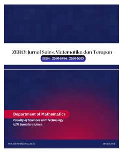
Zero : Jurnal Sains, Matematika, dan Terapan
E-ISSN : 2580-5754; P-ISSN : 2580-569X
Volume 10, Number 2, 2026
DOI: 10.30829/zero.v10i2.31161
Page: 818-830



Ensemble Anomaly Detection and SHAP-Based Attribution for Mapping Educational Inequality across Indonesian Provinces

[1]Emeylia Safitri
Statistics Study Program, Universitas Terbuka, South Tangerang, 15437, Indonesia

[2]I Gusti Ngurah Sentana Putra
Statistics and Data Science, IPB University, Bogor, 15580, Indonesia

[3]Yeni Rahkmawati
Department of Statistics, Universitas Lambung Mangkurat, Banjar Baru, 70714, Indonesia

[4]Ika Nur Laily Fitriana
Statistics Study Program, Universitas Terbuka, South Tangerang, 15437, Indonesia

[5]Nuramaliyah
Statistics Study Program, Universitas Terbuka, South Tangerang, 15437, Indonesia

[6]Ria Faulina
Statistics Study Program, Universitas Terbuka, South Tangerang, 15437, Indonesia

Article Info

Article history:

Accepted : 30 August 2026

Keywords:

Anomaly Detection;
Educational Inequality;
Isolation Forest;
SHAP.

ABSTRACT

Educational inequality across Indonesia's 38 provinces remains a challenge to equitable development. This study identifies anomalous provincial educational profiles by comparing Isolation Forest, Local Outlier Factor, and One-Class SVM across 50 parameter configurations in five analytical categories. The framework integrates ensemble majority voting, SHAP-based attribution, and leave-one-out robustness testing. Under the Silhouette-based selection criterion, Isolation Forest with contamination 0.05 ranked highest in four categories, with Silhouette Scores of 0.47-0.60 and Stability Scores of 1.0. However, sensitivity analysis showed that this criterion tends to favor configurations detecting fewer anomalies, so IF (0.05) should not be considered unambiguously superior. Highland Papua was consistently identified as anomalous across all categories. SHAP highlighted socioeconomic indicators, particularly rural child-labor participation, as important contributors to anomaly scores. Given the single-year design, 38 observations, and high feature-to-sample ratio, findings should be interpreted as exploratory descriptive mapping rather than causal or broadly generalizable inference.

Corresponding Author:

Emeylia Safitri,
Statistics Study Program
Universitas Terbuka, South Tangerang, Indonesia
Email: emeylia.safitri@ecampus.ut.ac.id

1. INTRODUCTION

Indonesia, the world's fourth most populous nation and an archipelagic state of 38 provinces, faces persistent challenges in equitable educational outcomes: substantial disparities persist between Java-centric regions and outer island territories despite significant infrastructure investment [1],[2], with remote regions, particularly Papua, facing compounding deficits [1]. 2025 Rapor Pendidikan data show the 15–24-year literacy rate in Highland Papua and Central Papua at only 90.5%, nine points below the 99.3% national average, while 7-12-year-olds not attending school in these provinces (18.0% and 16.7%) run more than four times the 3.9% national rate. The Rapor Pendidikan Indonesia framework (Kemendikdasmen) enables systematic analysis of these inequalities [2], yet most existing studies use clustering or regression that groups provinces into homogeneous categories rather than identifying genuinely anomalous profiles[1], [3]. Complementary work has applied predictive machine learning to Indonesian socioeconomic indicators: Alfia et al. [4] compared neural-network, regression, and tree-based models for poverty prediction, while Gonzales Martinez and Cooray [5] used spatial machine learning for poverty-targeting; both addressed supervised prediction of a single indicator rather than unsupervised, multidimensional anomaly identification.

Unsupervised anomaly detection offers a fundamentally different lens than clustering or single-target prediction: it identifies observations whose statistical profiles deviate significantly from the overall distribution [6], [7], [8]. Isolation Forest (IF) [9] isolates anomalies via random recursive partitioning with linear time complexity; Local Outlier Factor (LOF) [7] quantifies anomalousness through local density comparison; One-Class SVM [10] learns a kernel-space hyperplane encompassing most training observations. Each embodies distinct assumptions about anomaly structure, motivating systematic comparative evaluation, as categorized in recent surveys spanning statistical, density-based, distance-based, and ensemble families [11].

These three algorithms have been validated across domains sharing this study's central structural feature: a small share of genuine anomalies within a much larger normal population. Airlangga [8] compared LOF, IF, and One-Class SVM for Indonesian seismic anomaly detection, finding One-Class SVM led on silhouette score while IF led on Davies-Bouldin, a metric-ranking divergence that foreshadows the multi-metric protocol adopted in Section 2.4. Mutmainah and Yustanti [12] compared LOF and IF for faculty-performance anomaly detection, an administrative-monitoring setting structurally analogous to province-level educational monitoring; Oliveira et al. [13] applied unsupervised anomaly detection to household spending patterns; and Papastefanopoulos et al. [14] proposed a voting-based meta-learning ensemble, anticipating the ensemble design in Section 2.5. Kim and Park [15] reported IF dominance in streaming-data anomaly detection, consistent with the present findings, while Sidek et al. [6] found LOF sensitivity to dimensional complexity directly relevant given this study's high feature-to-sample ratio (Section 2.7). Chugh et al. [16] corroborated Isolation Forest's robustness under imbalanced conditions like the skewed anomaly proportions observed here.

Interpretability is a critical dimension for policy-relevant anomaly detection. SHAP [17] provides a unified framework for attributing model predictions to features. Dananjaya et al. [2] combined SHAP with clustering using the same Rapor Pendidikan Indonesia source, though restricted to Bali; Huang et al. [18], Delprato [19], and Zaman et al. [20] applied SHAP to PISA and geospatial education data internationally, confirming it as an established standard, though each addresses supervised student-level prediction rather than unsupervised, province-level anomaly identification. Silhouette [21], Davies-Bouldin [22], and Calinski-Harabasz [23] provide quantitative evaluation of anomaly-normal separation. To date, however, no study, particularly within the Indonesian education sector, has combined SHAP interpretability, ensemble majority voting, and Leave-One-Out (LOO) robustness testing within a single integrated analytical pipeline applied to multidimensional, province-level educational data; this narrower, LOO-anchored integration constitutes the specific methodological gap this study addresses. The contribution of this study is therefore not the anomaly detection algorithms themselves, which are individually well established, but their systematic integration into a four-stage analytical framework: comparative algorithm screening, ensemble consensus, explainable attribution, and leave-one-out robustness testing. Concretely, the LOO step provides a per-configuration within-procedure robustness assessment that ensemble voting alone does not. While ensemble voting measures agreement across parameter choices, LOO assesses whether a province's flagged status remains stable when individual observations are excluded from model training, providing a complementary measure of sensitivity to individual observations.

This study addresses this gap by proposing an integrated ensemble anomaly detection and SHAP-based attribution framework for mapping multidimensional educational inequality across Indonesian provinces. The framework evaluates five analytical model categories using IF, LOF, and One-Class SVM across 50 parameter configurations and a six-metric evaluation protocol. Ensemble majority voting, SHAP-based feature attribution [24], [25], dimensionality-reduction visualizations using Principal Component Analysis (PCA) [26], [27], t-distributed Stochastic Neighbor Embedding (t-SNE) [28], and Uniform Manifold Approximation and Projection (UMAP) [29], together with LOO robustness testing, are integrated to produce systematic and interpretable findings. The objectives of this study are: (1) to identify provinces exhibiting anomalous educational profiles across multiple dimensions; (2) to compare the performance of alternative algorithms and parameter configurations; and (3) to determine the educational indicators that contribute most strongly to the model-generated anomaly scores through SHAP-based attribution.

2. RESEARCH METHOD

2.1 Data

This study utilized provincial-level educational data from all 38 Indonesian provinces for 2025, sourced from the Rapor Pendidikan Indonesia released by Kemendikdasmen. The dataset encompasses comprehensive educational indicators spanning five analytical dimensions: physical school infrastructure and sanitation conditions, teacher workforce qualifications and student participation rates, ICT connectivity, and multidimensional educational outcome measures. The dataset covers all 38 provinces, including the newest Papua provincial subdivisions established through administrative reorganization.

2.2 Analytical Model Categories

Five model categories were constructed to enable multidimensional analysis: (1) the Overall Model, incorporating all available educational variables; (2) the Infrastructure Model, focusing on physical school conditions, sanitation facilities, and school ownership structure; (3) the Teacher and Student Model, analyzing school participation rates (APS), out-of-school children (ATS/OOSC) indicators, teacher qualification ratios, and student-teacher ratios across educational levels; (4) the ICT Access Model, evaluating digital connectivity indicators including internet and computer access across urban and rural contexts, motivated by well-documented digital divide patterns in education [30], [31]; and (5) the Educational Achievement Model, measuring outcome variables including literacy rates, school completion rates, mean years of schooling, and economic participation constraints on school attendance.

2.3 Anomaly Detection Algorithms

Three anomaly detection algorithms, Isolation Forest (IF), Local Outlier Factor (LOF), and One-Class SVM, were systematically evaluated with multiple parameter configurations. IF isolates anomalies via random feature partitioning, using shorter isolation paths to compute anomaly scores, with three contamination rates (auto, 0.05, 0.10). LOF detects outliers based on local density deviation from k-nearest neighbors, producing scores substantially greater than 1.0 for anomalies, tested with k = 5, 10, and 20. One-Class SVM learns a hyperplane separating normal data from the origin using kernel functions, with four kernels (linear, poly, rbf, sigmoid), resulting in 10 algorithm parameter combinations per model category and 50 total across all models.

2.4 Evaluation Framework

Six evaluation metrics were used to assess anomaly-normal separation quality. Silhouette Score [21] (ranging from -1 to 1, higher is better) measures how well each observation fits its cluster, while Davies-Bouldin Index [22] (lower is better) quantifies cluster similarity, and Calinski-Harabasz Index [23] (higher is better) computes the ratio of between cluster to within cluster dispersion. Stability Score evaluates consistent anomaly label assignments across 30 random repetitions using Jaccard similarity, and Relative Anomaly Score Consistency (RASC) measures the stability of relative score rankings via Spearman correlation. The Proportion of Anomalies (PoA) records the fraction of provinces identified as anomalous, with the best models selected by maximizing the Silhouette Score as the most direct measure of separation quality. Silhouette Score. For province i, the silhouette value is defined as:

$$s(i) = \frac{b(i)-a(i)}{\max\{a(i),b(i)\}} \tag{1}$$

where $a(i)$ is the mean distance from province i to other provinces sharing its label (anomaly or normal) and $b(i)$ is the mean distance from i to provinces of the other label. The reported Silhouette Score is the mean of $s(i)$ over all 38 provinces [21]. Davies–Bouldin Index. The index is defined as:

$$DBI = \frac{1}{2}\sum_{i=1}^{2} \max_{j\neq i}\left[\frac{\sigma_i+\sigma_j}{d(c_i,c_j)}\right] \tag{2}$$

for groups $i = 1, 2$ (anomaly, normal), where σ_i is the mean within-group scatter of group i, c_i its centroid, and $d(c_i, c_j)$ the Euclidean distance between centroids [22]. Calinski–Harabasz Index. The index is defined as:

$$CHI = \frac{\mathrm{Tr}(B)/(k-1)}{\mathrm{Tr}(W)/(n-k)} \tag{3}$$

with $k = 2$ groups, B the between-group and W the within-group dispersion matrices [23]. Isolation Forest anomaly score. The anomaly score is defined as:

$$s(x,n) = 2^{-\frac{E[h(x)]}{c(n)}}, \qquad c(n) = 2H(n-1) - \frac{2(n-1)}{n} \tag{4}$$

where $E[h(x)]$ is the average path length of x across the forest of isolation trees and c(n) is the average path length of an unsuccessful search in a binary search tree, with $H(\cdot)$the harmonic number [9]. Local Outlier Factor. The factor is defined as:

$$LOF_k(x) = \frac{\frac{1}{|N_k(x)|}\sum_{o\in N_k(x)} \mathrm{lrd}_k(o)}{\mathrm{lrd}_k(x)} \tag{5}$$

where $N_k(x)$ is the k-nearest-neighborhood of x and lrd_k is the local reachability density [7]. One-Class SVM objective. The optimization problem is:

$$\min_{w,\xi,\rho} \frac{1}{2} \| w \|^2 + \frac{1}{\nu n}\sum_i \xi_i - \rho \quad \text{s.t. } (w \cdot \phi(x_i)) \geq \rho - \xi_i,\ \xi_i \geq 0 \tag{6}$$

subject to $(w \cdot \varphi(x_i)) \geq \rho - \xi_i, \xi_i \geq 0$ for all i, where $\nu \in (0,1]$ bounds the expected anomaly fraction and φ is the kernel feature map [10]. Proportion of Anomalies (PoA). The proportion is defined as:

$$PoA = \frac{|A|}{n} \tag{7}$$

where A is the set of provinces flagged anomalous by a given configuration. Ensemble majority-voting rule. A province p is an ensemble anomaly if:

$$\frac{1}{M}\sum_{m=1}^{M} \mathbb{1}[p \in A_m] > 0.5 \tag{8}$$

where A_m is the anomaly set from configuration m and M = 10 configurations per model category. Stability Score. For $R = 30$ random repetitions producing anomaly sets $A_1, \dots, A_R$, the score is:

$$\text{Stability} = \frac{2}{R(R-1)}\sum_{r<r'} J(A_r, A_{r'}), \qquad J(A,B) = \frac{|A\cap B|}{|A\cup B|} \tag{9}$$

the mean pairwise Jaccard similarity across all repetition pairs. Relative Anomaly Score Consistency (RASC). The score is defined as:

$$RASC = \frac{1}{R}\sum_{r=1}^{R} \rho_s\,(\mathrm{rank}_r, \mathrm{rank}_{\mathrm{full}}) \tag{10}$$

the mean Spearman rank correlation between the anomaly-score ranking of provinces in repetition r and the full-data ranking. Leave-One-Out Consistency Score. The score is defined as:

$$LOO = \frac{1}{n}\sum_{i=1}^{n} J\,(A_{\mathrm{full}}\{i\},\ A_{-i}) \tag{11}$$

where A_{-i} is the anomaly set obtained after excluding province i and retraining on the remaining $n - 1$ provinces.

2.5 Ensemble Detection, Interpretability, and Robustness Testing

An ensemble detection approach using majority voting was implemented across all 10 algorithm parameter configurations per model, classifying a province as an ensemble anomaly if more than 50% of configurations independently detected it as anomalous, thereby reducing dependence on any single algorithm or parameter choice. SHAP TreeExplainer was applied to the best Isolation Forest models to quantify the contribution of each educational variable to province-level anomaly scores, producing feature-importance rankings that support exploratory interpretation of model behaviour. Leave-One-Out (LOO) robustness testing iteratively excluded one province at a time, retrained the best model on the remaining 37 provinces, and compared the resulting anomaly set with the full model result using Jaccard similarity. The LOO Consistency Score, calculated as the average Jaccard similarity across all excluded provinces, approaches 1.0 when detected anomalies represent robust data-driven patterns rather than artifacts of individual observations.

2.6 Software, Tools, and Ethical Considerations

All analyses were implemented in Python 3 using scikit-learn (Isolation Forest, LOF, One-Class SVM), the shap library (TreeExplainer), and standard scientific computing packages (NumPy, pandas, Matplotlib) for data processing and visualization. As this study relied exclusively on publicly available, province-level aggregate statistics released by Kemendikdasmen, with no individual-level or identifiable human-subject data, no formal ethical approval was required.

2.7 Methodological Justifications, Data Dictionary, and Limitations

Several methodological choices require clarification. The Silhouette Score was used as the primary selection criterion because it directly measures within-label compactness and between-label separation, although its use for anomaly detection may favor configurations identifying only a few extreme observations. This issue is particularly relevant for the 38-province dataset, where IF with contamination 0.05 mechanically flags only one or two anomalies. All variables were standardized using StandardScaler before model fitting. Sensitivity checks showed that DBI- and CHI-based reselection produced the same best configuration in four of five categories, while a semi-synthetic test showed that IF (0.05) recovered only 19-34% of injected anomalies, compared with 83-99% for IF (auto) and 100% for LOF (20). Thus, the IF (0.05) results should be interpreted as conservative, with ensemble findings providing complementary evidence. Table 1 further reports the feature counts for each model category to highlight the high-dimensional, low-sample setting.

Table 1. Feature Count and Feature-to-Sample Ratio per Model Category (n = 38 provinces)

Model	n Features	Feature-to-Sample Ratio (p : n)
Overall	117	3.08 : 1
Infrastructure	77	2.03 : 1
Teacher and Student	27	0.71 : 1
ICT Access	15	0.39 : 1
Educational Achievement	19	0.50 : 1

With 117 features for the Overall Model against 38 provinces (a feature-to-sample ratio of roughly 3:1), the Infrastructure Model at 77 features (2:1), and even the leaner Teacher and Student, ICT Access, and Educational Achievement Models at 27, 15, and 19 features respectively, this is a high-dimensional, low-sample-size regime in which distance- and density-based methods (LOF, One-Class SVM) degrade, PCA variance capture is unstable, and t-SNE/UMAP embeddings are, at this sample size, better understood as illustrative rather than statistically robust low-dimensional summaries. We report this explicitly rather than only in the concluding limitations, since it bears on how confidently the reported anomaly sets and their SHAP attributions should be read as generalizable patterns rather than as descriptive summaries specific to the 2025 cross-section.

For reproducibility, the complete dataset and data dictionary containing all 117 retained variables are publicly available at https://sl.ut.ac.id/edu-inequality-data. The dataset was compiled from the Rapor Pendidikan Indonesia 2025. All variables were taken directly from the source tables, and no additional ratios or composite indicators were constructed. Missing, blank, or non-numeric entries were converted to zero prior to analysis, which may affect anomaly scores where missingness is substantial. Across the five model categories, zero-filled cells accounted for 0.24%-3.02% of entries (Overall: 1.21%; Infrastructure: 0.24%; Teacher and Student: 3.02%; ICT Access: 1.58%; Educational Achievement: 1.25%), and Papua provinces carried a higher average share of these substitutions (2.7 cells per province) than non-Papua provinces (1.2 cells per province), suggesting that zero-filled entries were somewhat more frequent among Papua provinces.

A further limitation is the single cross-sectional year (2025) used throughout, which precludes distinguishing transient from structural anomalies; and province-level aggregation, which can mask substantial within-province heterogeneity. Because the entire framework is motivated by educational inequality, aggregating away sub-provincial variation is a genuine tension in the analysis rather than a closing caveat, and results should be read as describing province-level profiles rather than individual-level or district-level conditions.

3. RESULT AND ANALYSIS

3.1 Overview of Analysis

The educational-inequality anomaly detection analysis was carried out on data from all 38 Indonesian provinces for 2025 across five analytical model categories: Overall, Infrastructure, Teacher and Student, ICT Access, and Educational Achievement. Each model category was evaluated with three anomaly-detection algorithms Isolation Forest (three contamination rates: auto, 0.05, 0.10), Local Outlier Factor (three neighborhood sizes: 5, 10, 20), and One-Class SVM (four kernels: linear, poly, rbf, sigmoid) yielding 50 algorithm parameter combinations evaluated with six metrics (Silhouette Score, DBI, CHI, Stability Score, RASC, and PoA), further enriched with ensemble majority voting, SHAP TreeExplainer interpretability, PCA/t-SNE/UMAP dimensionality reduction visualization, and Leave-One-Out (LOO) robustness testing.

3.2 Algorithm Performance Comparison

The evaluation of anomaly-normal separation quality was conducted using the distribution of evaluation metrics across algorithms.

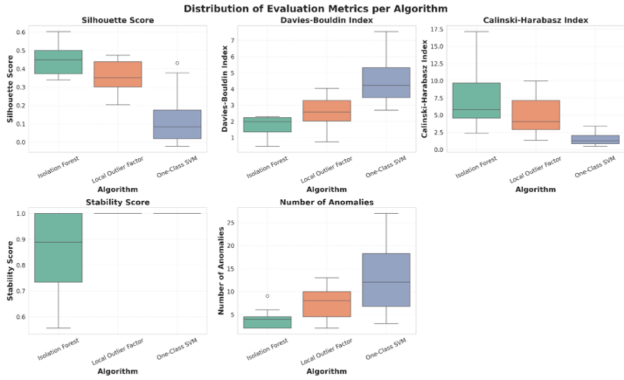


Figure 1. Distribution of Evaluation Metrics per Algorithm

Figure 1 presents the Silhouette Score, Davies-Bouldin Index (DBI), Calinski-Harabasz Index (CHI), Stability Score, and number of detected anomalies for Isolation Forest (IF), Local Outlier Factor (LOF), and One-Class SVM. Under the evaluated metrics, IF generally showed stronger separation and stability, achieving the highest median Silhouette Score, lowest DBI, highest CHI, and strongest Stability Score while producing relatively conservative anomaly counts. LOF and One-Class SVM showed comparatively weaker performance; however, these results should be interpreted within the high-dimensional, low-sample setting of this study rather than as evidence of their general inferiority. These findings align with recent international evidence. Chugh et al. [16] similarly found Isolation Forest more stable than density and boundary-based alternatives on highly imbalanced credit-card data, reinforcing its suitability here; the weaker performance of One-Class SVM is likewise consistent with Airlangga's [8] Indonesian seismic-anomaly benchmark, where kernel-based boundary methods were less stable than tree-based partitioning under moderate sample sizes, though that study found One-Class SVM competitive on silhouette score alone, a divergence attributable to differing feature dimensionality and contamination assumptions between domains.

3.3 Best Model Performance per Category

Table 2 presents the complete summary of best model performance per analysis category, and Figure 2 summarizes the best Silhouette Score achieved by each model along with the algorithm that produced that performance.

Table 2. Best Model Anomalies

Model	Algorithm	Parameter	Silhouette ↑	DBI ↓	CHI ↑	Stability ↑	RASC ↑	LOO ↑	Best-Model Anomalous Provinces
Overall	Isolation Forest	0.05	0.47	0.74	9.98	1.00	0.25	0.89	Central Papua, Highland Papua
Infrastructure	Local Outlier Factor	20	0.42	0.83	8.00	1.00	0.24	0.72	Central Papua, Highland Papua
Teacher & Student	Isolation Forest	0.05	0.59	0.46	17.16	1.00	0.34	0.97	Central Papua, Highland Papua
ICT Access	Isolation Forest	0.05	0.57	1.76	5.18	1.00	0.34	0.97	DI Yogyakarta, Highland Papua
Educational Achievement	Isolation Forest	0.05	0.60	0.56	17.10	1.00	0.32	0.97	Central Papua, Highland Papua

Note: DBI: Davies–Bouldin Index; CHI: Calinski–Harabasz Index; RASC: Relative Anomaly Score Consistency; LOO: Leave-One-Out Consistency Score. ↑ indicates that higher values are preferred; ↓ indicates that lower values are preferred.

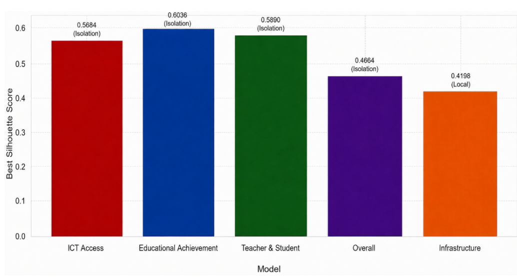


Figure 2. Best Silhouette Score per Model

Based on Table 2 and Figure 2, all models except Infrastructure selected IF (0.05) as best. Educational Achievement scored highest (Silhouette 0.60), followed by Teacher and Student (0.59), ICT Access (0.57), Overall (0.47), and Infrastructure (0.42, LOF (20)); Stability (1.0) was attained by all best models. Highland Papua and Central Papua were repeatedly identified among the most anomalous provincial profiles across the Overall, Teacher and Student, Educational Achievement, and Infrastructure models. Because anomaly scores from LOF and Isolation Forest are on different scales, their magnitudes should not be compared directly across algorithms. DI Yogyakarta was additionally flagged in the ICT Access Model not from disadvantage but its structurally advanced digital profile consistent with Zaman et al.'s [20] finding of administratively distinct urban centers as structural outliers and with digital-divide evidence elsewhere [30], [31]. While several anomalous profiles correspond to substantial educational disadvantages in Papua, DI Yogyakarta represents a structurally distinct ICT-access pattern worth further study.

3.4 Correlation Among Evaluation Metrics

Figure 3 presents the correlation matrix among the evaluation metrics used in this study. The correlation matrix shows a strong negative relationship between Silhouette Score and DBI (r = -0.879) and a strong positive relationship with CHI (r = 0.784), consistent with their respective measures of separation quality. The negative correlation between Silhouette Score and anomaly count (r = -0.753) indicates that the Silhouette criterion tends to favor configurations identifying fewer anomalies; therefore, this relationship should not be interpreted as independent evidence supporting the 0.05 contamination setting. RASC and Stability Score show weaker correlations with the other metrics because they capture ranking and label consistency rather than separation quality. These relationships should also be interpreted within the study's high-dimensional, low-sample setting.

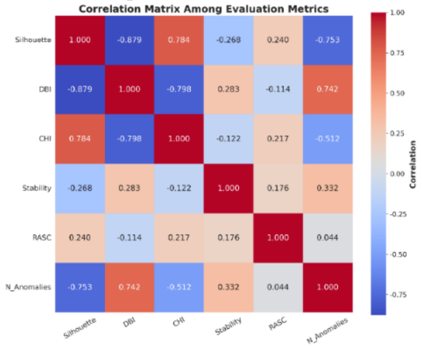


Figure 3. Correlation Matrix Among Evaluation Metrics

3.5 Anomaly Score Distribution for Best Models

Figure 4 displays the anomaly score distribution for the best model in each category. Red points indicate provinces classified as anomalous by the corresponding best-performing model. The dashed line indicates the 90th percentile of the displayed anomaly scores as a visual reference only and was not used as the model classification threshold.

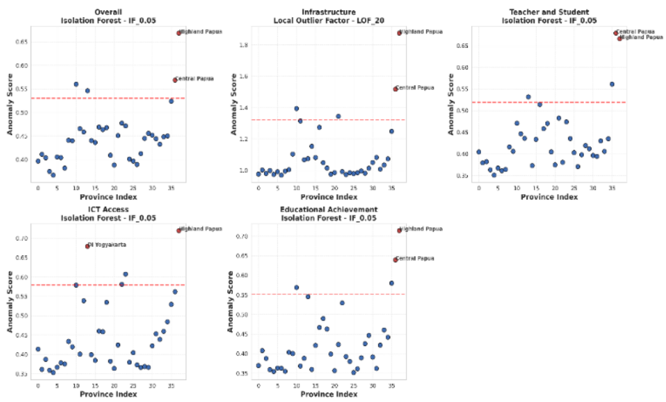


Figure 4. Anomaly Score Distribution for the Best Model per Category

Highland Papua consistently shows the highest anomaly score across nearly all models, followed by Central Papua. In the Overall Model, it reaches 0.68. In the Infrastructure Model (LOF (20)), it reaches 1.90 with an extremely elevated outlier factor; because LOF and Isolation Forest scores are on different scales, this figure is not directly comparable to the Overall Model's 0.68 and should be read as a separate, algorithm-specific severity indicator. In the ICT Access Model, DI Yogyakarta and Highland Papua are both flagged. DI Yogyakarta's presence reflects a structurally advanced digital profile rather than disadvantage.

3.6 Ensemble Anomaly Detection

To improve detection reliability and reduce dependence on any single algorithm, an ensemble approach based on majority voting across all 10 algorithm parameter combinations per model was applied. A province is classified as an ensemble anomaly if more than 50% of the configurations detect it as anomalous.

3.6.1 Vote Proportion per Province

Figure 5 displays each province's anomaly vote proportion across the five models; red bars exceed the 50% ensemble threshold. Highland Papua consistently exceeds this threshold across all five models, indicating that it has the most consistently anomalous profile across the five model categories. Flagged provinces vary by model: the ICT Access Model flags the most (five), spanning both severely underserved regions (Highland Papua) and structurally distinct access patterns (DI Yogyakarta, East Kalimantan, North Kalimantan, Central Java).

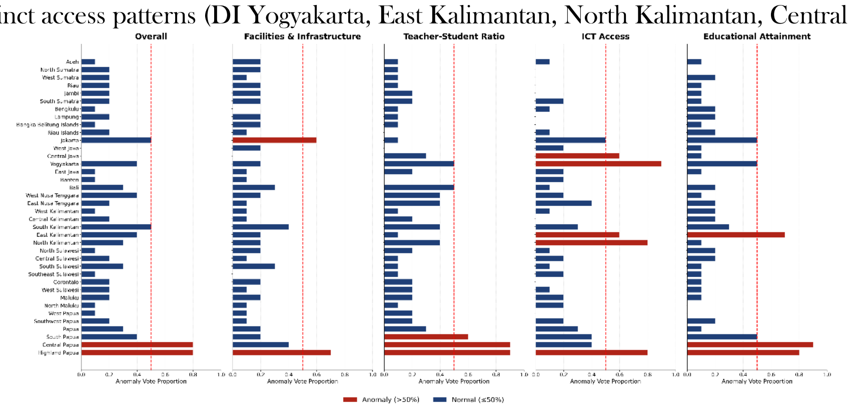


Figure 5. Ensemble Anomaly Vote Proportion per Province

3.6.2 Vote Matrix Heatmap

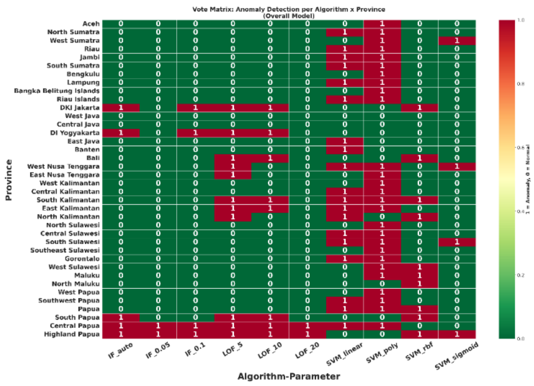


Figure 6. Vote Matrix Heatmap (Overall Model)

Figure 6 presents the vote matrix for the Overall Model. Highland Papua and Central Papua are consistently detected by most algorithms, especially IF and LOF, while One-Class SVM (linear and polynomial kernels) flags a substantially larger number of provinces, indicating lower selectivity under these parameterizations. DKI Jakarta and DI Yogyakarta are detected under IF (auto) and IF (0.10) but not consistently under IF (0.05), indicating that their classification is sensitive to the chosen parameterization.

3.6.3 Summary of Ensemble Anomalies

Table 3 summarizes the ensemble anomaly detection results across all model categories.

Table 3. Ensemble Anomaly Detection Results

Model	N	Anomalous Provinces (Ensemble, >50% Vote)
Overall	2	Central Papua, Highland Papua
Infrastructure	2	DKI Jakarta, Highland Papua
Teacher & Student	3	Papua Selatan, Central Papua, Highland Papua
ICT Access	5	Central Java, DI Yogyakarta, East Kalimantan, North Kalimantan, Highland Papua
Educational Achievement	3	East Kalimantan, Central Papua, Highland Papua

Note: Table 2 reports anomalies from the single best-performing configuration, whereas Table 3 reports anomalies based on majority voting across all ten configurations. Therefore, differences between the two tables reflect different decision rules rather than computational inconsistency.

Highland Papua is the only province flagged as an ensemble anomaly across all five models, indicating the most consistently anomalous multidimensional educational profile in this analysis; Central Papua appears in three models and East Kalimantan in two. DKI Jakarta's emergence as an Infrastructure ensemble anomaly reflects its megapolitan profile (private-school concentration, distinct classroom ratios, facility ownership) rather than backwardness. The ICT Access Model detects the most anomalies (five), reflecting two poles of digital inequality severely underserved and structurally advanced regions. The differences between Tables 2 and 3 are not computational inconsistency: Table 2 reports the single best-performing configuration while Table 3 reports consensus across all ten. DKI Jakarta was flagged by 6 of 10 configurations (60%) versus Central Papua's 4 of 10 (40%), so only DKI Jakarta crosses the majority threshold into Table 3. Because the LOO test (Section 3.9) evaluates only the single best-performing configuration (LOF (20) for Infrastructure, whose full-data set is Central Papua and Highland Papua), DKI Jakarta's status reflects the ensemble procedure specifically and should be read as parameter-sensitive rather than LOO-confirmed.

3.7 SHAP Feature Importance Analysis

Interpretability of the Isolation Forest models was carried out using SHAP, which quantifies the contribution of each variable to the anomaly score. Because SHAP TreeExplainer applies only to tree-based models, all reported feature-importance results including for the infrastructure category, whose best-performing configuration was LOF (20) rather than Isolation are derived from an Isolation Forest surrogate model; the infrastructure results in particular should therefore be read throughout as an approximation of, rather than a direct explanation of, that category's actual best-performing detector. The feature-importance results for each model category are summarized below, with the full plot shown only for the Overall Model.

3.7.1 Overall Model

The top-ranked feature was the percentage of learners aged 10-23 working in rural areas, followed by junior-secondary separate toilet ownership (0.0393), the number of private junior secondary schools (0.039), and primary/junior-secondary shared toilet ownership (0.0387, 0.0364) pointing to rural child labor and school-sanitation indicators as key correlates. Because these SHAP values were small and closely spaced, ranking differences should be read cautiously, not as evidence of causation.

3.7.2 Infrastructure Model

Because the best-performing Infrastructure detector was LOF (20), whereas SHAP TreeExplainer was applied to an Isolation Forest surrogate, the following values should be interpreted as approximate explanations of the surrogate model rather than direct explanations of the LOF (20) anomaly decisions. Within this surrogate model, the leading variables were the number of private primary schools (0.0551), primary student-per-classroom ratio (0.0507), senior-secondary separate toilet ownership (0.0467), vocational student-teacher ratio (0.0463), and senior-secondary water-source adequacy (0.0448), pointing to upper-secondary facility completeness as a marker of inequality.

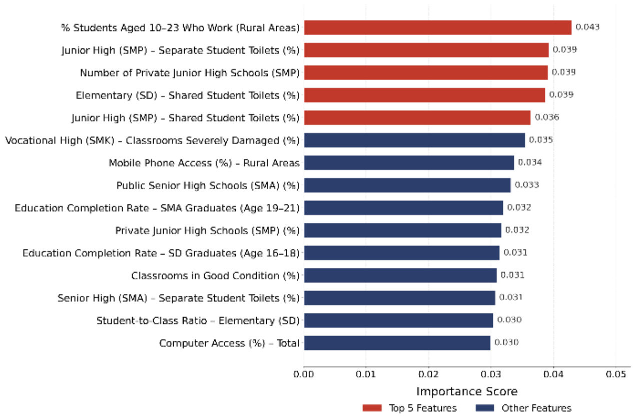


Figure 7. Feature Importance-Overall Model

3.7.3 Teacher and Student Model

The School Participation Rate (APS) for ages 13-15 ranked highest (0.1687), followed by APS for ages 7-12 (0.1108), senior-secondary student-teacher ratio (0.1050), Out-of-School Children rate for ages 7-12 (0.0991), and qualified primary-teacher percentage (0.0865). Participation rates at the primary/junior-secondary levels are the most discriminative dimension, with Central Papua and Highland Papua far below the national APS average.

3.7.4 ICT Access Model

Contributions were relatively balanced, led by total computer access (0.1595), village-level internet access (0.1594), total internet access (0.1455), village-level computer access (0.1455), and male mobile-phone access (0.1418). The prominence of village-level variables suggests that urban-rural differences in digital access are important contributors to the detected anomaly pattern.

3.7.5 Educational Achievement Model

Rural child labor had the highest SHAP importance (0.201), followed by literacy rate for ages 15-24 (0.135), junior-secondary completion for ages 16-18 (0.129), rural mean years of schooling (0.127), and rural household-duty participation (0.125), suggesting a link between economic/domestic burden and reduced schooling. Because SHAP attributes model behavior rather than causation on cross-sectional, aggregate data, this pattern is best read as a hypothesis consistent with a poverty-education cycle, warranting confirmation through longitudinal or quasi-experimental designs.

This pattern parallels recent international SHAP findings: Huang et al. [18] found self-efficacy and socioeconomic status outweighed school-resource indicators for PISA 2022 mathematics literacy across six East Asian systems, while Delprato [19] identified household poverty, paid child labor, and weak home ICT access as the strongest predictors of low PISA performance across ten Latin American cohorts closely mirroring this study's rural child-labor attribution. This consistency with independently derived international SHAP findings provides contextual support for the province-level patterns identified in the present study.

3.8 Dimensionality Reduction Visualization

To visualize the inequality pattern in a low-dimensional space, three dimensionality-reduction methods (PCA, t-SNE, UMAP) were applied using the ensemble majority-voting labels for coloring. Highland Papua and Central Papua consistently projected far from the main cluster, reinforcing the separation already established by the anomaly-score results. PCA variance capture ranged from 49.1% (Overall Model) to 73.2% (ICT Access Model, the highest of all categories); DKI Jakarta's Infrastructure-anomaly position was visually distinct; however, consistent with Section 3.6.3, it should not be read as independent confirmation given its ensemble-only, parameter-sensitive status. These visualizations are exploratory and are reported here only as a consistency check on the primary anomaly-detection results rather than as independent evidence.

3.9 Leave-One-Out Robustness Test

The LOO robustness test was conducted to evaluate how stable the anomaly detection results are when one province at a time is removed from the dataset. In each iteration, the model was retrained without one province, and the resulting anomalous-province set was compared with the full model result using Jaccard similarity. A Jaccard value close to 1 indicates very high consistency. Table 4 summarizes the LOO Consistency Score for each model.

Table 4. Leave-One-Out Robustness Test Results

Model	Algorithm	Anomalous Provinces (Best Model, Full Data)	LOO
Overall	IF (0.05)	Central Papua, Highland Papua	0.89
Infrastructure	LOF (20)	Central Papua, Highland Papua	0.72
Teacher & Student	IF (0.05)	Central Papua, Highland Papua	0.97
ICT Access	IF (0.05)	DI Yogyakarta, Highland Papua	0.97
Educational Achievement	IF (0.05)	Central Papua, Highland Papua	0.97

Four of the five models fall in the very-stable category (LOO ≥ 0.80): Teacher and Student, ICT Access, and Educational Achievement score highest, and Overall is also very stable, removing any single province barely changes the detected anomaly list. Infrastructure is the only merely stable model, likely reflecting greater cross-province heterogeneity in infrastructure variables. Highland Papua and Central Papua are detected in nearly 100% of LOO iterations for Overall, Teacher and Student, and Educational Achievement, indicating their anomalous status is highly stable under leave-one-out perturbation and is not driven by the inclusion of any single province in model fitting. DKI Jakarta does not appear in the Infrastructure LOO test because the protocol tracks only the best-performing configuration (LOF (20), whose full-data set is Central Papua and Highland Papua); its Infrastructure-anomaly status rests solely on the ensemble majority vote (6 of 10 configurations, Section 3.6), and its own leave-one-out robustness under that ensemble was not evaluated here, a direction for future extension. Unlike prior benchmarks such as Airlangga [8], Mutmainah and Yustanti [12], and Papastefanopoulos et al. [14], which reported point-in-time comparisons without an explicit stability check, this framework's LOO protocol adds direct quantitative evidence of within-procedure robustness those studies lacked.

3.10 Discussion

The results carry three implications. First, SHAP attribution in the Overall and Educational Achievement Models consistently highlighted rural child-labor participation and household-duty burden among the variables contributing most strongly to the model-generated anomaly scores. These patterns suggest that the observed educational disparities may involve socioeconomic constraints in addition to physical infrastructure. However, these associations should not be interpreted as causal effects. The Infrastructure Model requires additional caution because its best-performing detector was LOF (20), whereas the reported SHAP attribution was derived from an Isolation Forest surrogate model. Consequently, the Infrastructure feature rankings describe the behaviour of the surrogate model rather than directly explaining the LOF-based anomaly decisions. Second, the confounds discussed in Section 2.7 indicate that the IF (0.05) anomaly set in Table 2 represents a conservative selection rather than an exhaustive list, so the broader ensemble results in Table 3 should be viewed as complementary evidence reflecting agreement across parameter configurations. Third, the high-dimensional, low-sample setting, particularly the approximately 3:1 feature-to-sample ratio in the Overall Model, means that the weaker observed performance of LOF and One-Class SVM may reflect the structure of this dataset rather than their general unsuitability. Overall, the framework provides exploratory evidence for identifying unusual provincial educational profiles and dimensions warranting further investigation, but it does not establish causal determinants, definitive policy priorities, or generalizable relationships beyond the 2025 provincial cross-section.

4. CONCLUSION

This study applied an integrated ensemble anomaly-detection and SHAP-based attribution framework to map multidimensional educational inequality across Indonesia's 38 provinces. Across 50 algorithm-parameter configurations, Isolation Forest at contamination 0.05 achieved the highest Silhouette Score in four of five categories. Highland Papua was consistently identified as anomalous across all categories via ensemble voting, with best-model LOO consistency scores of 0.72-0.97, indicating a relatively stable profile. SHAP attribution highlighted socioeconomic indicators, particularly rural child-labor and household-duty participation, among the important contributors to model-generated anomaly scores, alongside physical-infrastructure indicators, though these are model-based associations rather than causal relationships. The infrastructure-category attribution in particular should be interpreted with additional caution, as it is derived from an Isolation Forest surrogate rather than that category's best-performing detector, LOF (20); given this and the study's setting of $n = 38$ with up to 117 features, these results are offered as an exploratory descriptive mapping rather than a generalizable inferential finding. DI Yogyakarta exhibited a distinctive ICT-access profile, while DKI Jakarta emerged as an Infrastructure anomaly only through ensemble voting and, not being confirmed under the best-model LOO procedure, should be interpreted cautiously. The principal contribution lies in the integration of established algorithms into a single workflow encompassing comparative screening across 50 configurations, ensemble consensus, SHAP attribution, and LOO robustness assessment, extending prior point-in-time comparisons and regional anomaly-mapping research with explicit robustness testing. The framework's cross-sectional, high-dimensional, and aggregation limitations are acknowledged, and two concrete extensions would most directly strengthen its evidentiary basis: pairing consecutive-year Rapor Pendidikan releases into a province-year panel to test whether anomalous status persists across years, and validating the SHAP-ranked socioeconomic drivers against matched district-level evidence from Highland Papua and Central Papua using granular Susenas microdata to test the framework's associations against finer-grained ground truth without a fully longitudinal redesign.

ACKNOWLEDGEMENT

The authors would like to thank Lembaga Penelitian dan Pengabdian Masyarakat (LPPM) Universitas Terbuka for funding this research.

5. REFERENCES

[1] S. Robiati, A. Hakim, G. Dharmawan, and C. Khotimah, “Application of K-Medoids for Regional Classification Based on Quality, Access, and Governance of Education in Indonesia,” *Proceedings of The International Conference on Data Science and Official Statistics*, vol. 2025, no. 1, pp. 1002–1016, Dec. 2025, https://doi.org/10.34123/icdsos.v2025i1.682.

[2] Md. W. P. Dananjaya, N. N. K. Krisnawijaya, G. H. Prathama, I. G. N. D. Paramartha, and A. W. O. Gama, “Analisis Determinan Karakter Siswa Menggunakan Explainable Machine Learning (SHAP) dan Klasterisasi Profil Sekolah Studi Kasus Rapor Pendidikan Provinsi Bali,” *Jurnal Kridatama Sains dan Teknologi*, vol. 7, no. 02, pp. 936–948, Dec. 2025, https://doi.org/10.53863/kst.v7i02.1988.

[3] T. Terttiaavini, A. Heryati, and T. S. Saputra, “Optimizing Socioeconomic Features for Poverty Prediction in South Sumatera,” *TIERS Information Technology Journal*, vol. 6, no. 1, pp. 16–32, Jun. 2025, https://doi.org/10.38043/tiers.v6i1.6244.

[4] L. A. Alfia, F. Nugroho, and Y. M. Arif, “Comparative Analysis of Artificial Neural Networks, Linear Regression, Random Forest, and Support Vector Machine for Predicting Poverty Levels in Indonesia,” *International Journal of Advances in Data and Information Systems*, vol. 6, no. 3, pp. 789–800, Dec. 2025, https://doi.org/10.59395/ijadis.v6i3.1467.

[5] R. G. Martinez and M. Cooray, “Enhancing Poverty Targeting with Spatial Machine Learning: An application to Indonesia,” 2025.

[6] Z. Sidek, S. S. S. Ahmad, and N. H. I. Teo, “Unsupervised outlier detection in high-dimensional text data: a comparative analysis,” *Bulletin of Electrical Engineering and Informatics*, vol. 14, no. 4, pp. 2997–3005, Aug. 2025, https://doi.org/10.11591/eei.v14i4.9573.

[7] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, “LOF,” in *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, New York, NY, USA: ACM, May 2000, pp. 93–104. https://doi.org/10.1145/342009.335388.

[8] G. Airlangga, “Advanced Machine Learning Techniques for Seismic Anomaly Detection in Indonesia: A Comparative Study of Lof, Isolation Forest, and One-Class SVM,” *Jurnal Lebesgue : Jurnal Ilmiah Pendidikan Matematika, Matematika dan Statistika*, vol. 5, no. 1, pp. 49–61, Apr. 2024, https://doi.org/10.46306/lb.v5i1.490.

[9] F. T. Liu, K. M. Ting, and Z.-H. Zhou, “Isolation-Based Anomaly Detection,” *ACM Transactions on Knowledge Discovery from Data*, vol. 6, no. 1, pp. 1–39, Mar. 2012, https://doi.org/10.1145/2133360.2133363.

[10] B. Schölkopf, R. Williamson, A. Smola, J. Shawe-Taylor, and J. Piatt, “Support vector method for novelty detection,” *Advances in Neural Information Processing Systems*, pp. 582–588, 2000.

[11] D. Samariya and A. Thakkar, “A Comprehensive Survey of Anomaly Detection Algorithms,” *Annals of Data Science*, vol. 10, no. 3, pp. 829–850, Jun. 2023, https://doi.org/10.1007/s40745-021-00362-9.

[12] M. Mutmainah and W. Yustanti, “Studi Komparasi Local Outlier Factor (LOF) dan Isolation Forest (IF) pada Analisis Anomali Kinerja Dosen,” *Journal of Informatics and Computer Science (JINACS)*, vol. 6, no. 02, pp. 532–540, Jul. 2024, https://doi.org/10.26740/jinacs.v6n02.p532-540.

[13] G. P. Oliveira, J. Castro Gertrudes, and R. B. Oliveira, “Spending pattern visualization using unsupervised machine learning,” in *Anais do XXXVIII Simpósio Brasileiro de Banco de Dados (SBBD 2023)*, Sociedade Brasileira de Computação - SBC, Sep. 2023, pp. 167–178. https://doi.org/10.5753/sbbd.2023.231577.

[14] V. Papastefanopoulos, P. Linardatos, and S. Kotsiantis, “Unsupervised Outlier Detection: A Meta-Learning Algorithm Based on Feature Selection,” *Electronics*, vol. 10, no. 18, p. 2236, Sep. 2021, https://doi.org/10.3390/electronics10182236.

[15] T. Kim and C. H. Park, “Anomaly Pattern Detection in Streaming Data Based on the Transformation to Multiple Binary-Valued Data Streams,” *Journal of Artificial Intelligence and Soft Computing Research*, vol. 12, no. 1, pp. 19–27, Jan. 2022, https://doi.org/10.2478/jaiscr-2022-0002.

[16] B. Chugh, N. Malik, D. Gupta, and B. S. Alkahtani, “A probabilistic approach driven credit card anomaly detection with CBLOF and isolation forest models,” *Alexandria Engineering Journal*, vol. 114, pp. 231–242, Feb. 2025, https://doi.org/10.1016/j.aej.2024.11.054.

[17] S. M. Lundberg and S. I. Lee, “A unified approach to interpreting model predictions,” *Advances in Neural Information Processing Systems*, vol. 2017-December, no. Section 2, pp. 4766–4775, 2017.

[18] Y. Huang, Y. Zhou, J. Chen, and D. Wu, “Applying Machine Learning and SHAP Method to Identify Key Influences on Middle-School Students’ Mathematics Literacy Performance,” *Journal of Intelligence*, vol. 12, no. 10, p. 93, Sep. 2024, https://doi.org/10.3390/jintelligence12100093.

[19] M. Delprato, “Identifying the post-pandemic determinants of low performing students in Latin America through interpretable Machine Learning methods,” *Engineering Applications of Artificial Intelligence*, vol. 179, p. 115154, Sep. 2026, https://doi.org/10.1016/j.engappai.2026.115154.

[20] B. Zaman *et al.*, "Modeling education impact: a machine learning-based approach for improving the quality of school education," *Journal of Computers in Education*, vol. 11, no. 4, pp. 1181–1214, Dec. 2024, https://doi.org/10.1007/s40692-023-00297-5.

[21] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, Nov. 1987, https://doi.org/10.1016/0377-0427(87)90125-7.

[22] D. L. Davies and D. W. Bouldin, "A Cluster Separation Measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-1, no. 2, pp. 224–227, Apr. 1979, https://doi.org/10.1109/TPAMI.1979.4766909.

[23] T. Caliński and J. Harabasz, "A dendrite method for cluster analysis," *Communications in Statistics*, vol. 3, no. 1, pp. 1–27, Jan. 1974, https://doi.org/10.1080/03610927408827101.

[24] J. Han, J. A. Guzman, and M. L. Chu, "Prediction of gully erosion susceptibility through the lens of the SHapley Additive exPlanations (SHAP) method using a stacking ensemble model," *Journal of Environmental Management*, vol. 383, p. 125478, May 2025, https://doi.org/10.1016/j.jenvman.2025.125478.

[25] S. M. Lundberg *et al.*, "From local explanations to global understanding with explainable AI for trees," *Nature Machine Intelligence*, vol. 2, no. 1, pp. 56–67, Jan. 2020, https://doi.org/10.1038/s42256-019-0138-9.

[26] E. Settanni and J. S. Srai, "It's a long way to the top (if you wanna biplot): a back-to-basics perspective on the implementation of principal component biplots in R," *Quality & Quantity*, vol. 60, no. 1, pp. 1173–1213, Jul. 2025, https://doi.org/10.1007/s11135-025-02266-9.

[27] H. Abdi and L. J. Williams, "Principal component analysis," *WIREs Computational Statistics*, vol. 2, no. 4, pp. 433–459, Jul. 2010, https://doi.org/10.1002/wics.101.

[28] L. van der Maaten and G. Hinton, "Visualizing Data using t-SNE," *Journal of Machine Learning Research*, vol. 9, no. 86, pp. 2579–2605, 2008.

[29] L. McInnes, J. Healy, N. Saul, and L. Großberger, "UMAP: Uniform Manifold Approximation and Projection," *Journal of Open Source Software*, vol. 3, no. 29, p. 861, Sep. 2018, https://doi.org/10.21105/joss.00861.

[30] K. V. Nayak and S. Alam, "The digital divide, gender and education: challenges for tribal youth in rural Jharkhand during Covid-19," *DECISION*, vol. 49, no. 2, pp. 223–237, Jun. 2022, https://doi.org/10.1007/s40622-022-00315-y.

[31] N. Singh, Dr. S. Singh, Dr. B. Goswami, Dr. S. Kumar, and B. Kumar, "Bridging The Digital Divide: A Comprehensive Analysis Of ICT Infrastructure In Rural Schools Of Jharkhand, India," *Educational Administration: Theory and Practice*, Jun. 2024, https://doi.org/10.53555/kuey.v30i6.7279.