



Available online at <http://scik.org>

Commun. Math. Biol. Neurosci. 2026, 2026:116

<https://doi.org/10.28919/cmbn/10276>

ISSN: 2052-2541

TEMPORALLY VALIDATED DISTRICT-LEVEL TUBERCULOSIS CASE-COUNT MODELING IN INDONESIA USING NEGATIVE BINOMIAL REGRESSION AND PSO-OPTIMIZED EXPLAINABLE ENSEMBLES

MAHDAYANI PUTRI YUNIZAR¹, SRI WINARNI^{2,*}, DEFI YUSTI FAIDAH²

¹Master's Program in Applied Statistics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Jl. Raya Bandung-Sumedang KM. 21, Jatinangor, Sumedang 45363, West Java, Indonesia

²Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Jl. Raya Bandung-Sumedang KM. 21, Jatinangor, Sumedang 45363, West Java, Indonesia

Copyright © 2026 the author(s). This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract. Indonesia has a high tuberculosis (TB) burden and heterogeneous district-level notification counts. We analyzed 5,140 observations from 514 districts/municipalities in 2016-2025 using separate inferential and predictive approaches. Contemporaneous count regression used a log-population offset and district-clustered robust uncertainty. Among Poisson, generalized Poisson, NB1, and NB2 models fitted to 2016-2024, NB2 had the lowest AIC (66,458.61); Poisson Pearson dispersion was 320.07. For prediction, previous-year regional characteristics were used to estimate the following year's recorded counts, excluding previous TB counts. Random Forest, XGBoost, and LightGBM baselines and PSO-optimized variants were compared with predictive negative binomial regression using five expanding-window folds for outcome years 2020-2024. XGBoost-PSO was selected by the lowest mean weighted absolute percentage error (WAPE; 32.74%). On the temporally separated 2025 outcome year, it achieved WAPE 19.39%, mean absolute error 327.25 cases, root mean squared error 714.36 cases, and mean Poisson deviance 121.56. RF-PSO had slightly lower WAPE (19.32%) and lower deviance (111.58), while XGBoost-PSO minimized RMSE. PSO reduced holdout WAPE relative to each baseline, but a separate persistence reference performed better than all seven candidates (WAPE 7.55%). Monte Carlo SHAP identified

*Corresponding author

E-mail address: sri.winarni@unpad.ac.id

Received September 8, 2026

population (57.55%) and density (15.52%) as the largest global contributions to 2025 predictions. After refitting the model through the 2025 outcome, the projected total for 2026 was 792,909.68 recorded cases, with Bogor District having the highest district-level projection. The 2025 outcomes had been examined in earlier versions, so this is a retrospective temporal evaluation. The 2026 point predictions lack calibrated prediction intervals and require prospective validation. The framework separates ecological rate associations, restricted covariate-based forecasting, and model-specific predictive explanations.

Keywords: tuberculosis; negative binomial regression; random forest; XGBoost; LightGBM; particle swarm optimization; SHAP; temporal validation

2020 AMS Subject Classification: 62P10; 92C60

1. INTRODUCTION

Tuberculosis (TB) control is part of Sustainable Development Goal Target 3.3, which calls for ending the epidemic by 2030 [1]. The World Health Organization estimated 10.7 million incident TB cases globally in 2024, with Indonesia contributing approximately 10% and ranking second after India [2]. In the present study dataset, recorded cases increased from 360,565 in 2016 to 867,480 in 2025. These figures describe diagnosed and reported cases, rather than estimated total incidence. Indonesian studies documented pandemic-related disruptions to TB detection and services [3, 4], making surveillance processes essential to interpreting changes in recorded counts.

National totals conceal variation between districts and municipalities. Arisanti et al. integrated negative binomial mixed models and Random Forest to investigate TB risk factors and forecast cases in West Java [5]. Farkhan et al. analyzed notified cases and deaths across 514 Indonesian districts using Bayesian space-time hierarchical models [6]. These studies motivate regional analysis; our study addresses a complementary question by comparing one-year-ahead count predictions from regional characteristics under a common temporal evaluation design. The dataset comprises 5,140 district-year observations from 514 districts/municipalities during 2016-2025, with nine regional characteristics.

Recorded TB cases are nonnegative counts, and negative binomial regression provides an alternative to Poisson regression when conditional variation exceeds equidispersion [7]. For association analysis, population exposure is incorporated through a log-population offset, and

coefficients are interpreted as incidence rate ratios (IRRs) of recorded notification rates. Accommodating overdispersion does not itself ensure accurate prediction of the conditional mean. In a basic log-linear model, nonlinear terms and interactions must be specified explicitly, whereas tree ensembles offer flexible prediction rules [8, 9, 10]. This motivates testing whether those rules reduce future-year prediction errors while retaining negative binomial regression for interpretable rate associations.

The predictive comparison includes negative binomial regression and Random Forest, XGBoost, and LightGBM before and after particle swarm optimization (PSO), yielding seven candidates [8, 9, 10]. Ensemble performance depends on hyperparameters controlling model complexity and training [11, 12]. PSO searches candidate configurations using individual and swarm experience [13, 14]. In this study, optimization minimizes mean validation weighted absolute percentage error (WAPE). Its benefit is evaluated against each corresponding baseline rather than assumed in advance.

Forecast evaluation must respect the information available when predictions are made and avoid leakage from future observations [15, 16]. We use characteristics in year t to predict recorded cases in year $t+1$, excluding previous TB counts from all fitted candidates. Five expanding-window folds for outcome years 2020-2024 guide model selection, and the 2025 outcomes are excluded from fitting and tuning in the current pipeline. WAPE, mean absolute error, root mean squared error, and Poisson deviance summarize complementary aspects of performance. SHapley Additive exPlanations (SHAP) provide global and local predictor contributions [17, 18]; these explain model predictions and are not treated as causal effects [19].

This study aims to estimate regional notification-rate associations, compare the one-year-ahead accuracy of seven models, and assess the benefit of PSO. The selected model is refitted to produce district-level projections for 2026, accompanied by maps and SHAP explanations. IRRs and SHAP address separate questions: adjusted associations with recorded notification rates and contributions to fitted count predictions. The resulting evidence can support regional surveillance review and consideration of examination and treatment capacity, interpreted alongside local case-finding, reporting, and epidemiological information.

2. MATERIALS AND METHODS

2.1. Study design, data sources, and analytical targets. We assembled a retrospective ecological panel for 2016-2025, using the district/municipality-year as the observational unit. The balanced panel comprised 5,140 observations from 514 districts/municipalities over 10 years. The response was the annual number of recorded TB cases. TB notifications, Bacillus Calmette–Guérin (BCG) immunization coverage, public health centers (puskesmas), general hospitals, and household clean and healthy behavior (PHBS) were obtained from Indonesian Ministry of Health sources, including national and subnational health profiles [20, 21]. Population, population density, adequate sanitation, adequate drinking-water access, and poverty were obtained from official district/municipality statistics published by Statistics Indonesia [22].

Recorded notifications represent the cases captured by surveillance rather than latent biological incidence. Diagnosis, access, public-private engagement, care pathways, and reporting practices influence the number of cases recorded [3, 4, 23, 24]. Consequently, health-service indicators may reflect both population need and case-finding capacity. Regression associations and SHAP explanations were interpreted within this ecological surveillance setting.

Two analytical targets were distinguished. The inferential analysis estimated contemporaneous associations between regional characteristics and notification rates in 2016-2024, using population as an exposure offset. The predictive analysis estimated the following year’s absolute recorded case count from the previous year’s regional characteristics. Population was a predictor in machine-learning models and an offset in predictive count regression; both approaches used information available at the same forecast origin.

Table 1 summarizes the operational definitions and analytical roles. The machine-learning information set contained nine regional characteristics. The count-regression information set contained the same population exposure and eight covariates. Calendar year and previous TB counts were excluded from fitted candidate models.

Calendar year and district identifiers align observations; neither enters the fitted predictor set. Previous TB case counts are excluded from all seven fitted candidates.

2.2. Data quality control and preprocessing. Quality control covered panel completeness, analytical missingness, geographic identifiers, zero counts, and predictor ranges. The final panel

TABLE 1. Study variables and analytical roles.

Variable	Operational definition and unit	Analytical role
Recorded TB cases	Annual recorded count; cases	Response Y_t for inference; Y_{t+1} for prediction
Population	Resident population; persons	Log-exposure offset in count regression; ML predictor
Population density	Residents per land area; persons/km ²	Covariate
BCG coverage	Reported BCG immunization coverage; %	Covariate
Public health centers	Puskesmas per 100,000 residents	Covariate
General hospitals	General hospitals per 100,000 residents	Covariate
Adequate sanitation	Households with adequate sanitation; %	Covariate
Adequate drinking water	Households with adequate drinking-water access; %	Covariate
Household PHBS	Households practicing clean and healthy behavior; %	Covariate
Poverty	Population below the poverty line; %	Covariate

contained 514 districts/municipalities in each of 10 years, with no missing values in analytical columns. Twenty-three zero-count district-years were retained. BCG coverage in the final analytical data ranged from 0% to 100%.

Health-service variables were expressed as numbers of facilities per 100,000 residents rather than raw counts. This scaling distinguishes regional service availability from total population exposure. Population remained on the count scale for the tree ensembles and entered through log population in count regression. Percentage variables retained their percentage-point units, and descriptive summaries used the final analytical scales.

Predictor redundancy was assessed in the inferential development data. The eight count-regression covariates had variance inflation factors (VIFs) between 1.14 and 2.03, with the maximum for adequate sanitation. Population was excluded from this VIF calculation because its coefficient was fixed at one through the offset. Temporal evaluation kept later outcomes separate from training observations and used year- t characteristics to predict year- $t+1$ notifications [15, 16].

2.3. Count-distribution assessment and inferential negative binomial regression. For district i in year t , let Y_{it} denote the recorded TB count, N_{it} the population exposure, and x_{it} the

vector of eight regional covariates. A Poisson model assumes that the conditional mean and variance are equal. All count-distribution candidates shared the same log-link mean specification with a population offset. Equations (2.1) and (2.2) give the Poisson assumption and inferential mean model.

$$(2.1) \quad E(Y_{it} | x_{it}, N_{it}) = \text{Var}(Y_{it} | x_{it}, N_{it}) = \mu_{it}$$

$$(2.2) \quad \log \mu_{it} = \log N_{it} + \beta_0 + \sum_{j=1}^8 \beta_j x_{itj}$$

The coefficient of log population was fixed at one. Consequently, the modeled mean per resident is $\exp(\beta_0 + \sum_j \beta_j x_{itj})$, and exponentiated covariate coefficients describe notification-rate ratios. These rate ratios concern recorded aggregate notifications and do not measure individual disease risk. The empirical variance-to-mean ratio provided an initial, unconditional dispersion assessment, as defined in Equation (2.3).

$$(2.3) \quad DI = \frac{s_Y^2}{\bar{Y}}$$

Poisson, generalized Poisson, NB1, and NB2 distributions were fitted to the same 4,626 contemporaneous observations for 2016-2024. NB1 permits variance proportional to the mean, whereas NB2 permits a quadratic mean-variance relationship [7, 25, 26]. For positive dispersion parameters phi and theta, their variance functions are shown in Equation (2.4). The generalized Poisson model provided an additional flexible count-distribution comparator.

$$(2.4) \quad V_{NB1} = \mu(1 + \phi), \quad V_{NB2} = \mu + \frac{\mu^2}{\theta}$$

Model adequacy was assessed using AIC, BIC, log-likelihood, Pearson dispersion, and numerical convergence. Optimization checks included convergence status, a positive-definite Hessian, and gradient diagnostics. Alternative starts and optimizers were examined where available in the fitting log. The distribution with the lowest AIC among adequately converged candidates was used for the inferential regression.

For an increase of Δ units in covariate j , the adjusted notification-rate ratio was $\exp(\Delta\beta_j)$. Uncertainty used district-clustered HC0 sandwich standard errors, allowing dependence among repeated observations from the same district while assuming independence across districts [27].

The confidence interval transformation is given in Equations (2.5) and (2.6). Clustering changes the standard errors, confidence intervals, and Wald tests, but does not alter fitted coefficients.

$$(2.5) \quad IRR_j(\Delta) = \exp(\Delta\beta_j)$$

$$(2.6) \quad CI_{95\%} = \exp\left\{\Delta\left[\hat{\beta}_j \pm 1.96SE_{cluster}\left(\hat{\beta}_j\right)\right]\right\}$$

Density was interpreted per 100 persons/km², facility availability per one facility per 100,000 residents, and percentage variables per one percentage point. Population itself had no estimated IRR because it was the exposure offset. Calendar year was not included in this specification. Interpretations concern adjusted ecological associations rather than causal effects.

2.4. Next-year prediction and baseline ensemble models. The predictive dataset paired each district's characteristics in year t with its recorded cases in year $t+1$. Thus, predicting the 2025 outcome used regional characteristics measured in 2024. Previous case counts were not used as predictors, and district identifiers served only to align observations and outputs. Equation (2.7) expresses the forecasting target.

$$(2.7) \quad \hat{Y}_{i,t+1} = f(X_{it}), \quad X_{it} = (N_{it}, x_{it1}, \dots, x_{it8})$$

The predictive negative binomial benchmark used the same forecast origin and covariate timing. Its population exposure was N_{it} , not the unobserved future population $N_{i,t+1}$. NB1 and NB2 were compared by training-set AIC within each fit, and the chosen distribution generated count predictions. Equation (2.8) gives the predictive count-regression mean.

$$(2.8) \quad \log \mu_{i,t+1|t} = \log N_{it} + \gamma_0 + \sum_{j=1}^8 \gamma_j x_{itj}$$

Random Forest, XGBoost, and LightGBM were fitted as baseline and PSO-optimized regression ensembles. Random Forest aggregates randomized regression trees, with its predicted count given by Equation (2.9). XGBoost and LightGBM add trees sequentially to a boosted score and used Poisson objectives for the nonnegative count outcome. Their positive mean prediction is obtained by exponentiating the score, as illustrated in Equation (2.10)

[8, 9, 10, 28, 29].

$$(2.9) \quad \hat{Y}_{RF}(X) = \frac{1}{B} \sum_{b=1}^B T_b(X)$$

$$(2.10) \quad F_M(X) = F_0(X) + \sum_{m=1}^M \eta f_m(X), \quad \hat{\mu}(X) = \exp\{F_M(X)\}$$

The ensembles share a target and information set but differ in tree construction and regularization. Random Forest reduces variance through resampling and randomized feature selection. Boosted trees balance fit, learning rate, tree complexity, sampling, and regularization [30, 11, 12]. Table 2 reports the implemented baseline configurations and PSO search domains.

TABLE 2. Baseline configurations and admissible PSO search domains.

Learner	Baseline configuration	PSO search domain
Random Forest	mtry 3; min.node.size 5; sample.fraction 0.632; num.trees 500; max.depth 0	mtry 1-9; min.node.size 1-40; sample.fraction 0.50-1.00; num.trees 300-1,200; max.depth 0-30
XGBoost	Poisson objective; eta 0.10; max_depth 3; min_child_weight 1; gamma 0; subsample 1; colsample_bytree 1; lambda 1; alpha 0; nrounds 100	eta 0.01-0.20; max_depth 2-12; min_child_weight 1-40; gamma 0-10; subsample 0.60-1.00; colsample_bytree 0.50-1.00; lambda 0-30; alpha 0-10; nrounds 100-1,600
LightGBM	Poisson objective; learning_rate 0.10; num_leaves 31; min_data_in_leaf 20; feature_fraction 1; bagging_fraction 1; lambda_l1 0; lambda_l2 0; nrounds 100; max_depth -1; min_gain_to_split 0	learning_rate 0.01-0.20; num_leaves 8-127; min_data_in_leaf 5-100; feature_fraction 0.50-1.00; bagging_fraction 0.50-1.00; lambda_l1 0-10; lambda_l2 0-30; nrounds 100-1,600; depth search encoding 0-12; min_gain_to_split 0-5

Learning rates were searched on the log10 scale. Unrestricted depth is coded as 0 for Random Forest and -1 in the fitted LightGBM configuration.

Discrete search dimensions were mapped to valid integer settings. Depth encodings were translated to the learner’s admissible settings. The Random Forest value zero denotes unrestricted depth, whereas the LightGBM fitted value minus one denotes unrestricted depth. Every baseline lay within the admissible domain and was explicitly included as a seeded particle, making the baseline-to-PSO comparison transparent.

2.5. Particle swarm optimization. Each PSO particle represented a candidate hyperparameter vector within the prespecified bounds. Positions were updated using the particle's previous velocity, personal best, and swarm best, as shown in Equations (2.11) and (2.12). The cognitive and social coefficients control attraction to individual and collective optima, respectively. Independent uniform random terms support exploration of the search space [12, 13]. The inertia weight decreased linearly from 0.9 at iteration 1 to 0.4 at iteration 20, with cognitive and social coefficients $c_1 = c_2 = 1.5$. Particle positions were represented on the normalized interval $[0, 1]$; velocities were clipped to $[-0.2, 0.2]$, and velocity components crossing a boundary were set to zero before positions were bounded.

$$(2.11) \quad v_{kd}^{(r+1)} = \omega v_{kd}^{(r)} + c_1 u_1 [p_{kd}^{best} - z_{kd}^{(r)}] + c_2 u_2 [g_d^{best} - z_{kd}^{(r)}]$$

$$(2.12) \quad z_{kd}^{(r+1)} = z_{kd}^{(r)} + v_{kd}^{(r+1)}$$

The saved search histories contained 12 particles, an initialization at iteration zero, and 20 subsequent iterations per restart. Two restarts were performed for each learner, giving 252 recorded candidate evaluations per restart and 504 per learner. The baseline was seeded in both restarts. The best configuration across the two restarts was retained for each algorithm. The base random seed was 14072025. Restart r used seed $14072025 + 100m + 10000r$, where $m = 1, 2$, or 3 for Random Forest, XGBoost, or LightGBM, respectively, and $r = 1$ or 2 . Checkpoint resumption restored the saved random-number-generator state. The current run restored checkpoint states; its model-lock record identified XGBoost-PSO before the 2025 evaluation step and recorded that the 2025 outcomes were not used in the implemented selection criterion.

PSO minimized the arithmetic mean of validation WAPE across the five outcome years 2020-2024. This gives each validation year equal weight rather than pooling errors across years with different national totals. The optimization objective is defined in Equation (2.13). The 2025 outcomes did not enter the search objective or primary model designation.

$$(2.13) \quad J(\vartheta) = \frac{1}{5} \sum_{h=2020}^{2024} WAPE_h(\vartheta)$$

Search benefits were assessed separately for each learner. Mean validation WAPE for the retained PSO configuration was compared with that learner's seeded baseline, and the same

comparison was repeated on the held-out 2025 year. Candidate histories and convergence curves document improvement within the implemented search budget. They do not establish that PSO located a global optimum or that it outperforms other search algorithms [12, 13].

2.6. Temporal validation and performance measures. Five expanding-window folds preserved chronological ordering. The first fold trained on predictor years 2016-2018 paired with outcome years 2017-2019, then used 2019 characteristics to predict 2020. Later folds expanded the training pairs by one year until the final development fold predicted 2024 from 2023 characteristics. Each validation year contained all 514 districts/municipalities, and Table 3 reports the exact predictor-outcome alignment [15, 16].

TABLE 3. Predictor-year and outcome-year alignment for development testing and forecasting.

Stage	Training X years	Training Y years	Forecast X year	Outcome year	Training pairs
Fold 1	2016-2018	2017-2019	2019	2020	1,542
Fold 2	2016-2019	2017-2020	2020	2021	2,056
Fold 3	2016-2020	2017-2021	2021	2022	2,570
Fold 4	2016-2021	2017-2022	2022	2023	3,084
Fold 5	2016-2022	2017-2023	2023	2024	3,598
2025 test	2016-2023	2017-2024	2024	2025	4,112
2026 projection	2016-2024	2017-2025	2025	2026	4,626

Each validation or test year contains 514 districts/municipalities. Inferential distribution assessment uses a separate contemporaneous 2016-2024 panel of 4,626 observations.

The primary model was designated using the lowest mean validation WAPE across the seven candidates. MAE, RMSE, and mean Poisson deviance were secondary measures describing other aspects of error. This rule selected XGBoost-PSO even though Random Forest-PSO performed better on several secondary development metrics. Scores on the 2025 outcome year were descriptive benchmarks and did not change the primary designation.

For the retrospective 2025 temporal assessment, models were fitted on 4,112 training pairs with predictor years 2016-2023 and outcome years 2017-2024. The test inputs were characteristics from 2024, and the test outcomes were the 514 recorded counts for 2025. For observed counts y_i and positive predicted means $\hat{\mu}_i$, WAPE, MAE, RMSE, and mean Poisson deviance

were calculated using Equations (2.14)-(2.17). Bias was the mean predicted-minus-observed count in Equation (2.18).

$$(2.14) \quad WAPE = 100 \frac{\sum_{i=1}^n |y_i - \hat{\mu}_i|}{\sum_{i=1}^n y_i}$$

$$(2.15) \quad MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{\mu}_i|$$

$$(2.16) \quad RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{\mu}_i)^2}$$

$$(2.17) \quad D_{Pois} = \frac{2}{n} \sum_{i=1}^n \left[y_i \log \left(\frac{y_i}{\hat{\mu}_i} \right) - (y_i - \hat{\mu}_i) \right]$$

$$(2.18) \quad Bias = \frac{1}{n} \sum_{i=1}^n (\hat{\mu}_i - y_i)$$

The logarithmic term in Poisson deviance was defined as zero for $y_i = 0$. Lower values of the four losses indicate smaller prediction errors, whereas negative bias indicates systematic underprediction. A separate persistence reference predicted each district's 2025 count by its recorded 2024 count. Persistence was not one of the seven fitted models and did not supply a previous-case-count feature to any candidate; it provided a practical external reference for evaluating the restricted covariate-based forecasting task.

2.7. SHAP analysis and projections for 2026. The selected XGBoost-PSO model was explained using Monte Carlo SHAP estimates produced with fastshap [17, 31]. The prediction wrapper returned counts, so contributions were expressed in predicted cases. Local accuracy was checked by reconstructing each exported prediction from its baseline and summed contributions. The decomposition and normalized global importance are defined in Equations (2.19) and (2.20). The script specified 200 Monte Carlo repetitions ($nsim = 200$), a random sample of up to 1,000 training observations as the background, $adjust = TRUE$, and $parallel = FALSE$. The background mean predicted count supplied the baseline. Seeds were 14074050 for the 2025 explanations and 14074051 for the 2026 explanations.

$$(2.19) \quad f(X_i) \approx \phi_0 + \sum_{j=1}^9 \phi_{ij}$$

$$(2.20) \quad I_j = \frac{1}{n} \sum_{i=1}^n |\phi_{ij}|, \quad RI_j = 100 \frac{I_j}{\sum_{k=1}^9 I_k}$$

All nine regional predictors entered the global ranking and local decomposition. Positive SHAP values raise a prediction above its baseline, and negative values lower it. Relative importance quantifies a share of total mean absolute model contribution; it is not a share of cases caused by a predictor. SHAP depends on the fitted model, predictor correlations, and background distribution, so it does not provide adjusted regression effects or causal identification [17, 19].

After the 2025 temporal evaluation, the chosen hyperparameters were retained and the forecasting model was refitted using all 4,626 available pairs: predictor years 2016-2024 and outcome years 2017-2025. Regional characteristics from 2025 then generated predictions for 2026. Forecast summaries used the unrounded predictions, with rounding only for district-level presentation. New SHAP outputs explained this refitted model and were summarized separately from the 2025 holdout explanations.

Predictions were matched to geoBoundaries district/municipality polygons using harmonized administrative names. The matching audit recorded 514 predictions matched one-to-one with 514 administrative polygons, without unmatched predictions or polygons. Six explicit naming aliases and five excluded nonadministrative objects were recorded in the map audit. The map displays predicted recorded case counts rather than notification rates, individual risk, or spatially modeled effects.

2.8. Software and reproducibility. The analysis used R 4.5.3 with glmmTMB 1.1.14 for count regression, sandwich 3.1.1 for robust uncertainty, ranger 0.18.0 for Random Forest, xgboost 1.7.11.1 for XGBoost, lightgbm 4.7.0 for LightGBM, and fastshap 0.1.1 for SHAP estimation. Data import and plotting used readxl 1.5.0 and ggplot2 4.0.3. The saved outputs included distribution comparisons, convergence diagnostics, temporal folds, candidate histories, optimized configurations, district-level predictions, and SHAP matrices. Reported performance tables were checked against the archived district-level predictions, and figures summarize the corresponding numerical outputs. Training in temporal fold k used seed $14072025 + k$; the 2025 evaluation and final forecast models used seed 14073025. All learners were configured

for one computational thread. The saved run manifest confirms these settings, and the saved SHAP objects confirm 200 repetitions and 1,000 background observations for each prediction year.

3. RESULTS AND DISCUSSION

3.1. Panel characteristics and count-distribution adequacy. The final balanced panel contained 514 district/municipality observations per year and 5,140 observations overall. Recorded cases had a mean of 1,177.09, a median of 515.5, a standard deviation of 2,121.07, and a range of 0-29,067. The sample variance was 4,498,934.57, yielding an unconditional variance-to-mean ratio of 3,822.10. Twenty-three verified zero-count district-years were retained, and the analytical columns contained no missing observations.

Table 4 summarizes the observed variables and their units.

TABLE 4. Descriptive statistics on the final analytical scales for 2016-2025.

Variable	Mean	SD	Median	Range
Recorded TB cases (cases)	1,177.09	2,121.07	515.50	0.00-29,067.00
Population (persons)	528,823.56	637,178.34	284,143.50	13,699.00-5,965,410.00
Population density (persons/km ²)	1,312.67	4,127.11	147.57	0.92-79,677.77
BCG coverage (%)	82.43	18.78	88.09	0.00-100.00
Public health centers (per 100,000)	7.06	6.21	5.21	0.27-77.95
General hospitals (per 100,000)	1.15	0.95	0.86	0.00-7.59
Adequate sanitation (%)	73.27	20.44	78.41	0.00-100.00
Adequate drinking water (%)	76.78	21.59	82.95	0.00-100.00
Household PHBS (%)	60.19	18.95	64.00	0.00-100.00
Poverty (%)	11.96	7.56	10.06	1.67-45.11

The response was strongly right-skewed, with the mean substantially above the median. Population and density also varied markedly across districts. General-hospital availability ranged from zero to 7.59 facilities per 100,000 residents, while public health-center availability ranged from 0.27 to 77.95 per 100,000. These heterogeneous exposures and service profiles motivate a population-offset regression for rate associations and flexible models for absolute-count prediction.

Annual recorded TB cases increased from 360,565 in 2016 to 867,480 in 2025. Counts fell from 568,987 in 2019 to 393,569 in 2020, then rose to 443,343 in 2021 and 724,416 in 2022. The decline and recovery coincide with documented pandemic-related changes in detection and reporting [3, 4]. Figure 1 displays this temporal pattern, which makes next-year validation more informative than randomly mixed evaluation.

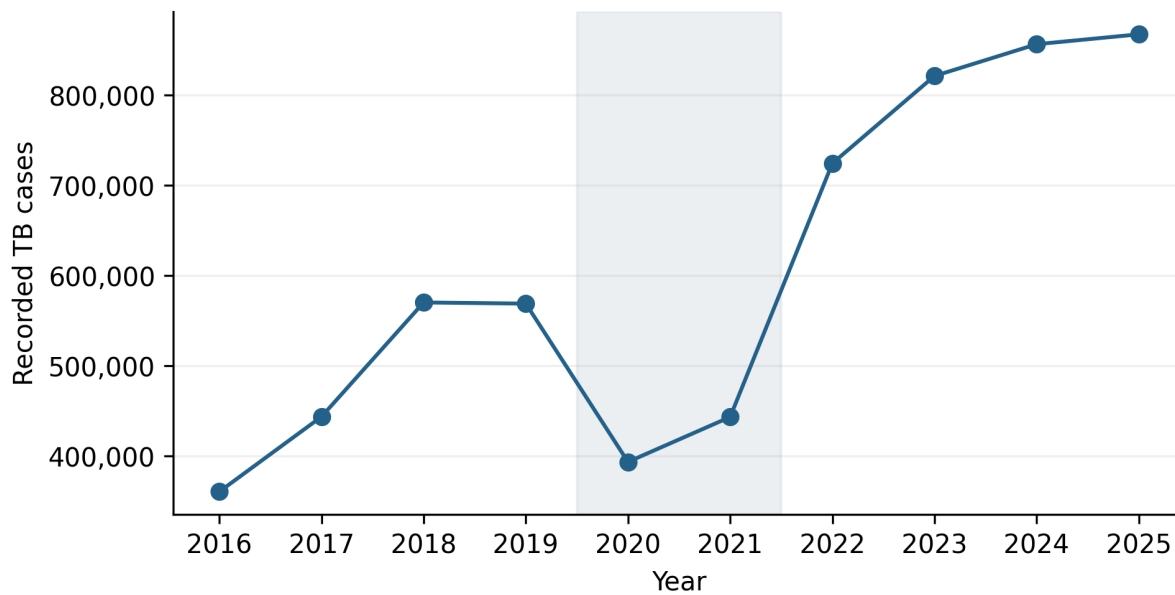


FIGURE 1. Annual recorded TB cases across 514 districts and municipalities in Indonesia from 2016 to 2025. Shading marks the 2020-2021 pandemic period.

The population-offset Poisson model remained severely overdispersed after covariate adjustment. Its Pearson statistic was 1,477,750.03 on 4,617 residual degrees of freedom, giving dispersion 320.07. Its AIC was 1,338,157.34, compared with 66,458.61 for NB2. Table 5 compares all four contemporaneous count-distribution candidates.

All candidates use the same eight covariates and log-population offset ($n=4,626$). All retained fits have status OK. Dispersion parameters have family-specific meanings and are not directly comparable across families.

NB2 had the lowest AIC and BIC, with AIC differences of 533.02 relative to generalized Poisson and 877.90 relative to NB1. Its estimated theta was 2.7290, equivalent to a quadratic overdispersion coefficient $1/\theta$ of approximately 0.3664. NB2 Pearson dispersion was 1.27, indicating a large reduction in extra-Poisson variation without establishing perfect model fit.

TABLE 5. Count-distribution comparison on contemporaneous development data from 2016-2024.

Distribution	AIC	BIC	Log-likelihood	Pearson dispersion	Delta AIC
NB2	66,458.61	66,523.00	-33,219.30	1.27	0.00
Generalized Poisson	66,991.63	67,056.02	-33,485.81	0.81	533.02
NB1	67,336.51	67,400.90	-33,658.25	1.09	877.90
Poisson	1,338,157.34	1,338,215.30	-669,069.67	320.07	1,271,698.74

The recorded convergence checks were satisfactory, including a positive-definite Hessian and a maximum scaled gradient of 0.000096 for the selected inferential NB2 fit.

This comparison supports NB2 for contemporaneous association analysis. The marked AIC advantage over Poisson reflects the need for a flexible variance structure [7, 25]. Information criteria compare fitted likelihoods on the same data and do not demonstrate superior future-year point prediction. The predictive count benchmark was therefore evaluated separately under the same chronological task as the ensembles.

3.2. Ecological notification-rate associations. The inferential NB2 regression included eight covariates and a log-population offset. District-clustered robust uncertainty was used because each district contributed repeated annual observations. Population exposure was accounted for without estimating a separate population coefficient. Table 6 reports the resulting IRRs, cluster-robust 95% confidence intervals, and Wald p values.

TABLE 6. Inferential NB2 notification-rate ratios with district-clustered robust confidence intervals.

Covariate	Contrast	IRR	95% CI	p value
Population density	per 100 persons/km ²	1.0033	1.0020-1.0045	<0.001
BCG coverage	per 1 percentage point	1.0038	1.0012-1.0064	0.005
Public health centers	per 1 facility/100,000	1.0129	0.9991-1.0268	0.067
General hospitals	per 1 facility/100,000	1.2355	1.1679-1.3070	<0.001
Adequate sanitation	per 1 percentage point	0.9995	0.9973-1.0016	0.635
Adequate drinking water	per 1 percentage point	1.0056	1.0041-1.0070	<0.001
Household PHBS	per 1 percentage point	1.0022	1.0005-1.0039	0.010
Poverty	per 1 percentage point	0.9949	0.9848-1.0051	0.322

HC0 sandwich standard errors are clustered by district/municipality (514 clusters). Population is an offset, and the intercept is not presented as a covariate IRR.

Population density was positively associated with recorded notification rates: an increase of 100 persons/km² corresponded to an IRR of 1.0033 (95% CI 1.0020-1.0045). One additional general hospital per 100,000 residents corresponded to an IRR of 1.2355 (95% CI 1.1679-1.3070), the largest multiplicative contrast among the reported one-unit increments. BCG coverage, adequate drinking-water access, and household PHBS also had positive adjusted associations. Public health-center availability ($p=0.067$), adequate sanitation ($p=0.635$), and poverty ($p=0.322$) had confidence intervals spanning one.

Health-service associations should be interpreted in relation to surveillance. A positive hospital-availability coefficient may reflect greater access to diagnosis, referral, screening, and reporting, as well as unmeasured regional differences [23, 24]. It does not imply that expanding hospital capacity increases transmission. The result is consistent with the need to distinguish recorded notification patterns from latent disease occurrence [6, 32].

The positive BCG association is likewise ecological rather than an estimate of vaccine effectiveness. Individual-participant evidence supports protective effects of infant BCG vaccination, particularly in childhood [33]. Regional coverage may also reflect program reach, population composition, and reporting processes. Aggregate notification-rate coefficients cannot resolve these mechanisms or replace individual-level vaccine-effect analyses.

Population density has been associated with TB in ecological studies [34, 25], and housing conditions have also been examined in relation to pulmonary TB [35]. In our dataset, drinking-water access and PHBS may additionally reflect regional development or surveillance capacity; this interpretation is a hypothesis rather than a tested mechanism. The modest effect sizes for percentage variables should be interpreted in their one-percentage-point units. Poverty's confidence interval included the null in the revised offset model, so an independently protective or harmful association was not established. These findings characterize adjusted aggregate associations and do not justify causal claims.

3.3. Temporal validation and PSO-selected configurations. Table 7 reports mean predictive performance across the five expanding-window folds. XGBoost-PSO achieved the lowest mean WAPE, 32.74%, followed by RF-PSO at 33.35% and LightGBM-PSO at 33.47%. XGBoost-PSO was therefore designated as the primary model under the stated selection rule. The difference between the two leading mean WAPEs was only 0.61 percentage points, so the ranking should not be interpreted as decisive statistical superiority.

TABLE 7. Mean predictive errors across expanding-window outcome years 2020-2024.

Model	Mean WAPE (%)	SD WAPE (pp)	Mean MAE (cases)	Mean RMSE (cases)	Mean deviance
XGB-PSO	32.74	8.62	411.94	949.06	247.88
RF-PSO	33.35	10.69	403.58	875.35	215.38
LGB-PSO	33.47	10.15	418.13	960.46	268.29
LightGBM	35.19	8.58	437.64	977.87	272.48
Random Forest	36.88	9.23	449.49	1,018.08	270.74
XGBoost	37.94	6.77	470.38	1,047.16	311.29
Negative binomial	42.64	3.47	539.54	1,242.92	437.53

The primary selection criterion is mean WAPE. MAE and RMSE are in cases. WAPE SD describes between-fold variation, not uncertainty in a model-comparison estimate.

RF-PSO achieved the lowest mean MAE (403.58), RMSE (875.35), and Poisson deviance (215.38). XGBoost-PSO had corresponding values of 411.94, 949.06, and 247.88. Thus, model rankings depended on the error criterion, and the primary designation followed WAPE rather than agreement across all metrics. Figure 2 summarizes the implemented selection criterion for all seven candidates.

The predictive negative binomial benchmark had mean WAPE 42.64%, exceeding all six ensemble variants. Training-set distribution selection used NB1 for the 2020 fold and NB2 for the remaining folds. This fold-specific selection differs from the contemporaneous inferential NB2 fit and maintains the available training data at each forecast origin. The results support the ensembles for the specified covariate-based forecasting task while retaining negative binomial regression for interpretable association analysis.

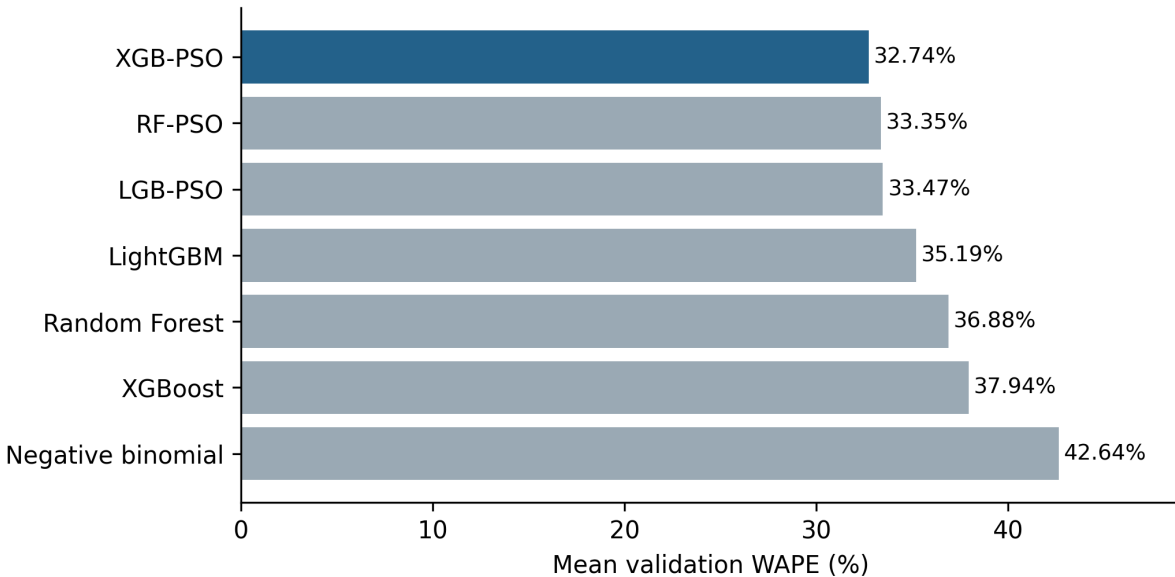


FIGURE 2. Mean validation WAPE across the five outcome years 2020-2024.

XGBoost-PSO was selected by the lowest mean WAPE.

Development performance varied substantially by year. For XGBoost-PSO, WAPE was 39.25% in 2020, 22.55% in 2021, 43.32% in 2022, 32.06% in 2023, and 26.51% in 2024. The large error in 2022 coincided with the substantial rebound in national notifications. Predictive relationships based on regional characteristics alone did not fully capture abrupt changes in case finding and reporting [3, 4, 15, 16].

All three learners improved their mean validation WAPE after PSO. Relative reductions were 9.57% for Random Forest, 13.71% for XGBoost, and 4.89% for LightGBM. These comparisons used the same folds and included the baselines as seeded candidates. Figure 3 shows the saved best-so-far WAPE trajectories for both restarts of each learner.

Table 8 reports the selected hyperparameter configurations.

RF-PSO selected seven candidate predictors per split, a minimum node size of one, 388 trees, and a sampling fraction of 0.9938. XGBoost-PSO selected a learning rate of 0.02990, depth 11, 786 boosting rounds, and both L1 and L2 regularization. LightGBM-PSO selected 31 leaves, 854 rounds, a learning rate of 0.11632, and unrestricted maximum depth. These data-specific configurations were retained before testing on the 2025 outcome year.

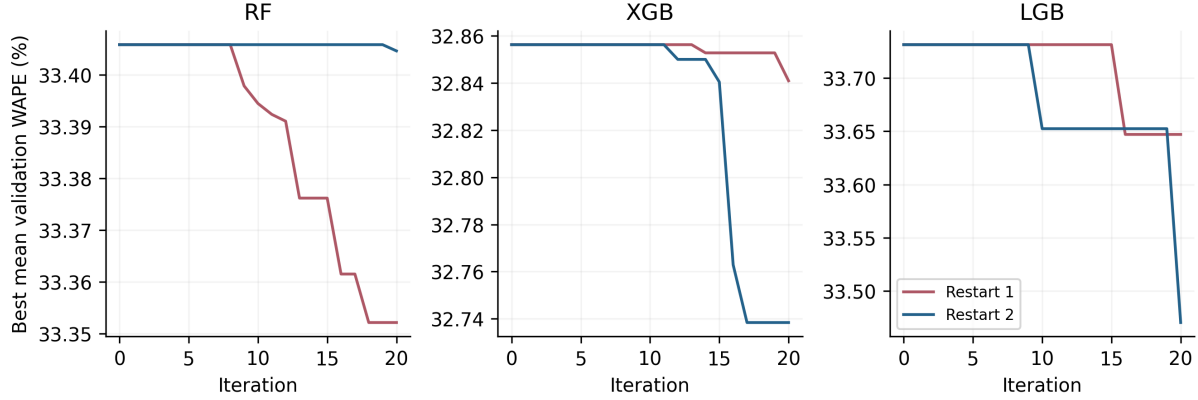


FIGURE 3. PSO convergence histories for Random Forest, XGBoost, and LightGBM. Each learner used two restarts with 12 particles and 20 iterations following initialization.

TABLE 8. Configurations retained after PSO under mean validation WAPE.

Model	Selected configuration
RF-PSO	mtry=7; min.node.size=1; sample.fraction=0.99380; num.trees=388; max.depth=0
XGB-PSO	max_depth=11; min_child_weight=19.43881; gamma=0; subsample=0.89065; colsample_bytree=1; lambda=8.64697; alpha=2.94168; nrounds=786; eta=0.02990
LGB-PSO	num_leaves=31; min_data_in_leaf=20; feature_fraction=0.95532; bagging_fraction=0.86426; lambda_l1=0.34120; lambda_l2=3.45142; nrounds=854; max_depth=-1; min_gain_to_split=0; learning_rate=0.11632

The validation results also show that tuning does not improve every year uniformly. XGBoost-PSO had higher WAPE than baseline XGBoost in 2022, despite reducing the mean across all five folds. The selected parameters represent the best recorded configuration under the implemented WAPE objective and budget. They are not generally optimal defaults, and no comparison with random search or Bayesian optimization was performed [11, 12, 13].

3.4. Retrospective temporal evaluation on the 2025 outcome year. The 2025 test used 2024 characteristics for all 514 districts/municipalities. XGBoost-PSO was fitted only on predictor-outcome pairs ending with the 2024 outcome, and its designation was unchanged after test evaluation. It achieved WAPE 19.39%, MAE 327.25 cases, RMSE 714.36 cases, and mean Poisson deviance 121.56. Table 9 compares the seven fitted candidates on the same test outcomes.

TABLE 9. Retrospective 2025 temporal-test errors using 2024 regional characteristics for 514 districts/municipalities.

Model	WAPE (%)	MAE (cases)	RMSE (cases)	Mean deviance	Bias (cases)
RF-PSO	19.32	326.01	725.23	111.58	-178.58
XGB-PSO (primary)	19.39	327.25	714.36	121.56	-228.93
LGB-PSO	21.69	366.05	855.04	137.71	-220.84
LightGBM	23.19	391.43	894.56	162.51	-233.96
Random Forest	24.21	408.55	992.52	168.02	-244.34
XGBoost	30.37	512.48	1,169.61	270.30	-323.53
Negative binomial	35.35	596.62	1,486.02	418.95	-432.65

The primary designation comes from development-fold WAPE. MAE, RMSE, and bias are in cases. Negative bias denotes underprediction.

RF-PSO had slightly lower WAPE (19.32%) and MAE (326.01 cases), as well as lower Poisson deviance (111.58). XGBoost-PSO had the lowest RMSE (714.36 cases), compared with 725.23 for RF-PSO. The WAPE difference was only 0.07 percentage points. These test-year differences describe metric-specific trade-offs and do not retrospectively replace the development-selected model.

All seven candidates underpredicted on average. XGBoost-PSO's bias was minus 228.93 cases per district, and its predicted national total was 749,812.01 compared with 867,480 observed cases. The shortfall was 117,667.99 cases, or 13.56% of the observed national total. Figure 4 combines the WAPE comparison with an observed-versus-predicted display for the primary model.

The largest absolute XGBoost-PSO errors occurred in high-burden districts in West Java. Cianjur had 12,410 observed cases versus 7,066.70 predicted, while Sukabumi had 12,845 observed versus 7,687.67 predicted. Bogor District had 28,595 observed versus 23,969.11 predicted. These underpredictions show that a favorable average ranking can coexist with substantial errors in districts where surveillance burden is greatest.

The predictive negative binomial model selected NB2 on the 4,112 training pairs. Its test WAPE was 35.35%, MAE 596.62, RMSE 1,486.02, and mean Poisson deviance 418.95. Its

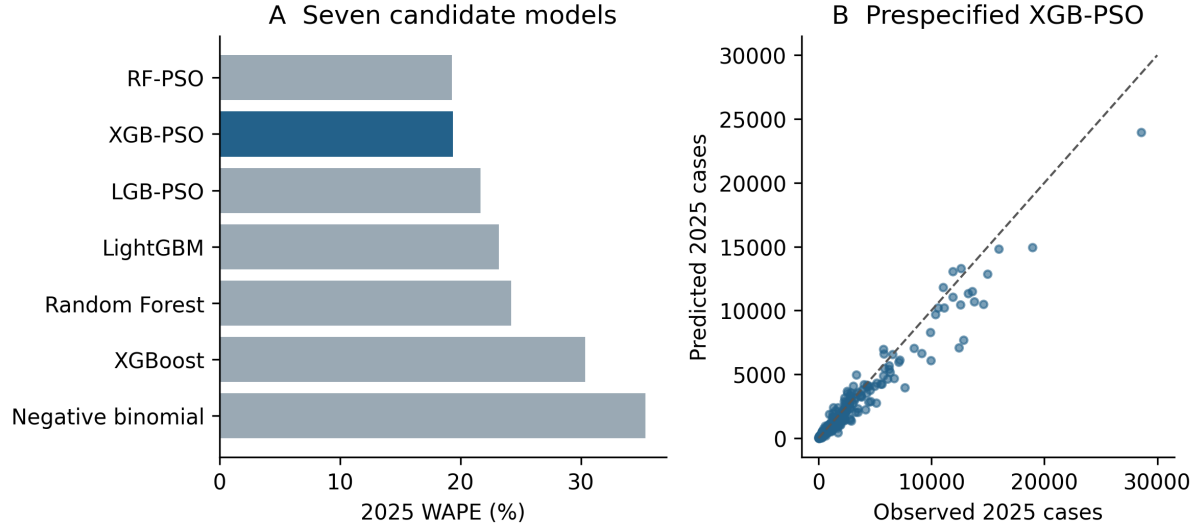


FIGURE 4. Retrospective 2025 temporal evaluation. Panel A compares WAPE for the seven candidates; Panel B plots XGBoost-PSO predictions against observed counts, with the dashed line indicating equality.

lower predictive accuracy does not invalidate contemporaneous IRR estimates, because association estimation and next-year prediction address different targets. The evaluation nevertheless quantifies the limitations of this log-linear mean specification for the restricted forecasting task.

3.5. Optimization gains and the persistence reference. PSO reduced 2025 WAPE relative to each learner’s baseline. The relative reduction was 20.20% for Random Forest, 36.14% for XGBoost, and 6.48% for LightGBM. XGBoost had the largest proportional improvement, from 30.37% to 19.39%. Table 10 separates development-fold improvements from 2025 temporal improvements.

TABLE 10. Baseline-to-PSO WAPE reductions in development and 2025 temporal testing.

Learner	CV baseline WAPE (%)	CV PSO WAPE (%)	CV reduction (%)	2025 baseline WAPE (%)	2025 PSO WAPE (%)	2025 reduction (%)
RF	36.88	33.35	9.57	24.21	19.32	20.20
XGB	37.94	32.74	13.71	30.37	19.39	36.14
LGB	35.19	33.47	4.89	23.19	21.69	6.48

All WAPE columns are percentages. Relative reduction is $100 \times (\text{baseline WAPE} - \text{PSO WAPE})/\text{baseline WAPE}$.

The separate persistence reference performed better than every fitted candidate on the 2025 outcome year. Its WAPE was 7.55%, MAE 127.42 cases, RMSE 222.19 cases, and mean Poisson deviance 29.43. Its bias was minus 21.28 cases per district. Thus, the ensemble results demonstrate improvements over covariate-based baselines, but not superiority over simply carrying forward each district's preceding recorded count.

This distinction matters because the fitted candidates deliberately excluded previous TB counts. Persistence has direct information about recent district-level notification burden, whereas the candidates estimate it through regional characteristics. The favorable persistence result coincided with a relatively small 1.28% increase in the national total from 2024 to 2025. Additional future-year evaluation is required to determine whether persistence remains advantageous during larger surveillance changes [15, 16].

The primary model can therefore support an interpretable assessment of predictions based on regional characteristics, but operational forecasting should consider the stronger persistence reference. The exclusion of previous counts is a substantive constraint of the present model comparison. It narrows the interpretation of predictive performance and helps explain why larger errors remain despite PSO gains. A future study could evaluate autoregressive and combined models under the same temporal design without treating the current results as evidence for their performance.

3.6. SHAP explanations of the selected model. Population dominated the global SHAP ranking for 2025 predictions, accounting for 57.55% of total mean absolute contribution. Density ranked second at 15.52%, followed by general hospitals per 100,000 residents at 5.31%. Together, these three features accounted for 78.37% of normalized global importance. Table 11 compares the holdout ranking with the explanations of the refitted 2026 forecasting model.

Mean absolute SHAP is expressed in predicted cases. Shares are normalized within each year and are not causal attributable fractions.

Figure 5 compares the global SHAP importance rankings for 2025 and 2026.

TABLE 11. Global SHAP importance for the selected XGBoost-PSO models.

Predictor	2025 mean SHAP (cases)	2025 share (%)	2026 mean SHAP (cases)	2026 share (%)
Population	1,005.63	57.55	959.89	56.62
Population density	271.16	15.52	284.48	16.78
General hospitals	92.74	5.31	98.32	5.80
Public health centers	81.10	4.64	73.82	4.35
Adequate drinking water	76.52	4.38	73.57	4.34
Poverty	69.60	3.98	72.72	4.29
BCG coverage	61.71	3.53	36.79	2.17
Household PHBS	51.14	2.93	66.10	3.90
Adequate sanitation	37.83	2.16	29.51	1.74

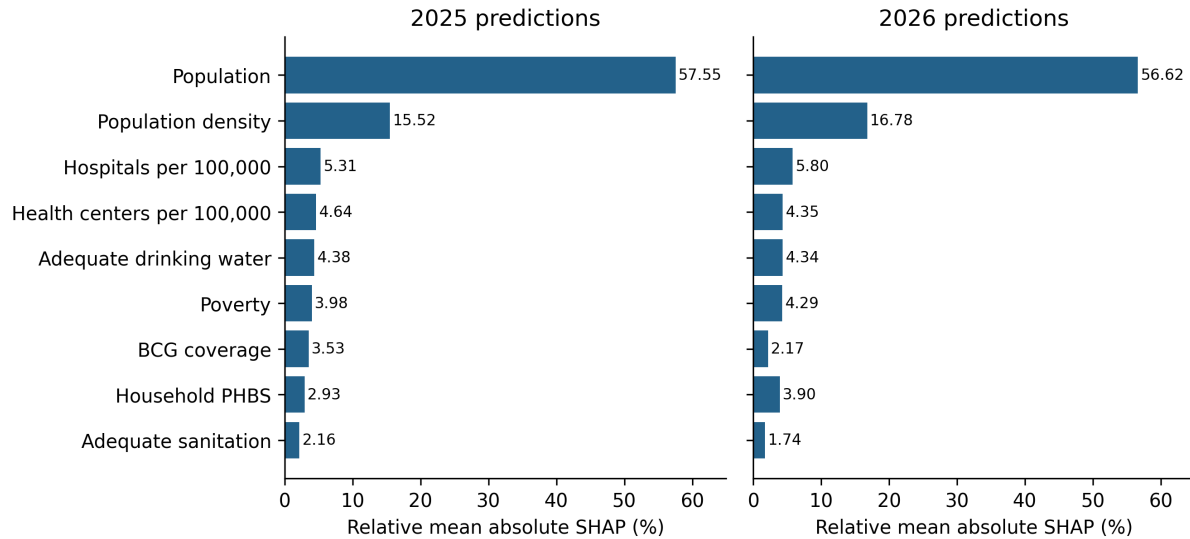


FIGURE 5. Relative mean absolute SHAP importance for XGBoost-PSO predictions in 2025 and 2026. Each ranking includes all nine regional predictors and is calculated separately for its fitted model and explained observations.

Population's dominance is consistent with predicting absolute notification counts. Larger populations can generate larger recorded totals even when notification rates are similar. Density and service availability provide additional information about regional profiles and surveillance capacity [6, 32, 34, 35, 23, 24]. Their SHAP importance describes the fitted model's use of information rather than a causal hierarchy of epidemiologic determinants.

Figure 6 shows the distribution of local contributions across the 514 test predictions. Population contributions have the widest spread, with large positive contributions for some districts

and negative contributions for many smaller predictions. Density contributions also span positive and negative values. Variation within each predictor's contribution distribution reflects its interaction with the remaining regional profile and the model's fitted structure.

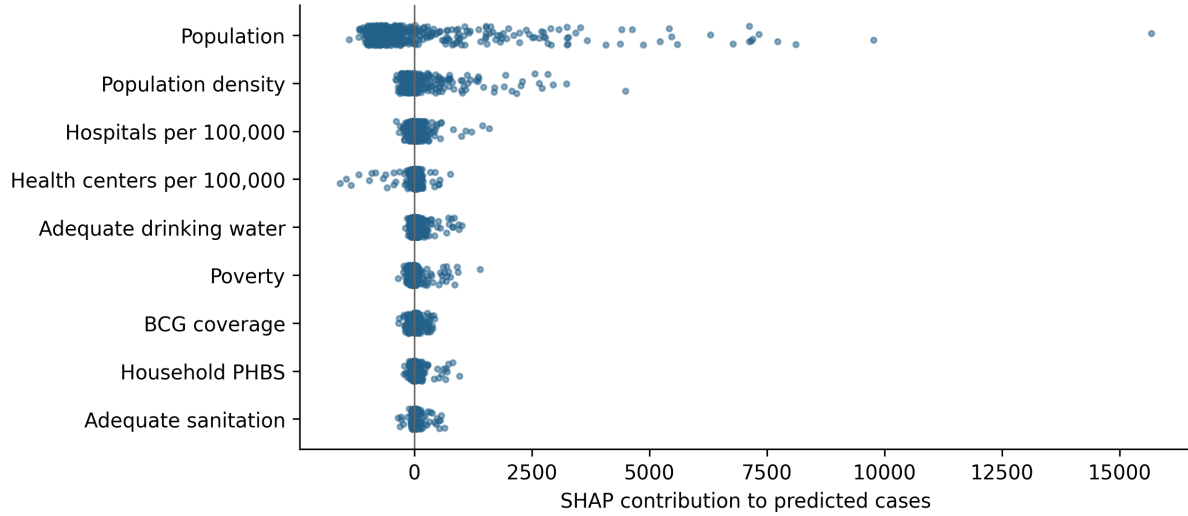


FIGURE 6. Local SHAP contribution distributions for the 514 temporally separated 2025 predictions. Horizontal positions are contributions in predicted cases; vertical jitter is used only to distinguish overlapping observations.

The 2025 explanation baseline was 1,221.44 predicted cases. The maximum absolute discrepancy between baseline plus summed SHAP values and the exported prediction was 0.000407 cases across all 514 observations. For the refitted 2026 model, the baseline was 1,136.00 and the maximum discrepancy was 0.000532 cases. These small reconstruction differences verify numerical local accuracy, but do not establish causal validity or eliminate uncertainty from Monte Carlo SHAP estimation [17, 19, 31].

3.7. District-level projections for 2026. The primary XGBoost-PSO model was refitted after the 2025 temporal evaluation using predictor-outcome pairs through the 2025 outcome. Characteristics measured in 2025 produced a total projected recorded burden of 792,909.68 cases for 2026. This was 8.60% below the observed 2025 total of 867,480. The projection is a model output rather than evidence that transmission or national notifications will decline by this amount.

The 2025 holdout had substantial aggregate underprediction, which limits confidence in the apparent decrease. The 2026 total must therefore be interpreted alongside the model's test-year

bias and the stronger persistence reference. No observed 2026 outcomes or forecast intervals were available in the reported evaluation. Table 12 summarizes projections from all seven models using the same forecast inputs.

TABLE 12. National projections of 2026 recorded TB cases across the seven fitted models.

Model	Projected 2026 total (cases)	Change from observed 2025 (%)
Negative binomial	690,187.47	-20.44
Random Forest	789,900.55	-8.94
RF-PSO	818,471.99	-5.65
XGBoost	746,802.11	-13.91
XGB-PSO (primary)	792,909.68	-8.60
LightGBM	782,663.83	-9.78
LGB-PSO	787,673.39	-9.20

The observed 2025 total is 867,480. Changes compare predicted 2026 totals with this observed reference and are not validated estimates of epidemiologic change.

Predicted burden varied substantially across districts. West Java had the largest projected provincial total, 204,487.01 cases, followed by East Java at 93,784.22 and Central Java at 82,019.21. The three provinces together accounted for 47.96% of the model's national projection. Figure 7 presents the district-level predicted absolute counts using the audited one-to-one polygon match.

Table 13 lists the ten districts/municipalities with the highest projected counts and their leading SHAP contributions.

Population is the largest absolute contributor in all ten districts. Contributions in parentheses are signed predicted cases; ranks use absolute SHAP magnitude. Counts are rounded only for display.

Bogor District had the highest projected count, 28,329.73, followed by Bandung City at 18,361.25 and Medan City at 15,763.43. Population was the largest absolute SHAP contributor in every district in the top ten. Density was the second-ranked contributor in nine of these districts, with poverty second in Bandung District. The third-ranked contribution varied across regional profiles, illustrating why a national global ranking cannot substitute for local explanations.

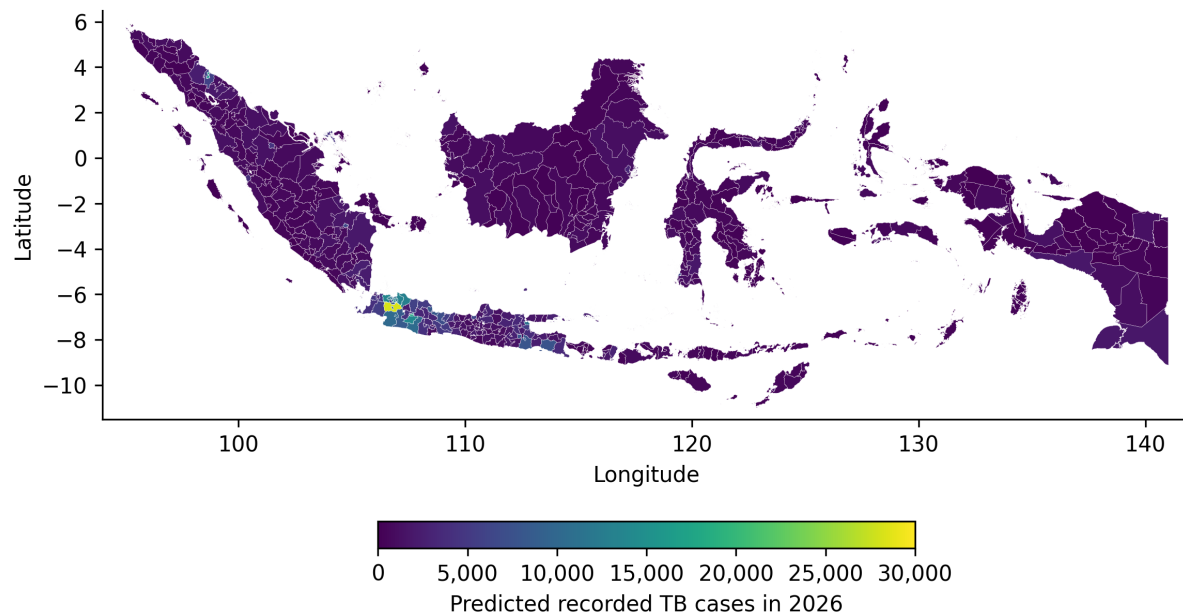


FIGURE 7. XGBoost-PSO projections of recorded TB case counts for 2026 across 514 Indonesian districts and municipalities. The map uses absolute predicted counts and does not represent notification rates or individual risk.

TABLE 13. The ten highest 2026 district projections and selected local SHAP contributions.

District or municipality	Projected cases (cases)	Second-ranked SHAP (cases)	Third-ranked SHAP (cases)
1. Bogor District	28,330	Population density (+5,375)	Poverty (+1,633)
2. Bandung City	18,361	Population density (+3,348)	Poverty (+1,273)
3. Medan City	15,763	Population density (+2,894)	Household PHBS (+1,705)
4. Bandung District	14,039	Poverty (+1,118)	Household PHBS (+571)
5. East Jakarta	13,994	Population density (+3,264)	Household PHBS (+1,466)
6. Tangerang District	13,784	Population density (+2,600)	Household PHBS (+1,146)
7. Bekasi District	13,464	Population density (+2,479)	Household PHBS (+962)
8. Karawang District	12,972	Population density (+2,179)	Adequate drinking water (+876)
9. Bekasi City	12,432	Population density (+2,575)	Public health centers (-1,147)
10. Tangerang City	12,317	Population density (+2,607)	General hospitals (+1,039)

For Bogor District, population contributed plus 18,212.21 predicted cases and density plus 5,375.13 relative to the model baseline of 1,136.00. Poverty contributed plus 1,633.30 cases, while BCG coverage had the largest negative contribution in that district. Population alone

accounted for 66.26% of the sum of absolute local SHAP contributions. This percentage summarizes the prediction decomposition and does not mean that population caused 66.26% of observed or projected TB cases.

In Bekasi City, population contributed plus 7,632.17 cases and density plus 2,575.06, whereas public health-center availability contributed minus 1,147.46. The negative facility contribution illustrates that local predictive directions can differ from a global importance ranking or a regression coefficient. Figure 8 compares the complete regional SHAP profiles for Bogor District and Bekasi City. Their baseline plus all nine contributions reconstructs each projected count within the small numerical discrepancies reported above.

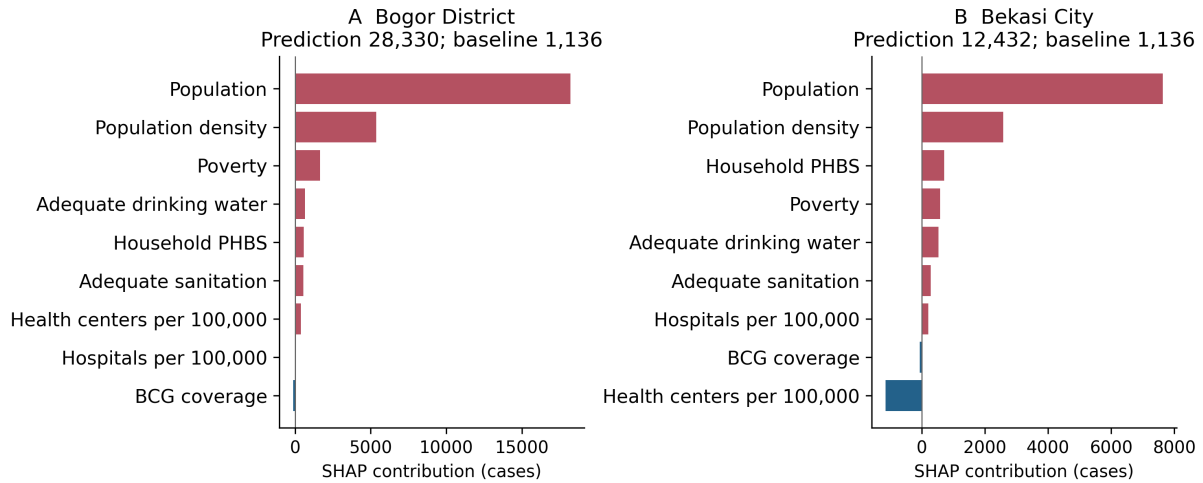


FIGURE 8. Local SHAP profiles for Bogor District and Bekasi City in the 2026 projections. Positive contributions increase the predicted count relative to the baseline; negative contributions decrease it.

3.8. Strengths, limitations, and implications. The main strength is the separation of inferential rate association, next-year count prediction, optimization, and model explanation. Distribution assessment compared four count families on the same contemporaneous development data, while predictive evaluation used earlier-year characteristics and later-year outcomes. All seven candidates excluded previous TB counts and were evaluated on identical observations. The 2025 temporal assessment, seeded baselines, two PSO restarts, saved candidate histories, and audited mapping provide transparent checks on the implemented workflow [15, 16, 36, 37].

Several limitations constrain association estimates. Notifications depend on access, diagnosis, case-finding intensity, completeness of reporting, and policy, as well as transmission [3, 4, 23, 24]. Ecological coefficients cannot be transferred to individual risk or interpreted as causal effects. District-clustered standard errors allow repeated observations within districts but do not model dependence between neighboring districts. The inferential model also omits an explicit temporal term, so unmodeled national time changes may affect the estimated regional associations.

The same expanding-window folds were used for tuning and reporting development performance. Consequently, their optimized averages may be favorable relative to fully nested temporal evaluation. Primary model selection was based on WAPE, while secondary metrics favored RF-PSO; the small ranking differences were not accompanied by uncertainty intervals for between-model differences. Geographic transfer was not tested because the same districts appeared in training and future-year evaluation. The retrospective 2025 temporal results establish performance for one later outcome year rather than general superiority across future periods or new regions [15, 16]. The 2025 outcomes had already been examined in earlier versions of the analysis. Although the current pipeline locked the primary model and hyperparameters from 2020-2024 validation before computing the 2025 metrics, the evaluation is retrospective and cannot be described as fully unseen prospective validation. Earlier knowledge of these outcomes may have influenced analytical development, so the test errors should be interpreted with this limitation.

The restricted feature set omitted previous case counts, calendar year, and several determinants of transmission and surveillance change. All fitted candidates showed negative 2025 bias, and persistence was substantially more accurate in that year. The 2026 projections therefore require prospective validation and should not be interpreted as a confirmed decline in national burden. Forecast uncertainty intervals, spatial dependence, and repeated evaluation on later years remain important extensions before operational adoption.

SHAP explanations are model-specific, sensitive to correlated covariates and background choices, and subject to Monte Carlo approximation. Small additivity errors verify reconstruction of the fitted prediction, not the correctness of a policy interpretation. High projected

counts and large contributions can guide surveillance review, but allocation decisions should also consider population denominators, reporting completeness, local epidemiology, and equity [17, 19, 37]. The model is most defensible as an aid to reviewing regional notification patterns alongside program knowledge and simpler forecasting references.

4. CONCLUSIONS

Recorded district-level TB counts in Indonesia were strongly overdispersed. NB2 provided the best fit among Poisson, generalized Poisson, NB1, and NB2 for contemporaneous 2016-2024 data, with a population offset and district-clustered robust uncertainty. Density, BCG coverage, general-hospital availability, adequate drinking-water access, and household PHBS had positive adjusted notification-rate associations. These associations concern aggregate recorded notifications and do not establish individual risk or causal effects.

XGBoost-PSO was selected by the lowest mean validation WAPE (32.74%) across five expanding-window folds using previous-year regional characteristics. On the temporally separated 2025 outcome year, its WAPE was 19.39%, MAE 327.25 cases, RMSE 714.36 cases, and mean Poisson deviance 121.56. RF-PSO performed slightly better on WAPE and MAE and better on deviance, while XGBoost-PSO minimized RMSE. PSO improved each ensemble's WAPE relative to its baseline, but persistence was more accurate than all seven fitted candidates on 2025.

SHAP identified population and density as the largest global contributors to XGBoost-PSO predictions. Refitting through the 2025 outcome yielded a projected 2026 total of 792,909.68 recorded cases, with Bogor District ranked highest. This projected total requires caution because of negative test-year bias and the absence of 2026 outcome validation. The framework supports ecological association analysis and interpretable regional forecasting while retaining explicit limits on predictive and causal conclusions.

ETHICS STATEMENT

This study used secondary, aggregate district/municipality-level data and involved no individual-level patient information, identifiable records, or direct interaction with human participants. Informed consent was therefore not applicable. All results are reported at aggregate or

district/municipality level.

DATA AND CODE AVAILABILITY

The harmonized dataset integrates publicly available district/municipality-level statistics from Statistics Indonesia with restricted health data obtained from the Ministry of Health of the Republic of Indonesia. Because the Ministry of Health data are subject to third-party access and reuse restrictions, the authors are not authorized to redistribute those data or the complete harmonized dataset. Researchers seeking access to the restricted data should submit a request through the Ministry of Health's official data-request service. Analysis code, temporal-fold summaries, model configurations, candidate histories, SHAP outputs, and other nonrestricted derived materials may be obtained from the corresponding author upon reasonable request. Requests should specify the intended scientific use and the materials required.

ACKNOWLEDGMENTS

The authors gratefully acknowledge Universitas Padjadjaran for institutional, academic, and administrative support. This research was funded by the Ministry of Higher Education, Science, and Technology of the Republic of Indonesia (Kementerian Pendidikan Tinggi, Sains, dan Teknologi; Kemdiktisaintek/DIKTI) through the RTM Program administered via BIMA (Basis Informasi Penelitian dan Pengabdian kepada Masyarakat), Contract No. 1706/UN6.3.1/PT.00/2026. The authors also acknowledge the Ministry of Health of the Republic of Indonesia, Statistics Indonesia, and the provincial and district health offices whose official aggregate publications supported construction and verification of the panel. The funder had no role in the study design, data analysis, interpretation, model selection, manuscript preparation, or decision to submit.

CONFLICT OF INTERESTS

The author(s) declare that there is no conflict of interests.

REFERENCES

- [1] United Nations, Goal 3: Ensure Healthy Lives and Promote Well-Being for All at All Ages, Target 3.3 and Indicator 3.3.2. Accessed 12 October 2026. <https://sdgs.un.org/goals/goal3>.

- [2] World Health Organization, Global Tuberculosis Report 2025, World Health Organization, 2025. <https://www.who.int/publications/i/item/9789240116924>.
- [3] H. Surendra, I.R.F. Elyazar, E. Puspaningrum, D. Darmawan, T.T. Pakasi, et al., Impact of the COVID-19 Pandemic on Tuberculosis Control in Indonesia: A Nationwide Longitudinal Analysis of Programme Data, *Lancet Glob. Health* 11 (2023), e1412–e1421. [https://doi.org/10.1016/S2214-109X\(23\)00312-1](https://doi.org/10.1016/S2214-109X(23)00312-1).
- [4] T. Lestari, Kamaludin, C. Lowbridge, E. Kenangalem, J.R. Poespoprodjo, et al., Impacts of Tuberculosis Services Strengthening and the COVID-19 Pandemic on Case Detection and Treatment Outcomes in Mimika District, Papua, Indonesia: 2014–2021, *PLOS Glob. Public Health* 2 (2022), e0001114. <https://doi.org/10.1371/journal.pgph.0001114>.
- [5] R. Arisanti, R.S. Pontoh, S. Winarni, N.A. Putri, S. Maurin, Identification and Modelling Tuberculosis Incidence Risk Factors in West Java with Negative Binomial Mixed Model Random Forest, *Cauchy* 10 (2025), 507–518. <https://doi.org/10.18860/cauchy.v10i2.29750>.
- [6] A. Farkhan, T.T. Pakasi, S. Sulisty, A. Salsabila, R.J. Maude, et al., Spatiotemporal Epidemiology and Associated Risk Factors of Tuberculosis Incidence and Mortality in Indonesia 2017–2022: a Nationwide Space-Time Hierarchical Analysis, *Popul. Health Metr.* 24 (2026), 12. <https://doi.org/10.1186/s12963-026-00458-5>.
- [7] J.A. Green, Too Many Zeros and/or Highly Skewed? A Tutorial on Modelling Health Behaviour as Count Data with Poisson and Negative Binomial Regression, *Health Psychol. Behav. Med.* 9 (2021), 436–455. <https://doi.org/10.1080/21642850.2021.1920416>.
- [8] L. Breiman, Random Forests, *Mach. Learn.* 45 (2001), 5–32. <https://doi.org/10.1023/A:1010933404324>.
- [9] T. Chen, C. Guestrin, XGBoost: A Scalable Tree Boosting System, in: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, New York, NY, USA, 2016, pp. 785–794. <https://doi.org/10.1145/2939672.2939785>.
- [10] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, et al., LightGBM: A Highly Efficient Gradient Boosting Decision Tree, in: *31st Conference on Neural Information Processing Systems (NIPS 2017)*, Long Beach, CA, USA, (2017).
- [11] P. Probst, M.N. Wright, A.-L. Boulesteix, Hyperparameters and Tuning Strategies for Random Forest, *WIREs Data Min. Knowl. Discov.* 9 (2019), e1301. <https://doi.org/10.1002/widm.1301>.
- [12] L. Yang, A. Shami, On Hyperparameter Optimization of Machine Learning Algorithms: Theory and Practice, *Neurocomputing* 415 (2020), 295–316. <https://doi.org/10.1016/j.neucom.2020.07.061>.
- [13] D. Freitas, L.G. Lopes, F. Morgado-Dias, Particle Swarm Optimisation: a Historical Review up to the Current Developments, *Entropy* 22 (2020), 362. <https://doi.org/10.3390/e22030362>.
- [14] J. Kennedy, R. Eberhart, Particle Swarm Optimization, in: *Proceedings of ICNN'95 - International Conference on Neural Networks*, Vol. 4, IEEE, 1995, pp. 1942–1948. <https://doi.org/10.1109/ICNN.1995.488968>.

- [15] V. Cerqueira, L. Torgo, I. Mozetič, Evaluating Time Series Forecasting Models: an Empirical Study on Performance Estimation Methods, *Mach. Learn.* 109 (2020), 1997–2028. <https://doi.org/10.1007/s10994-020-05910-7>.
- [16] H. Hewamalage, K. Ackermann, C. Bergmeir, Forecast Evaluation for Data Scientists: Common Pitfalls and Best Practices, *Data Min. Knowl. Discov.* 37 (2023), 788–832. <https://doi.org/10.1007/s10618-022-00894-5>.
- [17] S.M. Lundberg, G. Erion, H. Chen, A. DeGrave, J.M. Prutkin, et al., From Local Explanations to Global Understanding with Explainable AI for Trees, *Nat. Mach. Intell.* 2 (2020), 56–67. <https://doi.org/10.1038/s42256-019-0138-9>.
- [18] S.M. Lundberg, S.-I. Lee, A Unified Approach to Interpreting Model Predictions, in: 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.
- [19] M. Ghassemi, L. Oakden-Rayner, A.L. Beam, The False Hope of Current Approaches to Explainable AI in Health Care, *Lancet Digit. Health* 3 (2021), e745–e750. [https://doi.org/10.1016/S2589-7500\(21\)00208-9](https://doi.org/10.1016/S2589-7500(21)00208-9).
- [20] Kementerian Kesehatan Republik Indonesia, Profil Kesehatan Indonesia 2024, Kementerian Kesehatan Republik Indonesia, 2025. <https://www.kemkes.go.id/id/profil-kesehatan-indonesia-2024>.
- [21] Kementerian Kesehatan Republik Indonesia, Profil Kesehatan Indonesia 2025, Kementerian Kesehatan Republik Indonesia, 2026. <https://www.kemkes.go.id/id/profil-kesehatan-indonesia-2025>.
- [22] BPS-Statistics Indonesia, Badan Pusat Statistik, <https://www.bps.go.id>.
- [23] D.K. Sunjaya, C. Paskaria, D.M.D. Herawati, M. Pramayanti, R. Riani, et al., Initiating a District-Based Public–Private Mix to Overcome Tuberculosis Missing Cases in Indonesia: Readiness to Engage, *BMC Health Serv. Res.* 22 (2022), 110. <https://doi.org/10.1186/s12913-022-07506-4>.
- [24] B.W. Lestari, P.F. Hadisoemarto, N. Afifah, S. McAllister, D. Fattah, et al., Tuberculosis Care Provided by Private Practitioners in an Urban Setting in Indonesia: Findings from a Standardized Patient Study, *PLOS Glob. Public Health* 4 (2024), e0003311. <https://doi.org/10.1371/journal.pgph.0003311>.
- [25] M. Molemans, L. Kayaert, Q. Olislagers, S. Abrahams, N. Berkowitz, et al., Neighbourhood Factors and Tuberculosis Incidence in Cape Town: A Negative Binomial Regression and Spatial Analysis, *Trop. Med. Int. Health* 29 (2024), 599–611. <https://doi.org/10.1111/tmi.14001>.
- [26] glmmTMB Development Team, Family Functions for glmmTMB, Package Documentation, Accessed 12 October 2026. <https://glmmTMB.github.io/glmmTMB/reference/nbinom2.html>.
- [27] J.G. MacKinnon, M.Ø. Nielsen, M.D. Webb, Cluster-Robust Inference: a Guide to Empirical Practice, *J. Econometrics* 232 (2023), 272–299. <https://doi.org/10.1016/j.jeconom.2022.04.001>.
- [28] XGBoost Development Team, XGBoost Parameters: count:poisson Objective, Official Software Documentation, Accessed 12 October 2026. <https://xgboost.readthedocs.io/en/stable/parameter.html>.
- [29] LightGBM Development Team, Parameters: Poisson Objective, Official Software Documentation, Accessed 12 October 2026. <https://lightgbm.readthedocs.io/en/stable/Parameters.html>.

- [30] S. González, S. García, J. Del Ser, L. Rokach, F. Herrera, A Practical Tutorial on Bagging and Boosting Based Ensembles for Machine Learning: Algorithms, Software Tools, Performance Study, Practical Perspectives and Opportunities, *Inf. Fusion* 64 (2020), 205–237. <https://doi.org/10.1016/j.inffus.2020.07.007>.
- [31] B.M. Greenwell, Fast Approximate Shapley Values, *fastshap* Package Documentation, Accessed 12 October 2026. <https://bgreenwell.github.io/fastshap/reference/explain.html>.
- [32] I. Dwinata, T. Tsheten, Ansariadi, F. Wagnew, K.A. Alene, et al., Geospatial Codistribution of Tuberculosis and Diabetes Mellitus in Indonesia, *Infect. Dis. Poverty* 15 (2026), 37. <https://doi.org/10.1186/s40249-026-01432-x>.
- [33] L. Martinez, O. Cords, Q. Liu, C. Acuna-Villaorduna, M. Bonnet, et al., Infant BCG Vaccination and Risk of Pulmonary and Extrapulmonary Tuberculosis Throughout the Life Course: A Systematic Review and Individual Participant Data Meta-Analysis, *Lancet Glob. Health* 10 (2022), e1307–e1316. [https://doi.org/10.1016/S2214-109X\(22\)00283-2](https://doi.org/10.1016/S2214-109X(22)00283-2).
- [34] A.M. Liyew, A.C.A. Clements, T.Y. Akalu, B. Gilmour, K.A. Alene, Ecological-Level Factors Associated with Tuberculosis Incidence and Mortality: A Systematic Review and Meta-Analysis, *PLOS Glob. Public Health* 4 (2024), e0003425. <https://doi.org/10.1371/journal.pgph.0003425>.
- [35] J.-Y. Lee, N. Kwon, G.-Y. Goo, S.-I. Cho, Inadequate Housing and Pulmonary Tuberculosis: A Systematic Review, *BMC Public Health* 22 (2022), 622. <https://doi.org/10.1186/s12889-022-12879-6>.
- [36] G.S. Collins, K.G.M. Moons, P. Dhiman, R.D. Riley, A.L. Beam, et al., TRIPOD+AI Statement: Updated Guidance for Reporting Clinical Prediction Models That Use Regression or Machine Learning Methods, *BMJ* 385 (2024), e078378. <https://doi.org/10.1136/bmj-2023-078378>.
- [37] J. Wiens, S. Saria, M. Sendak, M. Ghassemi, V.X. Liu, et al., Do No Harm: A Roadmap for Responsible Machine Learning for Health Care, *Nat. Med.* 25 (2019), 1337–1340. <https://doi.org/10.1038/s41591-019-0548-6>.