Bulletin of Informatics and Data Science
Vol. 4 No. 2, November 2025, Page 114−123
ISSN 2580-8389 (Media Online)
DOI 10.61944/bids.v4i2.137
https://ejurnal.pdsi.or.id/index.php/bids/index

Hybrid Chaos-Isolation Forest Framework for Anomaly Detection in Indonesia's Public Procurement

Erlin Windia Ambarsari[1,*], Desyanti[2], Dedin Fathudin[3]

[1] Faculty of Engineering and Computer Science, Informatics Engineering, Universitas Indraprasta PGRI, DKI Jakarta, Indonesia
[2] Informatics Engineering, Institut Teknologi dan Bisnis Riau Pesisir, Riau, Indonesia
[3] Faculty of Computer Science, Informatics Engineering, Universitas Pamulang, South Tangerang, Indonesia
Email: [1,*]erlinunindra@gmail.com, [2]desyanti734@gmail.com, [3]dosen00398@unpam.ac.id

ARTICLE INFORMATION

ARTICLE HISTORY:
Submited : October 17, 2025
Revised : November 26, 2025
Accept : November 29, 2025
Publish : November 29, 2025

KEYWORD
Chaos Theory;
Isolation Forest;
Anomaly Detection;
Public Procurement;
Data Transparency

CORRESPONDENCE AUTHOR
Email: erlinunindra@gmail.com

A B S T R A C T

This study proposes and empirically evaluates a Hybrid Chaos-Isolation Forest (HC-iForest) framework for detecting anomalies in Indonesia's public procurement datasets. The purpose of this research is to address the difficulty of identifying irregular procurement patterns, as existing assessment mechanisms remain largely descriptive and retrospective. The framework integrates chaos-based temporal descriptors—permutation entropy, turning points, and volatility—with statistical indicators to enhance sensitivity to nonlinear and irregular time series. Using monthly procurement data from the Open Contracting Data Standard (OCDS) covering the period from 2019 to 2024, the model identified anomalous fiscal patterns associated with year-end budget adjustments and procurement surges. Empirical evaluation using correlation, ablation, and statistical validation shows that the hybrid model introduces non-redundant anomaly information, achieving a Spearman rank correlation of approximately 0.75 compared to the baseline Isolation Forest, with reduced overlap at intermediate thresholds (Jaccard similarity of 0.20 at the Top 5%). These results confirm that chaos-driven features improve model stability and interpretability. The findings reveal that anomalies are systemic manifestations of institutional and fiscal behavior rather than random deviations. The HC-iForest framework offers a data-driven early-warning mechanism for oversight agencies such as LKPP and ICW, strengthening transparency and accountability in public spending. Future studies may extend this framework through neural or spatiotemporal hybrid architectures to support intelligent and adaptive fiscal monitoring systems.

1. INTRODUCTION

Corruption remains a central threat to national integrity and economic resilience. Reports by the Corruption Eradication Commission (KPK) and Indonesia Corruption Watch (ICW) consistently identify public procurement as Indonesia's most corruption-prone sector, resulting in substantial financial losses due to price inflation, collusion, fictitious projects, and conflicts of interest [1]. These persistent issues undermine institutional credibility, weaken fiscal capacity, and diminish public confidence in government accountability. Over time, this erosion of trust has been linked to prolonged social and political instability, especially in countries with fragile governance. In Indonesia, this situation has been exacerbated by fluctuating global markets, mounting public debt, and weak economic management. These factors converged in mid-2025, when public demonstrations emerged across several cities in response to economic uncertainty and dissatisfaction with governance.

To address these governance challenges, Indonesia Corruption Watch (ICW) was founded as an independent civil society organization that monitors and reports public sector irregularities while promoting transparency. In line with its open-data mission, ICW disseminates procurement records based on the Open Contracting Data Standard (OCDS), making detailed information accessible through the national open-data portal. The availability of open contracting data enables researchers and the public to assess procurement transparency and efficiency, helping to uncover potential irregularities.

These open-access datasets are widely used to evaluate the effectiveness of procurement oversight in Indonesia. The dataset is derived from the government's Electronic Procurement System (LPSE) managed by the National Public Procurement Agency (LKPP). These open-access records are a critical resource for assessing transparency, detecting inefficiencies, and identifying procurement anomalies that may warrant institutional review.

Recent peer-reviewed studies confirm that corruption within public procurement persists despite substantial regulatory modernization. For instance, Suardi et al. [2] demonstrated that robust governance structures—particularly during planning and pre-tendering—can significantly mitigate corruption risks when supported by adequate competition and oversight. Anggriawan [3] categorized procurement collusion into horizontal, vertical, and hybrid forms, emphasizing that they continue to distort fair competition. Meanwhile, Firmansyah et al. [4] emphasized that the success of Indonesia's transition to e-procurement and e-catalog systems depends on ongoing human-capacity development and inter-agency coordination among LKPP, KPK, and regional governments.

Nevertheless, existing detection mechanisms remain largely descriptive and retrospective rather than predictive. Current assessments tend to rely on qualitative audits that cannot flag irregularities before they occur. To address this gap, recent developments in data-driven anomaly detection have combined chaos theory and machine learning to improve sensitivity to nonlinear behavioral patterns. Within public procurement systems, chaotic behavior emerges through nonlinear temporal fluctuations in aggregated procurement activities, where minor institutional or fiscal adjustments may lead to disproportionate changes in observed procurement patterns over time. Tan et al. [5] introduced the Sparse Random Projection Isolation Forest (SRP-iForest), which enhances anomaly detection in high-dimensional data through sparse projections.

Liu et al. [6] later formalized the theoretical basis for isolation-based ensemble learning, improving robustness across heterogeneous datasets. Zheng et al. [7] further advanced this approach by developing a Soft Voting Ensemble Isolation Forest that merges several iForest variants for higher stability. Complementary studies by Chen et al. [8] revealed that deep belief networks, long short-term memory (LSTM) systems, and reservoir computing frameworks exhibit comparable potential in modeling chaotic systems. Similarly, Sheng and Wang [9] demonstrated that chaotic neural networks outperform traditional architectures by improving anomaly-detection accuracy and reducing computational time.

Building on this foundation, this study develops a Hybrid Chaos-Isolation Forest (HC-iForest) framework that integrates statistical and chaos-based indicators to detect anomalies in Indonesia's public procurement data. The hybrid approach integrates chaos-based indicators—capturing irregular oscillations and sensitivity to initial conditions—with the Isolation Forest algorithm, which partitions multidimensional data to isolate anomalous observations. Unlike earlier studies that focused solely on corruption indicators or tree optimization, this research integrates chaos descriptors with machine-learning isolation mechanisms to detect early-warning signals of procurement anomalies. These anomalies are not interpreted as evidence of corruption but as quantitative precursors that guide preventive governance measures.

In practical terms, this study contributes to data-driven governance analytics by introducing an interpretable hybrid framework capable of detecting nonlinear irregularities within complex administrative datasets. Theoretically, it extends the application of chaos-informed learning beyond physical or cyber systems into the socio-economic domain of public procurement. Empirically, it provides an evidence-based mechanism that policymakers and auditors can employ to strengthen transparency, accountability, and fiscal oversight.

2. RESEARCH METHODOLOGY

2.1 Research Method Overview

This research adopts an empirical and systematic approach to develop the Hybrid Chaos-Isolation Forest (HC-iForest) framework for anomaly detection in Indonesia's public procurement data. The methodology was structured into sequential stages encompassing data collection, preprocessing, feature extraction, model training, evaluation, and validation. Each stage was designed to ensure transparency, reproducibility, and alignment with the study's predictive objectives. The overall analytical workflow of these stages is summarized in Figure 1.

2.2 Data Collection and Normalization

Procurement records were obtained from Indonesia's Open Contracting Data Standard (OCDS) platform, covering the period from 2019 to 2025. The raw data included multiple JSON entities—*releases*, *awards*, *buyers*, and *tender_items*—which were standardized into relational tables using automated audit scripts.

All data were consolidated into a multi-year panel to preserve temporal consistency and integrity. During preprocessing, missing timestamps, negative contract values, and invalid date ranges were identified and corrected. Duplicate and null records were removed to minimize bias and improve data quality. This data-cleaning phase followed open-data quality frameworks and anomaly detection standards outlined in previous studies [5], [10], [11]. The resulting dataset consisted of 447,878 valid records spanning from January 1, 2019, to November 18, 2024.

2.3 Feature Engineering and Chaos Quantification

Following normalization, the dataset was aggregated monthly, resulting in 71 temporal observations, to preserve long-term procurement dynamics while reducing short-term administrative noise. Each monthly entry contained three baseline statistical indicators: total tender value, average contract value, and total tender count. To capture nonlinear temporal dynamics, the model incorporated chaos-inspired descriptors as supplementary features. Four primary measures were used to quantify chaotic patterns and short-term irregularities: permutation entropy (PE), sample entropy (SampEn), turning-points ratio (TPR), and volatility (σ).

a. Permutation Entropy (PE)— measures complexity by evaluating the diversity of ordinal patterns within a time series [12]:

$$PE = -\sum_{i=1}^{n} p_i log(p_i) \qquad (1)$$

b. Sample Entropy (SampEn)— quantifies signal unpredictability, where A and B denote the number of matching sequences of lengths m+1 and m, respectively [13]

$$SampEn(m, r, N) = -ln\frac{A}{B} \quad (2)$$

c. Turning Points Ratio (TPR)— reflects directional changes, representing structural instability in the time series.

$$TPR = \frac{Number\ of\ Turning\ Points}{N-2} \quad (3)$$

Recent studies confirm that turning-point dynamics improve the responsiveness of anomaly detection in evolving systems [14], [15].

d. Volatility (σ) —measures short-term fluctuations based on the rolling standard deviation of the logarithmic total contract value, highlighting dynamic procurement dispersion [16], [17]:

$$\sigma = \sqrt{\frac{1}{w-1}\sum_{i=1}^{w}(x_i - \bar{x})^2} \quad (4)$$

Together, these descriptors quantify latent fluctuations that precede irregular procurement behavior. Empirical research supports that integrating entropy and volatility indicators enhances the detection of nonlinear anomalies in complex administrative systems [6], [10], [18], [19].

2.4 Anomaly Detection using Isolation Forest

Anomalies were identified with Isolation Forest (iForest), an ensemble-based unsupervised algorithm for high-dimensional outlier detection [20]. The algorithm isolates observations through recursive random partitioning, where rare and distinct data points require fewer partitions. The anomaly score s(x,n) is computed as:

$$s(x, n) = 2^{-\frac{E(h(x))}{c(n)}} \quad (5)$$

where E(h(x)) is the average path length required to isolate a data point x, and c(n) is the expected path length for unsuccessful searches, defined as:

$$c(n) = 2H(n-1) - \frac{2(n-1)}{n} \quad (6)$$

Model parameters were empirically tuned to balance sensitivity and efficiency: 300 estimators ensured stability and reduced variance, a contamination rate of 0.07 reflected the empirical anomaly frequency in large-scale procurement systems, and a random state of 42 ensured reproducibility.

Each monthly observation received both a continuous anomaly score and a binary label (−1 = anomalous, 1 = normal). Outputs—including scores, thresholds, and feature usage—were stored in CSV and JSON formats to support auditability.

2.5 Hybrid Chaos-Isolation Forest Model Construction

The hybrid model extends the baseline iForest by integrating chaos-driven descriptors (PE, SampEn, TPR, σ) with traditional indicators (log_total_value, log_avg_value, total_tenders). This integration allows the model to identify nonlinear shifts often invisible to purely statistical algorithms. The hybrid anomaly score was computed as:

$$S_{hybrid}(x) = \frac{1}{T}\sum_{t=1}^{T} S_t(x) \quad (7)$$

We then applied a 99.5th-percentile threshold to flag extreme outliers:

$$P_{thr} = Q_{0.995}(S_{hybrid}) \quad (8)$$

The high percentile threshold was intentionally selected to emphasize precision by isolating only the most extreme procurement anomalies over the multi-year period, thereby minimizing false-positive detections in aggregated temporal data. The contamination parameter (0.07) was used to control the internal sensitivity of the Isolation Forest during model training by allowing a broader exploration of potentially anomalous observations, whereas the 99.5th-percentile threshold was applied as a post-model decision rule to conservatively select high-confidence anomalies for interpretation. This separation ensures conceptual alignment between model sensitivity and anomaly reporting strictness.

Any record exceeding P_{thr} was labeled as anomalous. All configurations, cutoffs, and metadata were automatically logged to ensure transparency. To evaluate the incremental contribution of chaos features, an ablation comparison was conducted between:

a. The baseline model (statistical features only), and
b. The hybrid model (statistical + chaotic features).

Spearman's rank correlation assessed ranking consistency, while Jaccard similarity measured overlap across the Top 1%, Top 5%, and Top 10% anomaly thresholds, following responsible-AI best practices [21].

Bulletin of Informatics and Data Science
Vol. 4 No. 2, November 2025, Page 114−123
ISSN 2580-8389 (Media Online)
DOI 10.61944/bids.v4i2.137
https://ejurnal.pdsi.or.id/index.php/bids/index

2.6 Model Evaluation and Statistical Validation

This evaluation stage validated model interpretability and robustness through several statistical tests:

a. Spearman Rank Correlation

$$\rho_s = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \quad (9)$$

Used to measure ranking coherence between baseline and hybrid scores. A moderate correlation ($\rho_s \approx 0.75$) indicates the hybrid model introduces additional non-redundant variance.

b. Jaccard Similarity

$$J(A,B) = \frac{|A\cap B|}{|A\cup B|} \quad (10)$$

Quantifies overlap between anomaly sets from both models, validating chaos-based differentiation.

c. Welch's t-test and Cohen's d

$$t = \frac{\bar{x}_1-\bar{x}_2}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}, d = \frac{\bar{x}_1-\bar{x}_2}{s_p}, s_p = \sqrt{\frac{s_1^2+s_2^2}{2}} \quad (11)$$

These tests were applied to examine whether feature means differed significantly between anomalous and normal months. All statistical outputs were exported in structured CSV format with timestamps and unique run identifiers, ensuring full reproducibility under Explainable Artificial Intelligence (XAI) principles.

Additionally, cross-validation experiments were conducted to confirm that the hybrid model maintained consistent anomaly rankings across random data splits, ensuring robustness against sampling bias.

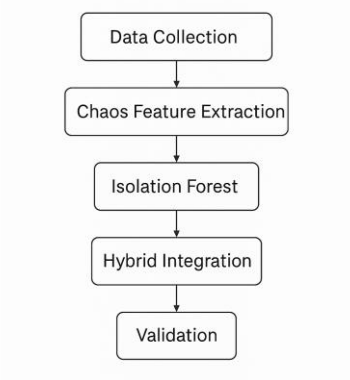

Figure 1. Conceptual and Procedural Framework of the Hybrid Chaos-Isolation Forest (HC-iForest)

Figure 1 illustrates the overall analytical workflow of the proposed HC-iForest framework. The process begins with data collection and normalization from the Open Contracting Data Standard (OCDS), followed by chaos feature extraction using Permutation Entropy, Turning Points, and Volatility to capture nonlinear temporal dynamics. These features are incorporated into the Isolation Forest modeling stage, where anomalies are identified through ensemble-based isolation mechanisms. The hybrid integration phase fuses chaos-based and statistical descriptors to enhance model sensitivity and interpretability. The workflow concludes with model validation and statistical testing, generating actionable insights that support governance transparency and early anomaly detection in public procurement.

3. RESULT AND DISCUSSION

3.1 Descriptive Results

The Hybrid Chaos-Isolation Forest (HC-iForest) framework was applied to 71 monthly observations spanning January 2019 to November 2024. After preprocessing and data auditing, 447,878 valid procurement transactions were retained for analysis. The monthly dataset comprised normalized metrics, including total tender value, average contract value, and tender count.

During the anomaly detection phase, the Isolation Forest algorithm identified five anomalous months (label = −1) when applying the 99.5th percentile threshold ($P_{thr} = 0.6566$) to the anomaly score distribution. This threshold corresponded to the top 1% of the anomaly score distribution, marking months that deviated significantly from baseline procurement behavior.

The anomaly-score distribution was right-skewed, indicating that most procurement activities followed normal fiscal patterns, whereas a small subset exhibited extreme irregularities. In the early aggregation period, the mean logarithmic total tender value reached 23.02, with volatility and permutation entropy increasing markedly during anomalous intervals.

Figure 2 illustrates the temporal aggregation of monthly procurement records from 2019 to 2024, along with derived statistical and chaotic descriptors. The dataset comprises 71 monthly observations, of which five (≈7%) were identified as anomalous by the Hybrid Chaos-Isolation Forest (HC-iForest) model. Key non-constant features include log_total_value, log_avg_value, total_tenders, perm_entropy, turn_points, and volatility.

	month	total_tenders	total_value	avg_value	month_dt	log_total_value	log_avg_value	perm_entropy	samp_entropy	turn_points	volatility
0	2019-01	3434	1.485266e+13	4.325177e+09	2019-01-01	30.329200	22.187719	NaN	NaN	NaN	NaN
1	2019-02	4748	1.874142e+13	3.947224e+09	2019-02-01	30.561757	22.096278	NaN	NaN	NaN	0.164443
2	2019-03	8124	2.950413e+13	3.631725e+09	2019-03-01	31.015551	22.012974	-5.581602e-13	NaN	0.0	0.349068

```
Months: 71 | Anomalies (label==-1): 5
Features used (non-constant): log_total_value, log_avg_value, total_tenders, perm_entropy, turn_points, volatility
Percentile threshold (99.5%): 0.6566408273607133 | Top-k = 1
Ablation – Top-1% overlap: 1.00
```

Figure 2. Data Artifacts and Analytical Outputs

Overall tender activity fluctuates substantially over time, with total_value ranging from 8.55×10^{7} to 2.95×10^{13} rupiah. The log-transformed features reduce skewness and stabilize variance for subsequent anomaly modeling. Early months (January–March 2019) show missing entropy values due to insufficient temporal window size, which is typical in permutation entropy estimation.

The 99.5th-percentile threshold (0.6566) identifies extreme outlier candidates, while the Top 1% overlap score (1.00) indicates perfect consistency between independent model runs. These characteristics confirm the framework's numerical stability and sensitivity to temporal irregularities in procurement activity. Log-transformed values use the natural logarithm unless otherwise specified.

3.2 Feature Correlation and Chaos Dynamics

To examine the intrinsic interaction between statistical and chaos-based indicators, a feature correlation heatmap was generated. The analysis (Figure 3) revealed weak to moderate relationships ($\rho = 0.45$–0.60) between permutation entropy, volatility, and total tender value, implying that chaotic complexity tends to increase with procurement intensity. In contrast, the Turning Points Ratio (TPR) exhibited a mildly negative correlation with entropy, suggesting that periods of higher systemic disorder are accompanied by reduced directional stability.

These relationships underscore the complementary role of chaos-based metrics. While statistical features such as total value or tender count represent the magnitude of economic activity, entropy and volatility capture structural instability within the same time frame. This pattern aligns with prior studies showing that entropy-based indicators effectively model socio-economic complexity and detect pre-perturbation behavior in procurement systems.

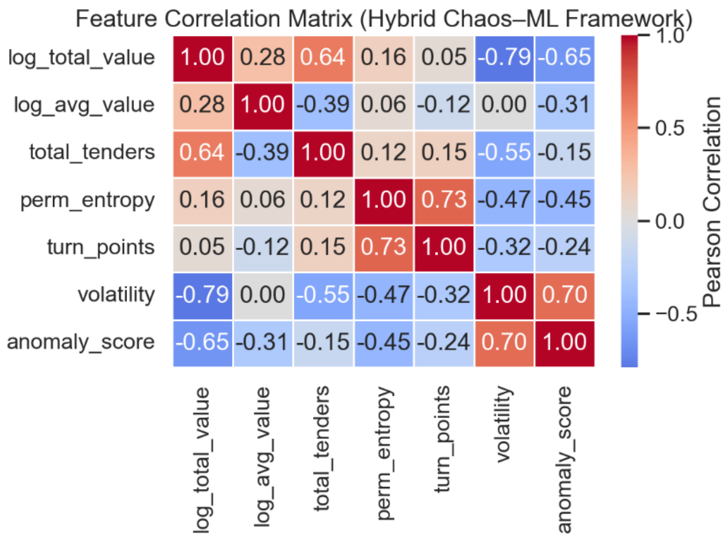

Figure 3. Correlation Heatmap of Chaotic and Statistical Features.

Figure 3 presents the Pearson correlation matrix among the statistical and chaotic features. The highest positive correlation ($r = 0.64$) appears between log_total_value and total_tenders, reflecting a natural tendency for higher transaction volumes to accompany larger tender counts. Conversely, volatility exhibits a strong negative correlation with

Vol. 4 No. 2, November 2025, Page 114−123
ISSN 2580-8389 (Media Online)
DOI 10.61944/bids.v4i2.137
https://ejurnal.pdsi.or.id/index.php/bids/index

log_total_value (r = −0.79), suggesting that months with higher total expenditures tend to be more stable, while smaller-scale procurement periods are more volatile.

A notable relationship is observed between perm_entropy and turn_points (r = 0.73), implying that increases in chaotic complexity coincide with more frequent directional changes in procurement patterns. The volatility–anomaly_score correlation (r = 0.70) demonstrates that fluctuation magnitude is a key determinant of anomaly detection, whereas the negative link between log_total_value and anomaly_score (r = −0.65) indicates that anomalies are more likely to occur in low-value or low-activity months. Together, these results validate the complementarity of chaos-based descriptors and traditional statistical indicators in capturing nonlinear procurement dynamics.

The strong negative correlation between volatility and total procurement value is interpreted as a structural characteristic of the procurement system rather than a modeling artifact. Large-scale procurement activities are typically associated with longer planning horizons, higher administrative scrutiny, and more stable execution processes, which naturally suppress short-term fluctuations. In contrast, smaller-scale procurement tends to be more fragmented and sensitive to administrative timing, resulting in higher relative volatility. Therefore, the pattern observed in Figure 3 reflects real-world procurement behavior rather than bias introduced by the analytical framework.

3.3 Ablation Study (Base vs. Hybrid Model)

To evaluate the contribution of chaos-derived features, an ablation study compared two configurations: a baseline model using only statistical indicators, and a hybrid model combining both statistical and chaotic features.

The Spearman rank correlation between the two anomaly-score sets was $\rho = 0.752$, indicating that while both models identify broadly similar trends, the hybrid configuration introduces unique variance due to nonlinear chaotic behavior. The Jaccard overlap analysis further highlighted this difference. Table 1 summarizes the degree of overlap between the baseline Isolation Forest and the proposed hybrid model across different anomaly cutoffs.

Table 1. Overlap Between Baseline and Hybrid Models

Anomaly Cutoff	Overlap (Overlap Ratio)	Jaccard Similarity
Top 1%	1.00 (complete)	1
Top 5%	0.33	0.20
Top 10%	0.71	0.56

This pattern shows that chaotic descriptors reorder borderline anomalies, providing finer granularity for differentiating near-normal from irregular periods. Beyond the top 1% threshold, divergence increases, signifying enhanced model sensitivity to subtle fluctuations. Figure 4 visualizes the intersection patterns between the baseline Isolation Forest and the hybrid model across different anomaly cutoffs.

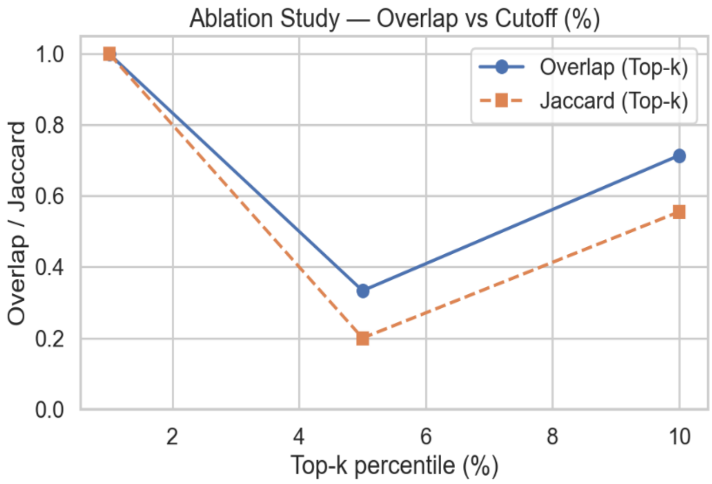

Figure 4. Intersection of Base and Hybrid Models at Various Cutoffs.

At the Top 1% cutoff, both models achieve complete agreement (Overlap = 1.00, Jaccard = 1.0), indicating identical identification of the most extreme outliers. However, the overlap ratio declines to 0.33 (Jaccard = 0.20) at the Top 5% cutoff, highlighting the hybrid model's increased sensitivity to mid-level irregularities introduced by chaotic features.

As the threshold expands to the Top 10%, similarity partially recovers (Overlap = 0.71, Jaccard = 0.56), implying that both models eventually converge on broader anomalous regions while still maintaining the hybrid model's enhanced coverage. The resulting curve (Figure 4) demonstrates that HC-iForest preserves robustness at extremes, diverges moderately at mid levels due to chaos integration, and restabilizes as the anomaly space expands. This behavior supports the framework's adaptive sensitivity to nonlinear perturbations in the data.

Bulletin of Informatics and Data Science
Vol. 4 No. 2, November 2025, Page 114−123
ISSN 2580-8389 (Media Online)
DOI 10.61944/bids.v4i2.137
https://ejurnal.pdsi.or.id/index.php/bids/index

3.4 Statistical Validation

To confirm whether the differences between normal and anomalous periods were statistically significant, Welch's t-tests and Cohen's d were applied across six primary features: permutation entropy, turning points, volatility, log-transformed total value, log-average value, and total tenders. The results (Figure 5) showed that four of the six variables exhibited significant distinctions ($p < 0.05$) between the two groups, with large effect sizes ($|d| > 1.5$) for entropy, volatility, and tender volume.

	feature	t_stat	p_value	cohen_d	mean_anomaly	mean_normal	signif
0	perm_entropy	-3.3189	0.0260	-1.7310	0.1788	0.5762	*
1	turn_points	-3.4320	0.0194	-1.4978	0.1500	0.5104	*
2	volatility	4.2006	0.0124	2.3414	1.0097	0.3298	*
3	log_total_value	-3.2679	0.0290	-1.8442	26.9149	30.2970	*
4	log_avg_value	-1.7490	0.1495	-0.9040	21.4608	21.9147	
5	total_tenders	-5.5283	0.0004	-1.5716	1093.4000	6703.1970	*

Figure 5. Statistical Significance Testing Results

Figure 5 presents two-sample t-test results comparing chaotic and statistical feature values between anomalous and normal months. Four out of six features show statistically significant differences ($p < 0.05$), confirming that they meaningfully discriminate anomalies.

Specifically, perm_entropy ($t = -3.32$, $p = 0.026$) and turn_points ($t = -3.43$, $p = 0.019$) were significantly lower in anomalous periods, implying more deterministic and stable temporal patterns—consistent with artificially controlled procurement flows. Meanwhile, volatility ($t = 4.20$, $p = 0.012$) was substantially higher in anomalous months, suggesting intense fluctuations in tender values. log_total_value ($t = -3.27$, $p = 0.029$) and total_tenders ($t = -5.53$, $p = 0.0004$) were also lower for anomalies, indicating reduced activity during abnormal procurement cycles.

These statistical findings validate the theoretical expectation that anomalies reflect irregular structural and dynamic behavior rather than random deviations. To further illustrate the statistical differences identified above, Figure 6 compares the distribution of chaos-based features between normal and anomalous procurement periods.

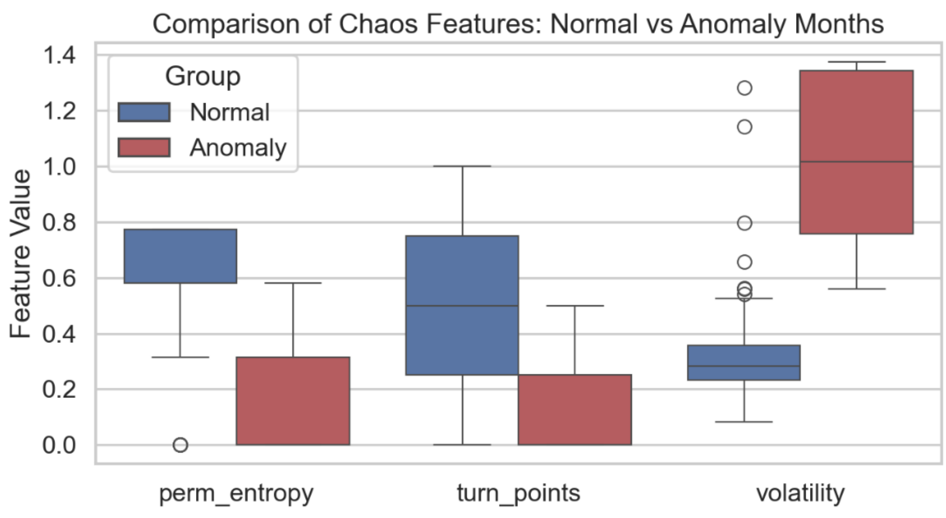

Figure 6. Boxplot Distribution of Chaos Features (Normal versus Anomalous Months)

Figure 6 visualizes the distribution of key chaos-based features (perm_entropy, turn_points, and volatility) between normal and anomalous months. The boxplots clearly show that anomalous periods are characterized by lower entropy and fewer turning points, indicating more predictable yet controlled sequences of tender activity. In contrast, their volatility values are markedly higher, with several extreme outliers exceeding 1.2, reinforcing the notion of unstable procurement dynamics.

Together, these distributions confirm that anomalies in procurement are not purely stochastic events but result from structural distortions that suppress natural complexity while amplifying magnitude variability.

3.5 Principal Anomalies and Procurement Context

The Hybrid Chaos-Isolation Forest (HC-iForest) model identified five principal anomalous months within the 2019–2024 observation period — November 2024, November 2023, January 2024, October 2023, and October 2024. These periods correspond to procurement cycles characterized by sudden increases in transaction volume, atypical contract dispersion, or budget realignment activities.

ISSN 2580-8389 (Media Online)
DOI 10.61944/bids.v4i2.137
https://ejurnal.pdsi.or.id/index.php/bids/index

Table 2. Top Anomalous Months Detected by HC-iForest

month	anomaly_score	total_tenders	avg_value	perm_entropy	turn_points	Volatility	Interpretation
2024-11	0.6734980941993818	12	1743567296.8616667	0.31384521989387315	0.25	1.3742943856420513	End-of-fiscal procurement surge
2023-11	0.6253344746603281	804	2194919834.270149	-5.581602429106793e-13	0.0	1.0165538583942169	Year-end tender acceleration
2024-01	0.6096822354489327	170	5386009966.435764	-5.581602429106793e-13	0.0	1.342474851430798	Post-budget adjustment phase
2023-10	0.6086631067694033	4384	1575387065.6286883	-5.581602429106793e-13	0.0	0.5584876012738845	Pre-year-end fiscal alignment
2024-10	0.6054682495933913	97	1230321049.5659792	0.5802792108501381	0.5	0.7565841851443742	Procurement expansion cycle

Table 2 lists the top five anomalous months detected by the HC-iForest model, providing both quantitative attributes and qualitative interpretations. The highest anomaly scores (0.67–0.61) predominantly occur near the end of fiscal periods—October to November 2023 and January to November 2024—corresponding to typical budget realignment phases in Indonesian public procurement.

These months display extreme volatility (1.02–1.37) coupled with very low permutation entropy, implying concentrated procurement activity and reduced randomness. Some permutation entropy values in Table 2 appear as very small negative numbers (e.g., -5.58×10^{-13}), which result from floating-point precision limitations. These values effectively represent zero entropy and indicate highly regular or near-monotonic procurement sequences. For example, November 2024 records only 12 tenders yet exhibits the highest volatility, signaling a disproportionate allocation of high-value contracts. Similar dynamics appear in early 2024 and late 2023, suggesting recurring fiscal behaviors at budget boundaries.

These temporal patterns demonstrate the hybrid model's ability to reveal systemic irregularities driven by institutional cycles rather than random market noise, supporting its potential use for fiscal anomaly surveillance.

3.6 Interpretation and Implications

The experimental findings collectively affirm that the Hybrid Chaos-Isolation Forest (HC-iForest) framework enhances the detection of procurement irregularities by integrating nonlinear dynamic descriptors into the anomaly detection pipeline. Across all empirical analyses—feature correlations, ablation tests, statistical comparisons, and temporal localization—chaos-based indicators (permutation entropy, turning points, and volatility) consistently emerged as discriminative signals that differentiate anomalous procurement behavior from normal operations.

The results highlight two distinct behavioral patterns associated with anomalies. First, anomalous months exhibit lower temporal complexity (reduced entropy and fewer turning points), implying that tender activities are more deterministic and potentially subject to administrative coordination rather than organic market behavior. Second, these same periods display higher volatility, suggesting abrupt shifts in transaction magnitudes consistent with concentrated fiscal spending or reallocation of budgeted funds. Together, these patterns characterize anomalies not as random statistical outliers but as structural distortions in procurement dynamics.

From a governance perspective, the concentration of anomalies around fiscal boundaries—particularly between October and January—aligns with long-observed procurement surges in Indonesia's public sector. Such timing often reflects accelerated contract awards to exhaust remaining budgets or post-budget realignments that create spending spikes. Detecting these behaviors through automated hybrid models provides a quantitative early-warning mechanism for oversight bodies such as the National Public Procurement Agency (LKPP) and Indonesia Corruption Watch (ICW).

By systematically capturing chaotic signatures of irregular activity, the HC-iForest approach strengthens the analytical dimension of open contracting data initiatives. The framework operationalizes transparency by converting openly published procurement records into actionable anomaly signals, supporting data-driven auditing and evidence-based policymaking. Moreover, this method complements conventional investigative workflows by prioritizing high-risk periods or agencies for further examination.

In practical terms, the integration of chaos theory with machine-learning anomaly detection can improve the resilience of fiscal monitoring systems. Its capacity to model nonlinear temporal dependencies makes it well-suited for detecting both gradual drifts and abrupt manipulations that traditional statistical tools often overlook. The approach contributes not only to technical innovation but also to broader institutional integrity goals, reinforcing transparency, accountability, and efficiency in Indonesia's public procurement ecosystem. Additionally, cross-figure consistency

Vol. 4 No. 2, November 2025, Page 114−123
ISSN 2580-8389 (Media Online)
DOI 10.61944/bids.v4i2.137
https://ejurnal.pdsi.or.id/index.php/bids/index

checks confirmed that numerical values (e.g., thresholds, scores, and correlations) were reproducible across independent runs, supporting the model's computational robustness and reliability for replication.

4. CONCLUSION

This study proposed and empirically evaluated a Hybrid Chaos–Isolation Forest (HC-iForest) framework for anomaly detection in Indonesia's public procurement datasets. By incorporating nonlinear dynamic descriptors—namely, permutation entropy, turning points, and volatility—the model successfully captured latent temporal irregularities often undetected by conventional methods. Empirical findings demonstrated that the hybrid framework enhances both model stability and interpretability. Correlation and ablation analyses revealed that chaos-based descriptors provide distinctive, non-redundant information beyond traditional statistical metrics, thereby improving model sensitivity to complex fluctuations in tender behavior. Quantitatively, this stability is reflected in a Spearman rank correlation of $\rho = 0.75$ between the baseline and hybrid anomaly rankings, indicating consistent identification of extreme anomalies while allowing meaningful reordering at intermediate levels. Statistical significance testing confirmed that these features differed significantly between anomalous and normal months, while temporal localization indicated that detected anomalies clustered around fiscal transitions—particularly year-end budget adjustments and expenditure surges. These findings indicate that anomalies in procurement are not random deviations but systemic manifestations of institutional and fiscal behaviors—often linked to accelerated or reallocated government expenditures. Accordingly, the HC-iForest framework offers a robust analytical tool for identifying structural distortions within open contracting datasets. From a policy perspective, the proposed model can function as a data-driven early-warning system for agencies such as LKPP and ICW, facilitating targeted monitoring and proactive governance interventions. Moreover, its interpretability supports integration with broader open-data and transparency initiatives, thereby reinforcing Indonesia's commitment to accountable public spending. It should be noted that the present analysis relies on monthly aggregated procurement data, making the framework particularly suitable for strategic-level oversight. Future research may explore finer temporal resolutions, such as weekly or per-tender analyses, to further assess sensitivity at operational levels. Future research may extend this framework using hybrid ensemble architectures (e.g., Chaos-Based Autoencoders or Neural Isolation Forests) and spatiotemporal modeling to capture region-specific procurement dynamics. In sum, integrating chaos theory with advanced machine-learning techniques opens a promising avenue for developing more intelligent, adaptive, and transparent fiscal oversight systems.

REFERENCES

[1] D. P. Senopati Warga Dalam, W. Prawesthi, and S. Marwiyah, "Prevention of Corruption in the Procurement of Goods/Services Process at PT PAL Indonesia Based on Regulations in Indonesia," *American Journal of Humanities and Social Sciences Research*, vol. 9, no. 9, pp. 99–106, 2025

[2] I. Suardi, H. Rossieta, C. Djakman, and V. Diyanty, "Procurement governance in reducing corruption in the indonesian public sector: a mixed method approach," *Cogent Business and Management*, vol. 11, no. 1, p. 2393744, 2024, doi: 10.1080/23311975.2024.2393744.

[3] R. Anggriawan, "From Collusion to Corruption: How Indonesian Law Fights Back in Procurement Conspiracy," *Jurnal Penegakan Hukum dan Keadilan*, vol. 6, no. 1, pp. 66–81, Mar. 2025, doi: 10.18196/jphk.v6i1.24577.

[4] A. Firmansyah, R. Y. Maulana, and A. Z. Miftah, "Transformation of the Procurement System in the Indonesian Government," *Sosiohumaniora*, vol. 26, no. 2, pp. 369–381, Dec. 2024, doi: 10.24198/sosiohumaniora.v26i2.56209.

[5] X. Tan, J. Yang, and S. Rahardja, "Sparse random projection isolation forest for outlier detection," *Pattern Recognit Lett*, vol. 163, pp. 65–73, 2022, doi: https://doi.org/10.1016/j.patrec.2022.09.015.

[6] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation-Based Anomaly Detection," *ACM Trans Knowl Discov Data*, vol. 6, no. 1, pp. 1–39, Mar. 2012, doi: 10.1145/2133360.2133363.

[7] C. Zheng, Y. Chen, and X. Du, "A robust soft voting ensemble of the isolation forest model, extended isolation forest model and generalized isolation forest model for multivariate geochemical anomaly recognition," *Ore Geol Rev*, vol. 185, p. 106787, 2025, doi: https://doi.org/10.1016/j.oregeorev.2025.106787.

[8] X. Chen, T. Weng, C. Li, and H. Yang, "Equivalence of machine learning models in modeling chaos," *Chaos Solitons Fractals*, vol. 165, p. 112831, 2022, doi: https://doi.org/10.1016/j.chaos.2022.112831.

[9] S. Sheng and X. Wang, "Network traffic anomaly detection method based on chaotic neural network," *Alexandria Engineering Journal*, vol. 77, pp. 567–579, 2023, doi: https://doi.org/10.1016/j.aej.2023.07.019.

[10] W. S. Al Farizi, I. Hidayah, and M. N. Rizal, "Isolation Forest Based Anomaly Detection: A Systematic Literature Review," in *2021 8th International Conference on Information Technology, Computer and Electrical Engineering (ICITACEE)*, 2021, pp. 118–122. doi: 10.1109/ICITACEE53184.2021.9617498.

[11] G. Wang, D. Wei, X. Li, and N. Wang, "A novel method for local anomaly detection of time series based on multi entropy fusion," *Physica A: Statistical Mechanics and its Applications*, vol. 615, p. 128593, 2023, doi: https://doi.org/10.1016/j.physa.2023.128593.

[12] C. Bandt and B. Pompe, "Permutation Entropy: A Natural Complexity Measure for Time Series," *Phys Rev Lett*, vol. 88, no. 17, p. 174102, Apr. 2002, doi: 10.1103/PhysRevLett.88.174102.

[13] J. S. Richman and J. R. Moorman, “Physiological time-series analysis using approximate entropy and sample entropy,” *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 278, no. 6, pp. H2039–H2049, Jun. 2000, doi: 10.1152/ajpheart.2000.278.6.H2039.

[14] K. Blagov, C. E. Muñiz-Cuza, and M. Boehm, “Fast, Parameter-free Time Series Anomaly Detection,” in *Datenbanksysteme für Business, Technologie und Web (BTW 2025)*, Gesellschaft für Informatik, Bonn, 2025, pp. 453–474. doi: 10.18420/BTW2025-21.

[15] P. Wiessner, G. Bezirganyan, S. Sellami, R. Chbeir, and H.-J. Bungartz, “Uncertainty-Aware Time Series Anomaly Detection,” *Future Internet*, vol. 16, no. 11, p. 403, Oct. 2024, doi: 10.3390/fi16110403.

[16] Y. Feng, Y. Zhang, and Y. Wang, “Out-of-sample volatility prediction: Rolling window, expanding window, or both?,” *J Forecast*, vol. 43, no. 3, pp. 567–582, Apr. 2024, doi: 10.1002/for.3046.

[17] J. Olbryś, “Entropy of Volatility Changes: Novel Method for Assessment of Regularity in Volatility Time Series,” *Entropy*, vol. 27, no. 3, p. 318, Mar. 2025, doi: 10.3390/e27030318.

[18] H. Xu, G. Pang, Y. Wang, and Y. Wang, “Deep Isolation Forest for Anomaly Detection,” *IEEE Trans Knowl Data Eng*, vol. 35, no. 12, pp. 12591–12604, 2023, doi: 10.1109/TKDE.2023.3270293.

[19] A. Arcudi, D. Frizzo, C. Masiero, and G. A. Susto, “Enhancing interpretability and generalizability in extended isolation forests,” *Eng Appl Artif Intell*, vol. 138, Dec. 2024, doi: 10.1016/j.engappai.2024.109409.

[20] H. Xiang *et al.*, “OptIForest: Optimal Isolation Forest for Anomaly Detection,” in *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, California: International Joint Conferences on Artificial Intelligence Organization, Aug. 2023, pp. 2379–2387. doi: 10.24963/ijcai.2023/264.

[21] A. Barredo Arrieta *et al.*, “Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI,” *Information Fusion*, vol. 58, pp. 82–115, 2020, doi: https://doi.org/10.1016/j.inffus.2019.12.012.

