI-WPM model, while Figure 5 displays the estimation outcomes generated using the ABR model.

SDWAI and machine learning maps show similar patterns. Areas with poorer access to safe drinking water are located in the northeast, which consists of hilly and mountainous terrain, while areas with better access are found in urban regions. The convergence of patterns across both approaches indicates that the spatial distribution of safe drinking water access is consistently captured by the index-based and data-driven models. Topographic constraints and limited infrastructure development in mountainous areas likely contribute to lower accessibility, whereas urban regions benefit from more developed piped water networks and supporting facilities. This consistency reinforces the reliability of the proposed framework in identifying spatial disparities in safe drinking water access at a fine resolution.

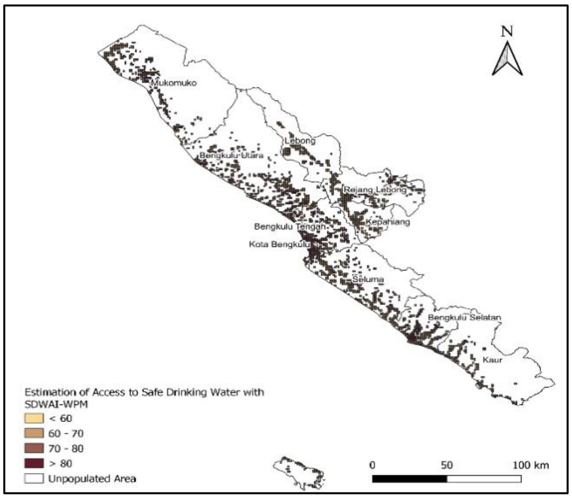

Figure 4. Results of the SDWAI-WPM estimation

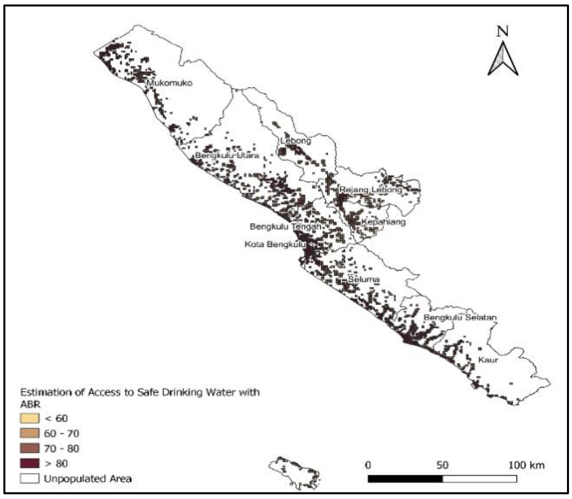

Figure 5. Results of the ABR estimation

3.5. SDWAI Ground Truth Analysis

Numerous previous studies have shown that ground-truth identification based on high-resolution imagery offers an evaluation perspective that cannot be fully represented by purely numerical, pixel-level validation. In this study, four randomly selected 1 km × 1 km grid cells were examined through visual interpretation using Google Earth imagery to characterize their geographic conditions.

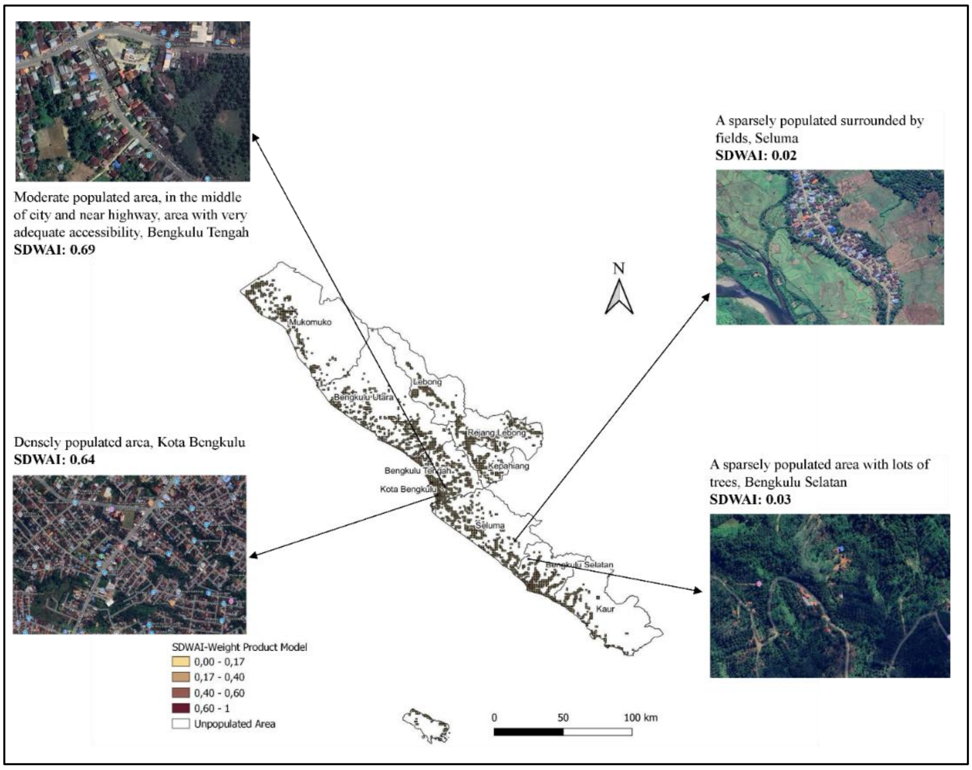

Figure 6. SDWAI (scaled index) ground-truth check

The results indicate that areas with high SDWAI scores tend to correspond to moderately to densely populated areas with more adequate accessibility (Figure 6). Urban areas generally exhibit higher SDWAI values, reflecting better availability of infrastructure and services. In contrast, areas with low SDWAI scores are more commonly associated with sparsely populated regions characterized by inadequate. These areas tend to be spatially deprived, with limited infrastructure and restricted access to basic services.

4. Conclusion

This study successfully achieved its objective of developing a 1 km × 1 km resolution estimation model of safe drinking water access by integrating multisource satellite imagery, point of interest (POI) data, and aquifer productivity maps. By combining Multi-Criteria Decision Analysis (MCDA) and machine learning approaches, the proposed framework enables the generation of fine-scale spatial estimates in regions where official statistics are available only at the regency/municipality level.

The Safe Drinking Water Access Index (SDWAI), developed using the Weighted Product Model (WPM), demonstrated a strong positive relationship with official data (r = 0.6491; R^2 = 0.3699), indicating that the MCDA-based approach provides a consistent and interpretable baseline for spatial disaggregation. The incorporation of machine learning further improved estimation performance, with Adaptive Boosting Regression (ABR) achieving the highest accuracy (r = 0.7706; R^2 = 0.4361). These results show that the complementary use of MCDA and machine learning enhances the robustness of the proposed estimation framework.

The proposed framework has strong potential for practical application, particularly in regions where safe drinking water statistics are available only at the regency/municipality level or are updated infrequently. The resulting 1 km × 1 km spatial estimates can support evidence-based planning by identifying priority areas for the development or expansion of drinking water facilities, enabling spatially targeted policy interventions, and informing spatial planning based on local clean water service needs. Moreover, this approach offers a cost-effective and scalable alternative to conventional survey-based methods, complementing official statistics for monitoring progress toward safe drinking water access targets in data-scarce regions.

Despite these contributions, this study has limitations. Validation is constrained by the availability of official data only at the regency/municipality level, resulting in an indirect assessment of the 1 km × 1 km estimates. In addition, field-based ground truth validation, which would ideally support fine-scale accuracy assessment, could not be conducted.

Ethics approval

Not required.

Acknowledgments

Not required.

Competing interests

All the authors declare that there are no conflicts of interest.

Funding

This study received no external funding.

Underlying data

This study uses secondary data obtained from BPS-Statistics Indonesia, Ministry of Public Works, Ministry of Energy and Mineral Resources, Satellite Imagery, and OpenStreetMap

Credit Authorship

Andrew Maruli Tua Tampubolon: Methodology, Software, Data Collection, Data Curation, Writing-Original Draft Preparation, Visualization. **Bony Parulian Josaphat:** Conceptualization, Supervision, Reviewing and Editing. **Asriadi Sakka:** Reviewing, Validation. **Yohanes Wahyu Trio Pramono:** Reviewing, Final Draft.

References

[1] V. Sheel, A. Kotwal, N. Dumka, V. Sharma, R. Kumar, and V. Tyagi, "Water as a social determinant of health: bringing policies into action," *J Glob Health Rep*, vol. 8, 2024, doi: 10.29392/001c.92160.

[2] G. Mujtaba, M. U. H. Shah, A. Hai, M. Daud, and M. Hayat, "A holistic approach to embracing the United Nation's Sustainable Development Goal (SDG-6) towards water security in Pakistan," Jan. 01, 2024, *Elsevier Ltd*. doi: 10.1016/j.jwpe.2023.104691.

[3] J. Abellán and J. A. Alonso, "Promoting global access to water and sanitation: A supply and demand perspective," *Water Resour Econ*, vol. 38, Apr. 2022, doi: 10.1016/j.wre.2022.100194.

[4] A. Umami, H. Sukmana, E. A. Wikurendra, and E. Paulik, "A review on water management issues: potential and challenges in Indonesia," *Sustain Water Resour Manag*, vol. 8, no. 3, Jun. 2022, doi: 10.1007/s40899-022-00648-7.

[5] M. L. Crouch, H. E. Jacobs, and V. L. Speight, "Defining domestic water consumption based on personal water use activities," *Aqua Water Infrastructure, Ecosystems and Society*, vol. 70, no. 7, pp. 1002–1011, Nov. 2021, doi: 10.2166/aqua.2021.056.

[6] G. E. Adjovu, H. Stephen, D. James, and S. Ahmad, "Overview of the application of remote sensing in effective monitoring of water quality parameters," Apr. 01, 2023, *MDPI*. doi: 10.3390/rs15071938.

[7] Kementerian Kesehatan, "Peraturan menteri kesehatan Republik Indonesia Nomor 492/Menkes/PER/IV/2010 tentang persyaratan kualitas air minum," Jakarta, 2010.

[8] Kementerian Kesehatan, "Peraturan menteri kesehatan Republik Indonesia Nomor 2 Tahun 2023 tentang peraturan pelaksanaan peraturan pemerintah Nomor 66 Tahun 2014 tentang kesehatan lingkungan," Jakarta, 2023. [Online]. Available: www.peraturan.go.id

[9] WHO and UNICEF, *Progress on drinking water, sanitation and hygiene: 2017 update and SDG baselines*. Geneva: World Health Organization and the United Nations Children's Fund, 2017.

[10] WHO and UNICEF, *Safely managed drinking water*. Geneva: World Health Organization and the United Nations Children's Fund, 2017.

[11] J. D. Tetteh *et al.*, "Spatial heterogeneity in drinking water sources in the Greater Accra Metropolitan Area (GAMA), Ghana," *Popul Environ*, vol. 44, no. 1–2, pp. 46–76, Sep. 2022, doi: 10.1007/s11111-022-00407-y.

[12] L. Ho, A. Alonso, M. A. Eurie Forio, M. Vanclooster, and P. L. M. Goethals, "Water research in support of the Sustainable Development Goal 6: A case study in Belgium," Dec. 20, 2020, *Elsevier Ltd*. doi: 10.1016/j.jclepro.2020.124082.

[13] UNICEF, "WASH Acts 2023," 2024.

[14] Kementerian PUPR, Kementerian PPN/Bappenas, USAID IUWASH PLUS, and SECO, "Peta jalan peningkatan kapasitas sumber daya manusia BUMD air minum," Jakarta, 2021.

[15] B. Ajisegiri *et al.*, "Geo-spatial modeling of access to water and sanitation in Nigeria," *Journal of Water Sanitation and Hygiene for Development*, vol. 9, no. 2, pp. 258–280, Jun. 2019, doi: 10.2166/washdev.2019.089.

[16] P. D. T. dan T. Kementerian Desa, "Peraturan menteri desa, pembangunan daerah tertinggal, dan transmigrasi Nomor 8 Tahun 2022 tentang prioritas penggunaan dana desa Tahun 2023," Jakarta, 2022.

[17] Hasbudin, W. O. Aswati, and Nirwana, "Evaluasi pengelolaan dana desa di Desa Bone Kancintala Kecamatan Bone Kabupaten Muna," *Jurnal Akuntansi dan Keuangan (JAK)*, vol. 6, no. 1, 2021, [Online]. Available: http://ojs.uho.ac.id/index.php/jak-uho/issue/archive

[18] Kementerian PPN/Bappenas, Pokja PPAS, and USAID IUWASH PLUS, "Meta data target indikator air minum: : Kupas Tuntas SDGs 6.1 Air Minum," Jakarta, 2021.

[19] BPS Kabupaten Tanjung Jabung Timur, "Statistik kesejahteraan rakyat Kabupaten Tanjung Jabung Timur 2022," Muara Sabak, 2022. [Online]. Available: www.freepik.com

[20] S. D. Kurniawan *et al.*, *Big data: Mengenal big data & implementasinya di berbagai bidang*. Jambi: PT. Sonpedia Publishing Indonesia, 2024. [Online]. Available: www.buku.sonpedia.com
[21] C. Abesser and M. Lewis, "A semi-quantitative technique for mapping potential aquifer productivity on the national scale: example of England and Wales (UK)," *Hydrogeol J*, vol. 23, no. 8, pp. 1677–1694, Dec. 2015, doi: 10.1007/s10040-015-1295-5.
[22] M. C. de O. Silva, R. S. Vasconcelos, and J. A. Cirilo, "Risk mapping of water supply and sanitary sewage systems in a City in the Brazilian semi-arid region using GIS-MCDA," *Water (Switzerland)*, vol. 14, no. 20, Oct. 2022, doi: 10.3390/w14203251.
[23] M. Guo, Q. Zhang, X. Liao, F. Y. Chen, and D. D. Zeng, "A hybrid machine learning framework for analyzing human decision-making through learning preferences," *Omega (Westport)*, vol. 101, Jun. 2021, doi: 10.1016/j.omega.2020.102263.
[24] C. M. Bishop, *Pattern recognition and machine learning*. New York: Springer, 2006.
[25] M. J. de S. Cordão, I. A. A. Rufino, P. Barros Ramalho Alves, and M. N. M. Barros Filho, "Water shortage risk mapping: a GIS-MCDA approach for a medium-sized city in the Brazilian semi-arid region," *Urban Water J*, vol. 17, no. 7, pp. 642–655, Aug. 2020, doi: 10.1080/1573062X.2020.1804596.
[26] M. Radulović *et al.*, "Assessment of groundwater potential zones using GIS and fuzzy AHP techniques—A case study of the titel municipality (Northern Serbia)," *ISPRS Int J Geoinf*, vol. 11, no. 4, Apr. 2022, doi: 10.3390/ijgi11040257.
[27] A. S. Suryani, "Persepsi masyarakat dalam pemanfaatan air bersih (Studi kasus masyarakat pinggir sungai di Palembang)," *Aspirasi*, vol. 7, no. 1, pp. 33–48, 2016, [Online]. Available: www.koran-sindo.
[28] B. Guo *et al.*, "Estimating socio-economic parameters via machine learning methods using luojia1-01 nighttime light remotely sensed images at multiple scales of China in 2018," *IEEE Access*, vol. 9, pp. 34352–34365, 2021, doi: 10.1109/ACCESS.2021.3059865.
[29] Y. Zheng, Q. Zhou, Y. He, C. Wang, X. Wang, and H. Wang, "An optimized approach for extracting urban land based on log-transformed dmsp-ols nighttime light, ndvi, and ndwi," *Remote Sens (Basel)*, vol. 13, no. 4, pp. 1–22, Feb. 2021, doi: 10.3390/rs13040766.
[30] Y. Wang *et al.*, "The impact of carbon monoxide on years of life lost and modified effect by individual- and city-level characteristics: Evidence from a nationwide time-series study in China," *Ecotoxicol Environ Saf*, vol. 210, Mar. 2021, doi: 10.1016/j.ecoenv.2020.111884.
[31] E. A. Adams and S. L. Smiley, "Urban-rural water access inequalities in Malawi: implications for monitoring the Sustainable Development Goals," *Nat Resour Forum*, vol. 42, no. 4, pp. 217–226, Nov. 2018, doi: 10.1111/1477-8947.12150.
[32] T. Prasetiawan, A. Nastiti, and B. S. Muntalif, "'Bad' piped water and other perceptual drivers of bottled water consumption in Indonesia," *Wiley Interdisciplinary Reviews: Water*, vol. 4, no. 4, Jul. 2017, doi: 10.1002/WAT2.1219.
[33] N. Dejito *et al.*, "Mapping access to water and sanitation in Colombia using publicly accessible satellite imagery, crowd-sourced geospatial information and randomrorests," in *35th Conference on Neural Information Processing Systems (NeurIPS 2021)*, Nov. 2021.
[34] K. Shi, Z. Chang, Z. Chen, J. Wu, and B. Yu, "Identifying and evaluating poverty using multisource remote sensing and point of interest (POI) data: A case study of Chongqing, China," *J Clean Prod*, vol. 255, May 2020, doi: 10.1016/j.jclepro.2020.120245.
[35] H. E. Winzeler, P. R. Owens, Q. D. Read, Z. Libohova, A. Ashworth, and T. Sauer, "Topographic wetness index as a proxy for soil moisture in a Hillslope Catena: Flow algorithms and map generalization," *Land (Basel)*, vol. 11, no. 11, 2022, doi: 10.3390/land11112018.
[36] P. Rao, P. Tassinari, and D. Torreggiani, "Exploring the land-use urban heat island nexus under climate change conditions using machine learning approach: A spatio-temporal analysis of remotely sensed data," *Heliyon*, vol. 9, no. 8, Aug. 2023, doi: 10.1016/j.heliyon.2023.e18423.
[37] Y. Bin Cai, Z. J. Wu, Y. H. Chen, L. Wu, and W. Bin Pan, "Investigate the difference of cooling effect between water bodies and green spaces: The study of fuzhou, China," *Water (Switzerland)*, vol. 14, no. 9, May 2022, doi: 10.3390/w14091471.
[38] S. Mishra and D. R. Mishra, "Normalized difference chlorophyll index: A novel model for remote estimation of chlorophyll-a concentration in turbid productive waters," *Remote Sens Environ*, vol. 117, pp. 394–406, Feb. 2012, doi: 10.1016/j.rse.2011.10.016.
[39] I. Yeo and R. A. Johnson, "A new family of power transformations to improve normality or symmetry," *Biometrika*, vol. 87, no. 4, pp. 954–959, 2000, [Online]. Available: http://biomet.oxfordjournals.org/

[40] J. Raymaekers and P. J. Rousseeuw, "Transforming variables to central normality," *Mach Learn*, vol. 113, no. 8, pp. 4953–4975, Aug. 2021, doi: 10.1007/s10994-021-05960-5.

[41] Sugiyono, *Metode penelitian pendidikan pendekatan kuantitatif, kualitatif dan R&D*. Alfabeta, 2010.

[42] Ch. M. Rao and K. Venkatasubbaiah, "Application of WSM, WPM and TOPSIS methods for the optimization of multiple responses," *International Journal of Hybrid Information Technology*, vol. 9, no. 10, pp. 59–72, Oct. 2016, doi: 10.14257/ijhit.2016.9.10.07.

[43] N. U. Moroff, E. Kurt, and J. Kamphues, "Machine Learning and Statistics: A study for assessing innovative demand forecasting models," in *Procedia Computer Science*, Elsevier B.V., 2021, pp. 40–49. doi: 10.1016/j.procs.2021.01.127.

[44] R. D. Stephens and S. H. Cadle, "Remote sensing measurements of carbon monoxide emissions from on-road vehicles," *J Air Waste Manage Assoc*, vol. 41, no. 1, pp. 39–46, 1991, doi: 10.1080/10473289.1991.10466823.

[45] A. Botchkarev, "Evaluating performance of regression machine learning models using multiple error metrics in Azure Machine Learning Studio," *SSRN Electronic Journal*, 2018, [Online]. Available: https://ssrn.com/abstract=3177507