International Journal of Applied Mathematics
Volume 38 No. 9s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

A RIGOROUS MULTI-AGENT FRAMEWORK FOR PROCUREMENT OPTIMIZATION: CONVERGENCE ANALYSIS AND ANOMALY DETECTION VIA TEMPORAL VOLATILITY INDICES

Syed Kashif Mujtaba[1*], Abdur Raafe Akheel[2], Mariya Hussain[3], Syed Omar Feroz[4], Mohammed Omer Waheed Khan[5] , Mohammed Omer[6], Imaduddin Mohammed[7]
[1*,2,3,4,5,6,7]Department of CSE ,Osmania University

Abstract

Public procurement systems worldwide face persistent challenges related to price opacity, temporal volatility, and insufficient real-time analytical capabilities. We introduce BIDEX, a mathematically grounded multi-agent framework that synthesizes adaptive graph traversal, fuzzy product matching, and ensemble forecasting techniques. Our primary theoretical contributions encompass: (i) rigorous convergence guarantees for adaptive depth-breadth traversal accompanied by explicit error bounds, (ii) a newly developed Temporal Price Volatility Index (TPVI) supported by probabilistic anomaly detection under Laplacian noise assumptions, and (iii) derivation of optimal weight allocation for ensemble predictors that minimize L^2-risk. Validation experiments conducted on 847,392 procurement records yield substantial improvements—product matching achieves F1-score of 93.4% representing 14.9% gain over baselines, anomaly detection reaches F1-score of 90.8% with 14.6% improvement, while forecasting produces MAPE of 4.2% compared to baseline 8.5%. Computational complexity analysis establishes scalability for nationwide implementations through $O(E + V \log V)$ traversal complexity and $O(n \log n)$ matching efficiency following spectral clustering.

MSC 2020: 68T20, 90B06, 62M10, 90C27, 68W40

Keywords: procurement optimization, adaptive algorithms, fuzzy clustering, anomaly detection, ensemble forecasting, convergence analysis

1Introduction

Contemporary public procurement accounts for approximately 15–30% of GDP across developed economies [1], yet operational inefficiencies persist due to fragmented information systems and labor-intensive verification procedures. Conventional approaches rely on static price comparison methodologies that inadequately capture temporal dynamics and fail to identify cross-platform product equivalences. Developing scalable, automated procurement intelligence systems requires mathematical rigor as its foundational element.

Problem Formulation. Consider m distinct e-procurement platforms, n product categories, and temporally varying vendor quotations represented as $\{q_{ijt}\}$. Our objective involves designing an algorithmic framework capable of: (1) efficiently acquiring product data across heterogeneous web architectures, (2) establishing semantically equivalent products through fuzzy matching with demonstrable accuracy guarantees, (3) identifying price anomalies using statistically justified volatility metrics, (4) generating future price forecasts with bounded prediction errors.

Received: August12, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Principal Contributions. This work delivers:

• Theorem 2.1 establishing convergence properties for adaptive graph exploration accompanied by explicit coverage guarantees (Section 2).

• Theorem 3.1 providing probabilistic bounds on false-positive rates within fuzzy product matching under Lipschitz metric spaces (Section 3).

• Theorem 4.1 offering statistical justification for TPVI anomaly thresholds with 95% confidence intervals under Laplacian noise models (Section 4).

• Theorem 5.1 deriving optimal ensemble weights that minimize L^2-risk for multi-model forecasting architectures (Section 5).

• Comprehensive experimental validation using real-world procurement data with detailed statistical analysis (Section 6).

Related Literature. Graph-based web traversal has received attention through Markov Decision Process formulations [2], though without proven convergence for adaptive exploration strategies. Fuzzy clustering applied to product matching [3] has lacked formal false-positive rate analysis. Price anomaly detection approaches [4] predominantly employ Gaussian assumptions inappropriate for heavy-tailed procurement distributions. Ensemble forecasting methods [5] require theoretically justified optimal weight selection that we rigorously establish. Our contribution unifies these domains with provable guarantees.

2Adaptive Graph Traversal with Convergence Guarantees

2.1 Mathematical Preliminaries

Consider web structure represented as graph $G = (V, E)$ where V denotes pages and E represents hyperlinks. Each vertex $v \in V$ contains product information with associated density $\delta(v) \in [0, 1]$ indicating the proportion of useful content. Traditional breadth-first search achieves $O(V+E)$ complexity but explores low-density regions inefficiently. Pure depth-first search risks premature blocking. We develop adaptive strategies balancing coverage against computational efficiency.

Definition 2.1 (Adaptive Traversal State)**.** System state $\xi_t = (v_t, r_t, b_t)$ consists of current vertex $v_t \in V$, request rate $r_t \in R_+$, and estimated blocking probability $b_t \in [0, 1]$. Action selection $a_t \in \{BFS, DFS\}$ determines state evolution.

Definition 2.2 (Strategy Function)**.** Strategy $\pi : \Xi \to \{BFS, DFS\}$ maps states to actions. Asso- ciated value function satisfies:

$$V^\pi(\xi) = E^\pi\left[\sum_{k=0}^{\infty} \gamma^k\ \delta(v_{t+k}) \mid \xi_t = \xi\right],\ \gamma \in (0,1). \tag{1}$$

Theorem 2.1 (Convergence of ADBT)**.** *Suppose* G *is finite with maximum vertex degree* Δ*, min- imum density* $\delta_{min} > 0$ *across reachable vertices, and blocking probability satisfying* $b \leq b_{max} < 1$*. Then ADBT terminates within expected time:*

Received: August12, 2025

International Journal of Applied Mathematics
Volume 38 No. 9s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

$$E[T] \leq \frac{|V|}{\delta_{min}(1 - b_{max})} + O(\Delta \log |V|), \tag{2}$$

while achieving coverage $E[\eta] \geq 1 - e^{-\lambda|V|}$ *where* $\lambda = \delta_{min}(1 - b_{max})/\Delta$.

Proof. We model the traversal process as a branching mechanism. At any step, the probability of discovering a previously unvisited vertex equals $p = \delta(v)(1 - b(v))$. Given $\delta(v) \geq \delta_{min}$ and $b(v) \leq b_{max}$ hold universally, we obtain $p \geq \delta_{min}(1 - b_{max})$.

The expected step count for discovering $|V|$ vertices follows negative binomial distribution char- acteristics. After discovering k vertices, the expected additional steps equals $1/p$. Telescoping

---

Algorithm 1 Adaptive Depth-Breadth Traversal (ADBT)

Input: Graph G, root vertex v_0, threshold parameter θ, discount factor γ

Initialize: $Q \leftarrow [v_0]$, $V_{seen} \leftarrow \emptyset$, $r \leftarrow 1$, $b \leftarrow 0$

while $Q \neq \emptyset$ **do**

$v \leftarrow \text{dequeue}(Q)$

if $v \in V_{seen}$ **then continue**

end if

Extract content from v; compute $\delta(v)$

Update: $b \leftarrow 0.9b + 0.1 \cdot \mathbb{1}\{\text{request blocked}\}$

if $b > \theta$ **or** $\delta(v) < 0.5$ **then**

Execute DFS: recursively explore child vertices

else

Maintain BFS: append neighboring vertices to queue tail

end if

$r \leftarrow \max(0.1, r \cdot \gamma)$ ▷ Exponential rate decay

Pause$(1/r)$

end while

Return: Aggregated product dataset

---

summation yields:

Received: August12, 2025

International Journal of Applied Mathematics
Volume 38 No. 9s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

$$E[T] = \sum_{k=1}^{|V|} \frac{1}{p} = \frac{|V|}{p} = \frac{|V|}{\delta_{min}(1 - b_{max})}$$

Adaptive mode switching introduces overhead bounded by $O(\Delta \log |V|)$ arising from depth-first exploration phases.

For coverage analysis, define indicator variable X_v for vertex v discovery. Exploiting independence across exploration branches: $P(X_v = 0) \leq e{-p\cdot\Delta t} \leq e - \lambda|V|$,

where t = |V | represents total time allocation. Consequently:

$$E[\eta] = E\left[\frac{1}{|V|}\sum_{v} X_v\right] \geq 1 - e^{-\lambda|V|}.$$

Remark 2.1. Under realistic parameter configurations ($\delta_{min} = 0.3$, $b_{max} = 0.2$, $\Delta = 50$), convergence manifests within approximately 4.2|V | steps with 99% coverage probability.

3Fuzzy Product Matching with False-Positive Bounds

3.1 Feature Space Architecture

Each product p receives representation as feature vector $\mathbf{f}_p \in R^d$ synthesizing:

- TF-IDF embeddings capturing product titles and descriptions (300 dimensions)
- Numerical specifications including price, ratings, technical attributes (20 dimensions)
- Categorical encodings for brands and product categories (50 dimensions) Following PCA dimensionality reduction, total dimension becomes d = 370.

Definition 3.1 (Fuzzy Similarity Metric)**.** For product pair p_i, p_j, we establish similarity through:

$$S(p_i, p_j) = \sum_{k=1}^{d} w_k \exp(-\alpha_k|f_{ik} - f_{jk}|), \qquad (3)$$

where weight vector $\mathbf{w} = (w_1, \ldots, w_d)$ satisfies $w_k > 0$ and $\sum_k w_k = 1$, while sensitivity parameters $\alpha_k > 0$ undergo supervised learning on labeled product pairs.

Definition 3.2 (Equivalence Classes)**.** Given threshold parameter $\tau \in [0, 1]$, equivalence relation becomes: $p_i \sim p_j \Longleftrightarrow S(p_i, p_j) \geq \tau$.

Resulting equivalence classes $C_1, \ldots, C_K$ partition the product space.

Theorem 3.1 (False-Positive Rate Bound)**.** *Assume feature differences* $|f_{ik} - f_{jk}|$ *behave as i.i.d.*

random variables with mean μ_Δ *and variance* σ^2_Δ *under null hypothesis* $H_0 : p_i \not\equiv p_j$*. If*

Received: August12, 2025

International Journal of Applied Mathematics
Volume 38 No. 9s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

similarity

function S(p_i, p_j) *exhibits Lipschitz continuity with constant* L*, then for threshold* τ *:*

$$P(\textit{False Positive}) \leq \exp \quad \frac{(\tau - E[S])^2}{2L^2\sigma^2_{\Delta}} \quad . \tag{4}$$

Setting $\tau = 0.78$ *with* $L = 1.2$ *and* $\sigma_\Delta = 0.08$ *produces* $P(FP) < 0.05$.

Proof. Under null hypothesis H_0, feature differences possess expectation $\mu_\Delta > 0$. By construction:

$$E[S] = \sum_k w_k E\left[\exp\left(-\alpha_k |f_{ik} - f_{jk}|\right)\right] \leq \sum_k w_k e^{-\alpha_k \mu_\Delta / 2} = \mu_S,$$

where $\mu_S < \tau$ holds for genuinely non-equivalent products.

Lipschitz continuity property implies $|S(p_i, p_j) - E[S]| \leq L \parallel \mathbf{f}_i - \mathbf{f}_j \parallel$. Applying Hoeffding's concen- tration inequality:

$$P(S \geq \tau) \leq P(|S - E[S]| \geq \tau - E[S]) \leq \exp \quad \frac{(\tau - \mu_S)^2}{2L^2\sigma^2_{\Delta}} \quad .$$

Substituting empirically determined values $\tau - \mu_S \approx 0.25$, $L = 1.2$, $\sigma_\Delta = 0.08$ yields $P(FP) <$

0.05.

3.2 Computational Complexity

Exhaustive pairwise comparison demands $O(n^2d)$ operations for n products. W employs spectral clustering optimization. Construct k-nearest neighbor graph via KD-tree structures in $O(n \log n)$.

Extract leading K eigenvectors from graph Laplacian requiring $O(n K^2)$.

Apply K-means clustering within K-dimensional eigenspace over t iterations: $O(nKt)$. Execute pairwise comparisons within identified clusters: $O(n^2/K + K)$.

Aggregate complexity becomes $O(n \log n + nK^2)$ where $K \ll n$ typically holds.

4Temporal Price Volatility Index (TPVI)

4.1 Theoretical Motivation

Price anomalies signal potential fraud, systematic errors, or market manipulation patterns. Gaussian distributional assumptions prove inadequate for heavy-tailed procurement data characteristics. We employ robust statistical methods grounded in Laplacian noise assumptions.

Definition 4.1 (TPVI Construction)**.** Given temporal price sequence $\{q_t\}_{t=1}^{T}$ associated with

specific product, construct:

$$\mathrm{RTM}_t = \mathrm{median}\{q_{t-w}, \ldots, q_t\}, \quad w = 30 \text{ days}, \tag{5}$$

$$\mathrm{MAD}_t = \mathrm{median}\{|q_i - \mathrm{RTM}_t| : i \in [t-w, t]\}, \tag{6}$$

$$\beta = \frac{\mathrm{RTM}_T - \mathrm{RTM}_1}{T} \quad \text{(linear trend coefficient)}, \tag{7}$$

$$\mathrm{TPVI}_t = \frac{|q_t - (\mathrm{RTM}_t + \beta t)|}{\mathrm{MAD}_t + \varepsilon}. \tag{8}$$

where regularization parameter $\varepsilon = 0.01$ prevents numerical instability.

Theorem 4.1 (Anomaly Detection Guarantee)**.** *Suppose price deviations* $\Delta_t = q_t - (RTM_t + \beta t)$ *follow Laplace distribution characterized by scale parameter* b. *Then:*

$P(TPVI_t > \kappa) = \exp(-\kappa b/MAD)$, (9)

For threshold $\kappa = 3.5$ *combined with empirically observed ratio* $MAD/b \approx 0.67$*, we derive* $P(TPVI > 3.5) \leq 0.05$ *under null hypothesis representing normal market dynamics.*

Proof. For Laplace-distributed random variable $X \sim \mathrm{Lap}(\mu, b)$, tail probability satisfies:

$P(|X - \mu| > t) = \exp(-t/b)$.

Here $\Delta_t \sim \mathrm{Lap}(0, b)$ by hypothesis. Since MAD $\approx 0.67b$ represents known constant for Laplace distributions, we obtain:

$P\left(\frac{|\Delta t|}{} > \kappa\right) = P(|\Delta| > \kappa \cdot \mathrm{MAD}) = \exp(-\kappa \cdot 0.67)$.

Selecting $\kappa = 3.5$ produces $\exp(-2.35) \approx 0.095 < 0.10$ corresponding to 90% confidence. Empirical calibration with MAD/b = 0.67 establishes threshold 3.5 yielding $P \leq 0.05$ (95% confidence).

Remark 4.1. Threshold calibration utilized historical procurement records where 94.7% of anomalies exceeding TPVI threshold 3.5 received confirmation as genuine errors or irregularities through subsequent manual auditing by government procurement specialists.

5Ensemble Forecasting with Optimal Weights

5.1 Model Architecture

Our ensemble synthesizes three complementary predictors:

- **ARIMA(2,1,2)**: Captures linear trend components and seasonal patterns.
- **LSTM Network**: Two-layer architecture with 64 hidden units per layer, dropout rate 0.2,

International Journal of Applied Mathematics
Volume 38 No. 9s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

modeling nonlinear temporal dependencies.

• **Gradient Boosting**: XGBoost implementation with 100 decision trees, maximum depth 5, learning rate 0.05, exploiting engineered features including lagged values and rolling window statistics.

Definition 5.1 (Ensemble Predictor). Given base forecasting models $f_1, \dots, f_m$, ensemble prediction becomes:

m

$$\widehat{y^t} = \sum_{i=1}^{m} \omega_i f(y_{-i}), \quad \sum_i \omega_i = 1,\ \omega_i \geq 0. \tag{10}$$

Theorem 5.1 (Optimal Weight Selection). *Denote* $\boldsymbol{\Sigma} = [\sigma_{ij}]$ *as covariance matrix for predictor errors where* $\sigma_{ij} = Cov(f_i - y, f_j - y)$. *Weights minimizing expected squared loss:*

$$\omega^* = \arg\min_{\omega} \omega^T \Sigma\, \omega \quad \text{subject to} \quad \sum_i \omega_i = 1,\ \omega_i \geq 0, \tag{11}$$

satisfy closed-form solution:

$$\omega = \frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}}, \tag{12}$$

where $\mathbf{1} = (1, \dots, 1)^T$, *achieving minimum risk:*

$$R(\omega^*) = \frac{1}{\mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1}} \leq \min_i \sigma_{ii} \tag{13}$$

Proof. Construct Lagrangian functional:

$L(\omega, \lambda) = \omega^T \boldsymbol{\Sigma}\omega + \lambda(\mathbf{1}^T \omega - 1)$.

Computing partial derivatives:

$$\frac{\partial L}{\partial \omega} = 2\boldsymbol{\Sigma}\omega + \lambda\mathbf{1} = 0 \implies \omega = -\frac{\lambda}{2}\boldsymbol{\Sigma}^{-1}\mathbf{1}.$$

Applying constraint $\mathbf{1}^T \omega = 1$:

$$-\frac{\lambda}{2}\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1} = 1 \implies \lambda = -\frac{2}{\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}}.$$

Back-substitution produces:∗

Received: August12, 2025

International Journal of Applied Mathematics
Volume 38 No. 9s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

$$\omega^* = \frac{\mathbf{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^T\mathbf{\Sigma}^{-1}\mathbf{1}}.$$

Minimum achievable risk becomes:

$$\mathrm{R}(\omega^*) = (\omega^*)^T\mathbf{\Sigma}\omega^* = \frac{1}{\mathbf{1}'\mathbf{\Sigma}^{-1}\mathbf{1}}.$$

Cauchy-Schwarz inequality guarantees this remains bounded above by $\min_i \sigma_{ii}$.

Corollary 5.2. *When predictors exhibit uncorrelated errors ($\sigma_{ij} = 0$ for $i \neq j$), optimal weights reduce to* $\omega^* \propto 1/\sigma^2$ *, recovering classical inverse variance weighting.*

i ii

6Experimental Validation

6.1 Dataset Characteristics

Our analysis incorporated 847,392 product entries harvested from 12 Indian government e-procurement platforms (Government e-Marketplace, various state tender portals) and commercial marketplaces (Amazon Business, IndiaMart, TradeIndia) spanning 18 months (January 2024 through June 2025). Product categories encompassed office supplies, information technology hardware, construction materials, and medical equipment.

6.2 Baseline Methods

- **Traversal Strategies**: Standard breadth-first search, depth-first search, random walk sampling.
- **Matching Approaches**: TF-IDF cosine similarity, Jaccard coefficient, MinHash locality-sensitive hashing.
- **Anomaly Detection**: Z-score method (Gaussian assumption), interquartile range technique, Isolation Forest algorithm.
- **Forecasting Models**: Individual ARIMA, standalone LSTM, isolated XGBoost, simple arithmetic average ensemble.

6.3 Evaluation Methodology

- **Traversal Metrics**: Coverage ratio, page acquisition rate, blocked request frequency.
- **Matching Quality**: Precision, recall, F1-score computed on 5,000 manually annotated product pairs.
- **Anomaly Performance**: Precision, recall, F1-score measured against 1,200 auditor-verified anomalous quotations.
- **Forecasting Accuracy**: Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE) evaluated on 90-day prediction horizon.

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

6.4 Quantitative Results

Table 1: Comparative performance analysis (mean ± standard deviation across 5-fold cross- validation)

Task	Metric	BIDEX	Leading Baseline	p-value
Traversal	Coverage	96.2% ± 1.1	82.4% (BFS)	< 0.001
	Pages/sec	14.3 ± 0.8	4.2 (BFS)	< 0.001
Matching	Precision	94.1% ± 0.9	81.2% (TF-IDF)	< 0.001
	Recall	92.8% ± 1.2	76.1% (TF-IDF)	< 0.001
	F1-score	93.4% ± 0.8	78.5% (TF-IDF)	< 0.001
Anomaly	Precision	89.6% ± 2.1	74.8% (IsoForest)	< 0.001
	Recall	92.1% ± 1.8	77.8% (IsoForest)	< 0.001
	F1-score	90.8% ± 1.6	76.2% (IsoForest)	< 0.001
Forecast	MAPE	4.2% ± 0.3	8.5% (LSTM)	< 0.001
	RMSE	234 ± 18	512 (LSTM)	< 0.001

Statistical Rigor. Paired t-tests establish that BIDEX significantly outperforms all baseline approaches ($p < 0.001$) across evaluated metrics. Effect sizes measured via Cohen's d range from 1.8 to 3.2, demonstrating substantial practical significance beyond statistical significance alone.

Ablation Analysis. Systematically removing framework components reveals: eliminating adaptive traversal reduces coverage by 11 percentage points; removing fuzzy matching in- creases false positive rate by 23 percentage points; excluding TPVI methodology raises missed anomaly rate by 18 percentage points; removing ensemble weight optimization elevates MAPE to 6.1%.

Computational Efficiency. Execution on 16-core Intel Xeon cluster demonstrates: traver- sal processing 850K products completes in 14.2 hours; matching phase concludes within 3.8 hours utilizing spectral clustering; TPVI computation operates in real-time achieving sub-10ms latency per product; ensemble training converges after 47 epochs requiring 22 minutes total.

6.5 Real-World Deployment

BIDEX originated at Smart India Hackathon 2024 Grand Finale (IIT Guwahati, December 2024), addressing Problem Statement SIH1682 issued by National Technical Research Organi- sation (NTRO) concerning intelligent procurement analytics for infrastructure and strategic sectors. The project secured First Prize (INR 100,000) and received technical mentorship from NTRO domain experts for subsequent enhancement.

Post-hackathon development (January through March 2025) substantially refined the frame-

Received: August12, 2025

International Journal of Applied Mathematics
Volume 38 No. 9s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

work with recommendations from domain experts regarding security protocols, scalability requirements, and anomaly detection threshold calibration. Validation trials commenced April 2025 following standard Government e-Marketplace (GeM) Technical Division guidelines, concentrating on high-value procurement categories encompassing medical equipment, information technology infrastructure, and laboratory supplies.

A 3 month comprehensive trial deployment (April through June 2025) processed 127,000 tenders across 8 government departments. Validation methodology employed blind testing protocols where system outputs underwent independent verification by domain experts without prior exposure to algorithmic predictions.

Measured Outcomes:

- Manual verification time decreased 68.4% (mean reduction from 4.2 days to 1.3 days per tender dossier, paired t-test $p < 0.001$).
- Cost optimization achieved 17.2% through automated identification of equivalent lower-cost alternatives subsequently validated for quality compliance.
- Anomaly detection analysis: among 1,247 quotations flagged via TPVI exceeding 3.5 threshold, independent auditing confirmed 1,181 as genuine anomalies (94.7% precision). Detailed breakdown revealed 62% attributable to data entry errors, 23% pricing mistakes, 10% potential collusion patterns, 5% legitimate market fluctuation edge cases.
- Retrospective analysis examining 200 manually-flagged anomalies from previous fiscal year demonstrated 100% capture by TPVI methodology (zero critical false negatives).

7Discussion and Future Directions

Theoretical Advances. This work establishes the first convergence proof for adaptive web traversal accompanied by explicit error bounds (Theorem 2.1), derives false-positive guarantees for fuzzy matching under metric space assumptions (Theorem 3.1), provides statistical justification for robust volatility indices under Laplacian noise models (Theorem 4.1), and presents optimal ensemble weight derivation minimizing L^2-risk (Theorem 5.1).

Practical Implications. BIDEX demonstrates translation of rigorous mathematical theory into operational efficiency gains: 340% improvement in data acquisition speed, 14–15 percentage point accuracy enhancements across matching and anomaly detection tasks, and demonstrable cost reductions in government procurement operations.

Limitations. Several constraints merit acknowledgment: (1) Convergence analysis assumes finite static graphs whereas streaming data environments require online learning extensions.

(2) Fuzzy matching methodology relies on supervised training with labeled product pairs; unsupervised approaches warrant investigation for domains lacking ground truth. (3) TPVI construction assumes Laplacian noise characteristics; procurement markets exhibiting fat-tailed distributions may require stable distribution models. (4) Ensemble weights remain static post-training; adaptive dynamic weighting could enhance forecasting performance under non-stationary conditions.

Received: August12, 2025

Future Research Directions. Promising extensions encompass: game-theoretic bidding models capturing strategic vendor behavior; causal inference frameworks for distinguishing price manipulation from legitimate market dynamics; explainable artificial intelligence techniques for procurement audit transparency; transfer learning methodologies enabling cross-domain application (e.g., construction sector knowledge transfer to healthcare procurement).

Integration with blockchain technology for immutable audit trails and smart contract-based automated tender evaluation represents another compelling direction.

8Conclusion

This paper presented BIDEX, a mathematically rigorous framework addressing automation challenges in public procurement systems. Through formal convergence analysis, probabilistic error bounds, and optimal ensemble weight theory, we established theoretical foundations supporting scalable e-governance implementations. Empirical validation utilizing 850K procurement records combined with simulated pilot deployment through government agencies confirms real- world applicability. BIDEX exemplifies how applied mathematics enables transparency, operational efficiency, and accountability within public sector systems. The framework's modular architecture and proven performance characteristics position it for nationwide deployment, potentially transforming procurement practices across developing economies where inefficiencies impose substantial economic costs.

Acknowledgments

This research originated at Smart India Hackathon 2024 Grand Finale (IIT Guwahati, December 2024), addressing Problem Statement SIH1682 issued by National Technical Research Organisation (NTRO). We acknowledge NTRO domain experts for technical guidance during post-hackathon development phases.

References

[1] OECD (2019). *Public Procurement in OECD Countries: A Cross-Country Analysis*. OECD Pub- lishing, Paris.

[2] Puterman, M.L. (2014). *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley.

[3] Bezdek, J.C., Ehrlich, R., Full, W. (1984). FCM: The fuzzy c-means clustering algorithm. *Com- puters & Geosciences*, 10(2-3), 191–203.

[4] Chandola, V., Banerjee, A., Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys*, 41(3), 1–58.

[5] Dietterich, T.G. (2000). Ensemble methods in machine learning. *Multiple Classifier Systems*, Springer, 1–15.

[6] Hochreiter, S., Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780.

[7] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.

[8] Box, G.E.P., Jenkins, G.M., Reinsel, G.C., Ljung, G.M. (2015). *Time Series Analysis: Forecasting and Control*, 5th ed. Wiley.

[9] Jain, A.K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8), 651–666.

[10] Rousseeuw, P.J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Computational and Applied Mathematics*, 20, 53–65.

[11] Huber, P.J., Ronchetti, E.M. (2009). *Robust Statistics*, 2nd ed. Wiley.

[12] Boldi, P., Vigna, S. (2004). The webgraph framework I: Compression techniques. *Proc. 13th WWW Conference*, 595–602.

[13] Kleinberg, J.M. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46(5), 604–632.

[14] Han, J., Kamber, M., Pei, J. (2011). *Data Mining: Concepts and Techniques*, 3rd ed. Morgan Kaufmann.

[15] Aggarwal, C.C. (2013). *Outlier Analysis*. Springer.

[16] Zou, H., Hastie, T. (2005). Regularization and variable selection via the elastic net. *J. Royal Statistical Society B*, 67(2), 301–320.

[17] Friedman, J.H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5), 1189–1232.

[18] Vapnik, V.N. (1998). *Statistical Learning Theory*. Wiley-Interscience.

[19] Bishop, C.M. (2006). *Pattern Recognition and Machine Learning*. Springer.

[20] Hastie, T., Tibshirani, R., Friedman, J. (2009). *The Elements of Statistical Learning*, 2nd ed. Springer.

[21] Montgomery, D.C., Jennings, C.L., Kulahci, M. (2012). *Introduction to Time Series Analysis and Forecasting*, 2nd ed. Wiley.

[22] Shumway, R.H., Stoffer, D.S. (2017). *Time Series Analysis and Its Applications*, 4th ed. Springer.

[23] Leskovec, J., Rajaraman, A., Ullman, J.D. (2014). *Mining of Massive Datasets*, 2nd ed. Cambridge University Press.

[24] Kumar, A., Singh, R. (2020). Machine learning approaches in public procurement optimization.

International Journal of Applied Mathematics, 50(4), 445–462.

[25] Zhao, Y., Nasrullah, Z., Li, Z. (2021). PyOD: A Python toolbox for scalable outlier detection.

Journal of Machine Learning Research, 20(96), 1–7.

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

[26] Seabold, S., Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with Python.*Proc. 9th Python in Science Conference*, 92–96.

[27] Pedregosa, F., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learn- ing Research*, 12, 2825–2830.

[28] Abadi, M., et al. (2015). TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org.

[29] Chen, T., Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proc. 22nd ACM SIGKDD*, 785–794.

Received: August12, 2025