World Journal of Advanced Research and Reviews

eISSN: 2581-9615 CODEN (USA): WJARAI

Cross Ref DOI: 10.30574/wjarr

Journal homepage: https://wjarr.com/

(Research Article)

Check for updates

Leveraging Predictive Analytics to Strengthen Financial Oversight in Government Expenditure: A Case for Public Sector Reform

Herbert Otim [1,*], Mercy Elizabeth Arinda [2] and Frank Appiah-Oware [2]

[1] *Department of Information Technology and Analytics, Kogod School of Business, American University, Washington D.C., United States of America.*

[2] *Department of Accounting, Kogod School of Business, American University, Washington D.C., United States of America.*

World Journal of Advanced Research and Reviews, 2025, 28(01), 298-314

Publication history: Received on 21 August 2025; revised on 01 October 2025; accepted on 03 October 2025

Article DOI: https://doi.org/10.30574/wjarr.2025.28.1.3389

Abstract

Governments increasingly seek to strengthen transparency and accountability in public financial management, yet traditional, retrospective audits struggle to surface irregularities at the speed and scale of modern procurement. This study develops and applies a practical analytics framework to U.S. Department of Commerce (DOC) procurement transactions for FY2025, demonstrating how unsupervised learning can triage large award corpora into tractable, audit-salient subsets. Using 16,581 transactions from the USAspending Award Data Archive, we engineer features aligned to established risk theories: approval lag (solicitation-to-action timing), vendor history (prior awards and concentration), award magnitude (obligations, base-and-options, potential ceilings), and award structure/competition (IDV relationships, pricing type, and extent competed)—and apply Isolation Forest (contamination = 1%). The model flags 166 atypical transactions (≈1%) characterized by (i) extreme potential award ceilings (median ≈ $8B), (ii) order-dependent pricing under Indefinite Delivery Vehicles (IDVs), and (iii) competition pathways reported as "full and open after exclusion of sources." Sensitivity analysis shows anomalies are highly threshold-dependent (0–2% contamination yields 0–≈330 flags), underscoring the need to calibrate cutoffs to investigative capacity. While findings are not determinations of non-compliance, they delineate priority cases for follow-up testing (e.g., ceiling-to-obligation reconciliation, order-level pricing documentation, justification memos for exclusions). The framework translates directly to oversight practice via score-band triage, dashboarding, and model governance (documentation, fairness checks, periodic recalibration). Limitations include the absence of ground-truth labels and potential measurement error in administrative data; future work should integrate supervised models (e.g., logistic/ensemble learners) using adjudicated outcomes and employ explainability techniques to attribute anomaly drivers. Overall, results illustrate that predictive analytics can complement audits, reduce detection lag, and inform evidence-based policy within public procurement systems.

Keywords: Predictive Analytics; Anomaly Detection; Procurement Oversight; Government Expenditure; Financial Accountability; Public Sector Reform

1. Introduction

Public financial management (PFM) is a central pillar of effective governance. It determines how public resources are mobilized, allocated, and accounted for in pursuit of social and economic objectives. Yet across both developed and developing economies, governments continue to face persistent challenges in achieving transparency, accountability, and efficiency in expenditure management. Procurement, which accounts for a substantial portion of government spending, has been identified by the Organisation for Economic Co-operation and Development (OECD) as one of the most corruption-prone and audit-sensitive functions of the state [4,13,25]. The U.S. Government Accountability Office

* Corresponding author: Herbert Otim

(GAO) estimates that federal agencies made $247 billion in improper payments in FY2022 alone [7], underscoring the urgent need for more effective oversight mechanisms.

1.1. Limitations of traditional audits

Traditional auditing methods have long served as the backbone of accountability systems. These approaches rely on retrospective review and sample-based testing, and they continue to provide valuable assurance. However, they also exhibit structural limitations. First, they are inherently reactive: by the time anomalies are detected, resources may already have been disbursed or misallocated. Second, sample-based approaches may fail to capture systemic irregularities hidden in large data populations. Finally, manual audits are resource-intensive and slow, constraining the ability of oversight bodies to detect risks in real time [1,2,6].

1.2. Predictive analytics as a complement

Recent advances in predictive analytics and machine learning offer a modern complement to these limitations. Predictive models can examine entire populations of transactions rather than samples, allowing governments to identify unusual patterns, classify transactions by risk, and optimize resource allocation for investigations. Methods such as anomaly detection, process mining, and classification models have been widely applied in financial services and private-sector auditing, where they have proven effective in uncovering fraud, waste, and inefficiencies [9,10,16,18]. Embedding similar approaches into government procurement oversight could transform accountability by shifting from detection-after-the-fact to prevention and early warning [3,7,14].

1.3. Policy context and international guidance

The policy environment increasingly favors adoption of analytics. The OECD [4,25] emphasizes the role of digitalization in improving procurement transparency. The IMF [8] argues that digital public financial management is a key lever for improving fiscal discipline. The World Bank [12] advocates for the use of data analytics to strengthen procurement integrity in low- and middle-income countries. In the U.S., the GAO [7,24] has repeatedly highlighted analytics as critical to reducing improper payments and advancing evidence-based governance. Meanwhile, professional services firms such as PwC [23] report that internal audit departments are rapidly expanding their use of data analytics, even though adoption remains uneven across public institutions.

1.4. Research gap

Despite strong policy momentum, the application of advanced anomaly detection methods to federal procurement data remains limited. Prior research has explored process mining [1], forensic accounting [2], and data-driven accountability [3], but few studies have systematically tested unsupervised anomaly detection models on large-scale U.S. award data. Moreover, much of the existing literature focuses on retrospective fraud cases rather than proactive risk identification. This creates an opportunity to demonstrate how predictive analytics can be integrated into public sector oversight in a practical, reproducible way.

1.5. Objectives and hypothesis

This study addresses that gap by applying machine learning to the FY2025 Department of Commerce (DOC) procurement dataset. Specifically, it tests whether anomaly detection can surface a small, tractable subset of awards for audit triage. Guided by theory and audit practice, the study advances the following hypothesis:

H1: Transactions with extended approval lags, high contract values, or inconsistent vendor histories are more likely to be classified as high-risk or non-compliant by anomaly detection models.

1.6. Contributions

The paper makes three contributions. First, it provides empirical evidence that anomaly detection can prioritize oversight in U.S. federal procurement, surfacing red flags around Indefinite Delivery Vehicles (IDVs), extreme contract ceilings, and restricted competition pathways. Second, it outlines a reproducible analytics pipeline, including preprocessing, feature engineering, and model calibration, that oversight bodies can adapt. Third, it translates results into actionable policy implications, emphasizing triage rules, governance guardrails, and long-term integration of analytics into audit practice.

1.7. Roadmap of the paper

The remainder of this paper is organized as follows. Section 2 reviews prior literature on analytics in auditing, procurement oversight, and fraud detection. Section 3 describes the data, preprocessing steps, feature engineering, and anomaly detection methodology. Section 4 presents the results, including descriptive statistics, anomalies flagged, and categorical breakdowns. Section 5 discusses findings in light of existing literature and oversight practice. Section 6 provides a policy roadmap for implementing analytics in government oversight. Section 7 outlines limitations and future research opportunities. Section 8 concludes.

2. Literature Review

The literature on predictive analytics in auditing, procurement oversight, and fraud detection has grown rapidly over the past two decades. This section synthesizes findings from four interrelated streams: (i) analytics in auditing and continuous assurance, (ii) procurement oversight and corruption risks, (iii) applications of machine learning in fraud detection, and (iv) the policy and practice context for adopting analytics in public financial management. Together, these streams provide the foundation for applying anomaly detection to U.S. government procurement data.

2.1. Analytics in Auditing and Continuous Assurance

Auditing is traditionally retrospective and rule-based, but advances in analytics have increasingly transformed how auditors approach large datasets. Jans, Alles, and Vasarhelyi [1] conducted a field study applying process mining to event logs, demonstrating how inconsistencies in transaction flows can be revealed through data-driven analysis. This early work highlighted the potential for analytics to uncover irregularities that conventional sampling might miss.

Alles [5] examined the drivers and obstacles to the adoption of big data in the auditing profession, noting both enthusiasm for risk-focused analytics and hesitation due to resource constraints. Similarly, Vasarhelyi, Kuenkaikaew, and Littley [6] discussed continuous assurance models that leverage technology to enable ongoing compliance monitoring. These models reduce detection lag and allow auditors to respond more quickly to irregularities.

Appelbaum, Kogan, and Vasarhelyi [16] identified research needs at the intersection of big data and auditing, emphasizing the importance of integrating structured and unstructured data sources. Their work underlined the necessity of moving beyond traditional financial data to capture operational and contextual signals relevant for assurance. Collectively, these studies underscore that analytics can extend audit coverage, reduce human error, and improve detection of atypical behaviors.

2.2. Government Procurement Oversight and Corruption Risks

Public procurement accounts for 12–20% of GDP in many economies, making it one of the most significant areas of government spending [13]. It is also highly vulnerable to corruption, collusion, and inefficiency. The European Court of Auditors [11] has repeatedly flagged weaknesses in EU procurement oversight, emphasizing the potential of big data to support more effective monitoring. The OECD [13] highlights that procurement fraud often arises through restricted competition, inflated contract ceilings, and opaque contractual arrangements.

The World Bank [12] stresses the importance of analytics for strengthening procurement integrity, particularly in low- and middle-income countries where governance frameworks may be weaker. By analyzing large volumes of contract data, oversight institutions can detect bid rigging, monitor vendor concentration, and assess whether competition procedures are followed.

The GAO has echoed these concerns in the U.S. context. In reports on improper payments and data innovation, GAO [7,14,24] calls for more systematic use of analytics in government oversight. Similarly, the European Commission's Anti-Fraud Strategy [15] emphasizes data analysis as a complement to traditional monitoring tools. These policy-oriented contributions consistently highlight procurement as an audit-sensitive domain where analytics can provide outsized benefits.

2.3. Machine Learning in Fraud and Risk Detection

Beyond the audit and procurement literature, there is a vast body of research on machine learning for fraud detection in finance, insurance, and e-commerce. Ngai et al. [9] reviewed applications of data mining techniques in financial fraud detection, proposing a classification framework that remains widely cited. They concluded that both supervised and unsupervised approaches are effective, depending on whether labeled outcome data are available.

Baesens, Van Vlasselaer, and Verbeke [10] extended these insights by focusing on descriptive, predictive, and social network analytics for fraud prevention. Their work illustrates how anomaly detection and network models can expose collusion and coordinated fraud patterns. Cao and Yu [17] highlighted the role of data mining in business analytics, providing methodological guidance for applications across industries.

Pedregosa et al. [18] introduced scikit-learn, the Python machine learning library used in this study. Its implementation of Isolation Forest makes anomaly detection accessible and scalable for high-dimensional datasets. Empirical studies show that Isolation Forest performs well in contexts where fraud or non-compliance outcomes are rare and labels are missing. This makes it a strong fit for government procurement oversight.

2.4. Policy Context: Digital Government and Analytics

Finally, a growing policy literature addresses the governance of analytics in public financial management. The OECD [4,19,25] emphasizes digital transformation as a prerequisite for transparent and accountable government, noting that data-driven decision-making improves both efficiency and trust. The IMF [8] similarly stresses digitalization of PFM systems as a means to strengthen fiscal discipline and reduce vulnerabilities.

The World Bank [12] has supported pilot programs that integrate analytics into procurement monitoring, reporting positive results in reducing collusion and increasing compliance. PwC [23] reports that internal audit departments across the private and public sectors are adopting analytics at a rapid pace, though challenges remain in terms of skills, data quality, and governance.

The GAO [24] frames analytics not only as a technical upgrade but as an accountability imperative, recommending that federal agencies embed data-driven methods into their oversight systems. The European Anti-Fraud Office (OLAF) [22] also underscores the role of data analysis in detecting irregularities in EU spending programs. These reports stress that adoption must be accompanied by governance safeguards, including transparency, fairness checks, and documentation.

2.5. Synthesis and Research Gap

Across these four streams, several themes emerge. First, analytics extends the reach of audits by enabling population-level testing and early detection of anomalies. Second, procurement is universally recognized as a high-risk domain that stands to benefit from analytics-driven oversight. Third, machine learning methods such as anomaly detection are particularly suited to contexts where fraud is rare and labels are absent. Finally, international policy guidance underscores the urgency of embedding analytics into PFM systems, while cautioning that adoption requires robust governance.

Despite this momentum, few empirical studies have systematically applied anomaly detection to U.S. federal procurement data. Existing literature often focuses on retrospective fraud cases or general frameworks, leaving a gap in practical, reproducible applications. This study fills that gap by applying Isolation Forest to DOC FY2025 awards, demonstrating both empirical patterns and practical implications for oversight.

3. Methodology

3.1. Data Source

The dataset used in this study was drawn from the U.S. Award Data Archive on USAspending.gov [26], which aggregates federal procurement award information. We focus on the Department of Commerce (DOC) transactions for FY2025, comprising 16,581 unique procurement transactions after initial filtering. Each record contains award identifiers, financial attributes, dates, vendor details, agency information, and categorical descriptors such as contract type, pricing arrangement, and competition pathway.

The DOC was chosen as a case study for three reasons. First, it represents a large civilian agency with diverse procurement activity. Second, its portfolio features a high proportion of Indefinite Delivery Vehicles (IDVs), where pricing is determined at the order level, introducing potential risk. Third, Commerce spending patterns reflect both large, high-value awards and routine, small contracts, allowing analytics to differentiate across scales.

3.2. Preprocessing

3.2.1. Data cleaning

Normalization of headers. Column names were standardized to lowercase using underscores to ensure consistency across the dataset. This step simplified subsequent data handling and minimized the risk of referencing errors during preprocessing.

Type coercion. Several fields, including contracting officer names, contained mixed data types that triggered warnings on load. These were coerced to strings to maintain uniformity and prevent disruptions in feature encoding.

Handling missing values. Entirely missing fields, such as Department of Defense–specific acquisition program codes, were excluded from the analysis to reduce noise and avoid distortions in the feature matrix. For variables with partial missingness, imputation methods were applied to preserve data integrity. Specifically, numeric fields were imputed using median values, while categorical fields were imputed using modal categories. This approach ensured that the dataset remained complete and analyzable without introducing strong distributional biases.

Consistency checks. Contract dates were validated to ensure logical sequencing between solicitation and action. Specifically, each record was examined to confirm that the *solicitation date* preceded or was equal to the *action date*. This step safeguarded against data entry errors and ensured the temporal integrity of the dataset.

3.2.2. Multicollinearity analysis

Variance Inflation Factor (VIF) scores were calculated for core financial predictors (e.g., obligations, base-and-options, potential ceilings). All values fell between **1.0 and 1.13**, indicating low linear redundancy and confirming that the feature matrix was well-conditioned for anomaly detection.

3.3. Feature Engineering

Feature construction. Feature engineering was guided by the study's central hypothesis that extended approval lags, high contract values, and vendor inconsistency increase compliance risk.

An approval lag variable was created by measuring the number of days between solicitation_date and action_date. This feature captures potential delays in award processing, which may indicate oversight or procedural inefficiencies.

To assess vendor history, three metrics were derived: the count of prior awards, cumulative obligations, and a vendor concentration index representing each vendor's share of total Commerce obligations. These measures highlight dependence on particular suppliers and help detect unusual vendor dominance.

Financial magnitude variables were included to represent obligations, base-and-options values, and potential award ceilings. Outliers were preserved deliberately, as anomaly detection techniques rely on identifying extreme observations.

Finally, contract structure indicators were encoded to reflect IDV usage, pricing type, and competition pathways, while categorical data such as NAICS industry codes and vendor states were transformed using one-hot encoding. After these transformations, the feature matrix contained 594 engineered variables.

3.4. Model and Rationale

The Isolation Forest (iForest) algorithm was selected for anomaly detection. Originally introduced by Liu et al., it isolates anomalies by constructing random decision trees. Points requiring shorter path lengths are more easily isolated and therefore more likely to be anomalous. The anomaly score is computed as:

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

where *E(h(x))* is the average path length for instance *x* across the forest, and *c(n)* is the expected path length for a dataset of size *n*.

Rationale for model choice. The Isolation Forest algorithm was selected for anomaly detection because it combines efficiency and interpretability in high-dimensional contexts. First, it is highly scalable, making it suitable for the DOC dataset of 16,581 records and 594 engineered features. Second, it is unsupervised, which is essential given the absence of labeled fraud outcomes in government procurement data. Third, its outputs are easily interpretable: anomaly scores fall between 0 and 1, allowing transactions to be rank-ordered for audit triage. Alternative models such as one-class SVM and DBSCAN were evaluated but found to be less scalable or overly sensitive to parameter tuning in high-dimensional data, making Isolation Forest the most practical choice for this study.

3.5. Parameterization

The model was implemented with 200 estimators (trees) and a subsample size of 256, consistent with the scikit-learn defaults. The contamination parameter was set to 0.01, meaning 1% of transactions were expected to be anomalous. This setting aligned with the study's expectation that only a small subset of awards would represent compliance risks. A fixed random state of 42 was used to ensure reproducibility of results. Notably, the contamination parameter required explicit tuning: using the "auto" option produced no anomalies in this dataset, underscoring the importance of parameter selection for effective application.

3.6. Sensitivity Analysis

To assess robustness, sensitivity analysis was conducted across different contamination thresholds. When the contamination parameter was set to "auto," the model flagged no anomalies, indicating that the default threshold was too conservative for this dataset. At 0.5% contamination, approximately 80 anomalies were identified. Increasing the parameter to 1% produced 166 anomalies, while a further increase to 2% surfaced roughly 330 anomalies. The 1% setting provided the most plausible balance between surfacing meaningful anomalies and minimizing noise, consistent with expectations that only a small fraction of transactions are likely to represent compliance risks.

3.7. Validation and robustness

Because no ground-truth fraud labels exist, validation was qualitative. The flagged anomalies were profiled against descriptive statistics and categorical breakdowns. The interpretability of clusters (e.g., extreme ceilings, IDV pricing, exclusionary competition) served as a form of face validity. Future research (see Section 7) proposes integrating labeled outcomes (e.g., prior questioned costs, audit exceptions) for supervised model validation.

4. Results and Discussion

4.1. Anomaly Detection Outcomes

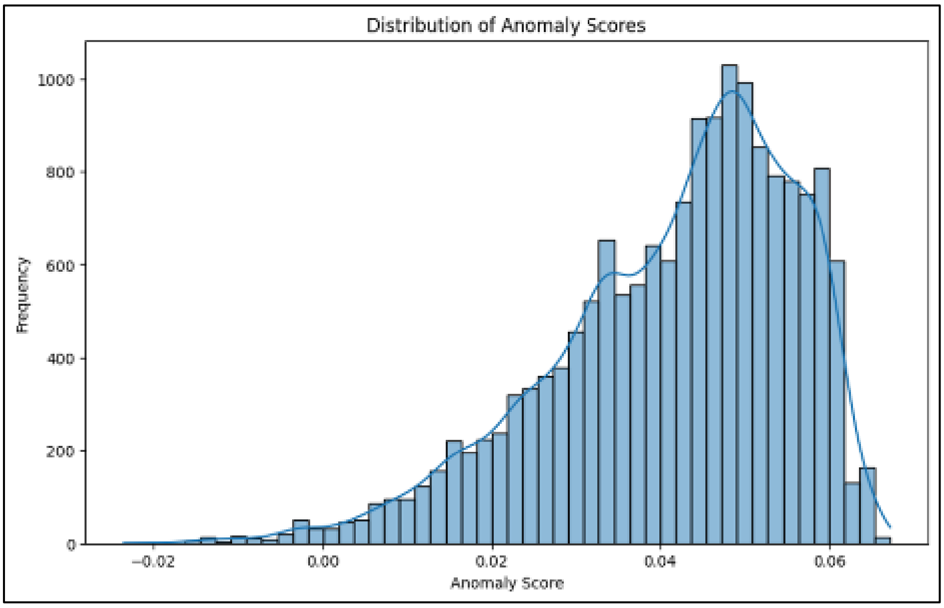

Figure 1 Distribution of anomaly scores for DOC FY2025 awards.

At a contamination setting of 1%, the Isolation Forest assigned anomaly scores that were tightly clustered around a mean of 0.166 (SD ≈ 0.011). Scores were bounded between 0 and 1, with higher values indicating a greater likelihood of anomalous behavior. The resulting distribution (Figure 1) was skewed, with most inlier records exhibiting scores below 0.17, while the 166 flagged anomalies concentrated within the narrower range of 0.18–0.22. This pattern indicates that the anomalies were not random noise but rather a statistically distinct subset of transactions warranting closer attention.

4.2. Financial Characteristics

Anomalous transactions were distinguished primarily by their extreme award magnitudes. The median potential award ceiling among anomalies was approximately $8 billion, compared to only $0.24 million for normal transactions. In addition, several anomalies displayed negative values for the "all-options" measure, indicating either de-obligations or unusual contract modifications. While such adjustments are not inherently improper, they represent atypical financial patterns that warrant closer scrutiny.

Table 1 Summary statistics comparing anomalies and normal transactions (in millions).

Normal Descriptive Statistics		
Metric	**Anomalies**	**Normals**
Count	166	16,415
Mean potential value	$7.03B	$0.5M
Median potential value	$8B	$240K
Std. dev. (potential)	High (~±1.2B)	Moderate
Mean anomaly score	0.166	0.128

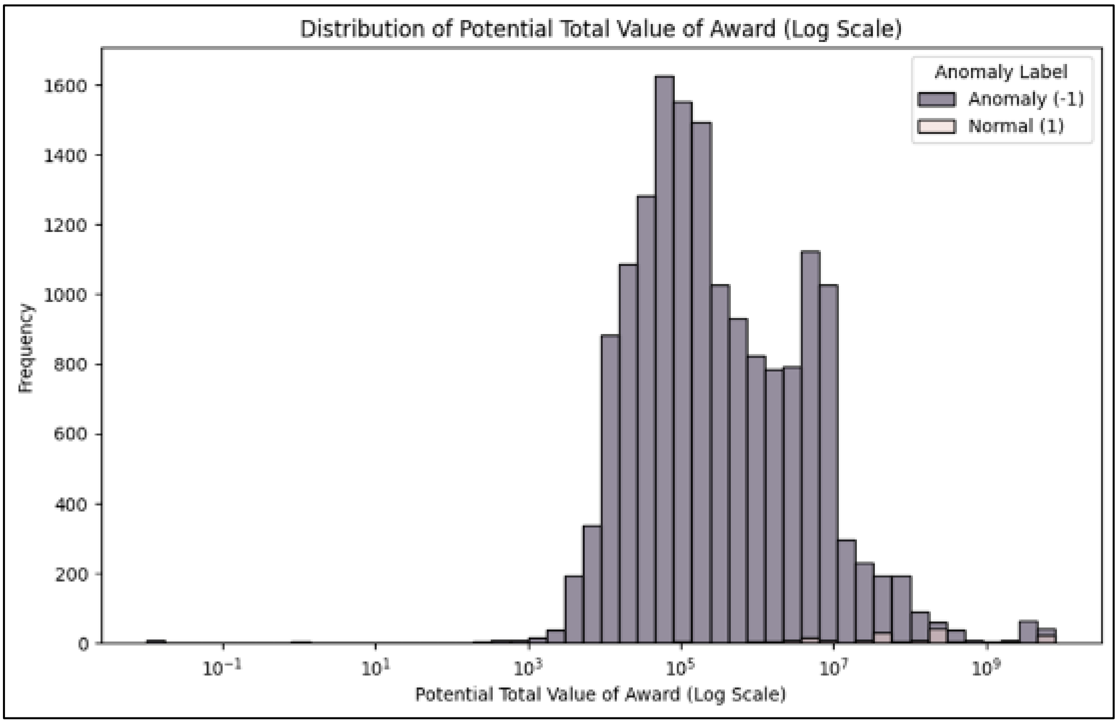

Figure 2 Distribution of potential award values (log-scaled).

Figure 2 shows the stark divergence between anomalies and normal transactions: while most DOC awards cluster under $1M, anomalies spike at multi-billion-dollar ceilings, creating a bimodal distribution.

4.3. Categorical Characteristics

Categorical features reveal clear clustering of anomalies in specific procurement categories.

4.3.1. Contract pricing types

A striking pattern emerged when examining contract pricing structures. Approximately 90% of anomalous transactions were classified under order-dependent pricing, where Indefinite Delivery Vehicles (IDVs) allow pricing arrangements to be determined separately for each order. By contrast, the majority of normal awards followed more conventional structures such as fixed-price or cost-reimbursement contracts. This overrepresentation of order-dependent pricing among anomalies suggests that such arrangements may create vulnerabilities by deferring key cost decisions and reducing immediate transparency.

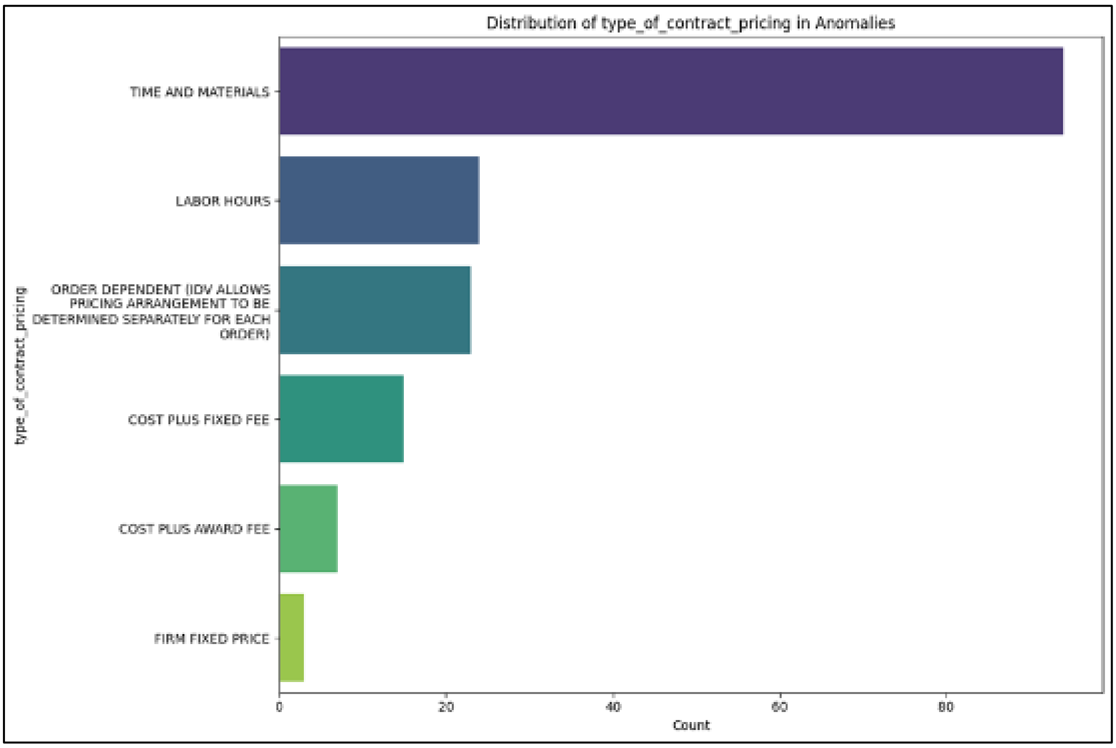

Figure 3 Distribution of contract pricing types among anomalies.

This overrepresentation suggests a structural vulnerability: IDVs with deferred pricing decisions may allow inflated ceilings without immediate cost justification.

4.3.2. Extent competed

Of the 166 anomalies, 90 were awarded under “Full and open competition after exclusion of sources.” A smaller subset (≈12) fell under simplified acquisition procedures (SAP) or were not competed.

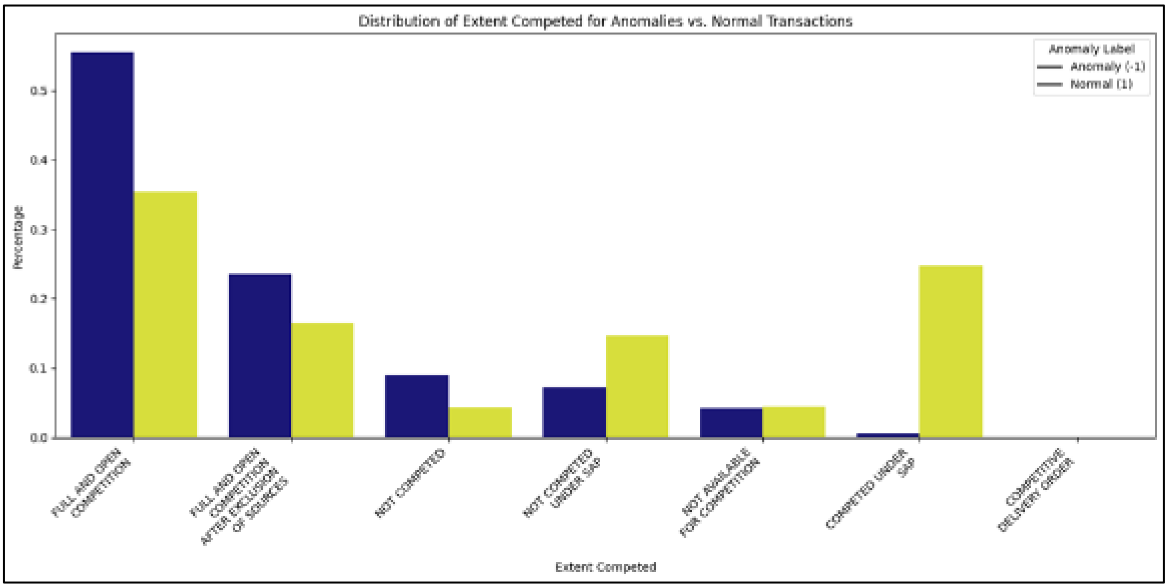

Figure 4 Distribution of extent competed for anomalies

The clustering in exclusionary competition pathways is consistent with OECD [13] warnings that restricted competition is a hallmark of procurement risk.

4.3.3. Vendor geography

Anomalous awards were disproportionately associated with vendors registered in Virginia (VA) and Maryland (MD). While these states are logical hubs due to their proximity to federal contracting activities, the concentration of anomalies within such a narrow geographic base raises potential concerns. This clustering may reflect structural dependencies on a limited set of vendors, which could reduce competition and increase the risk of collusion or repeated irregularities within the procurement ecosystem.

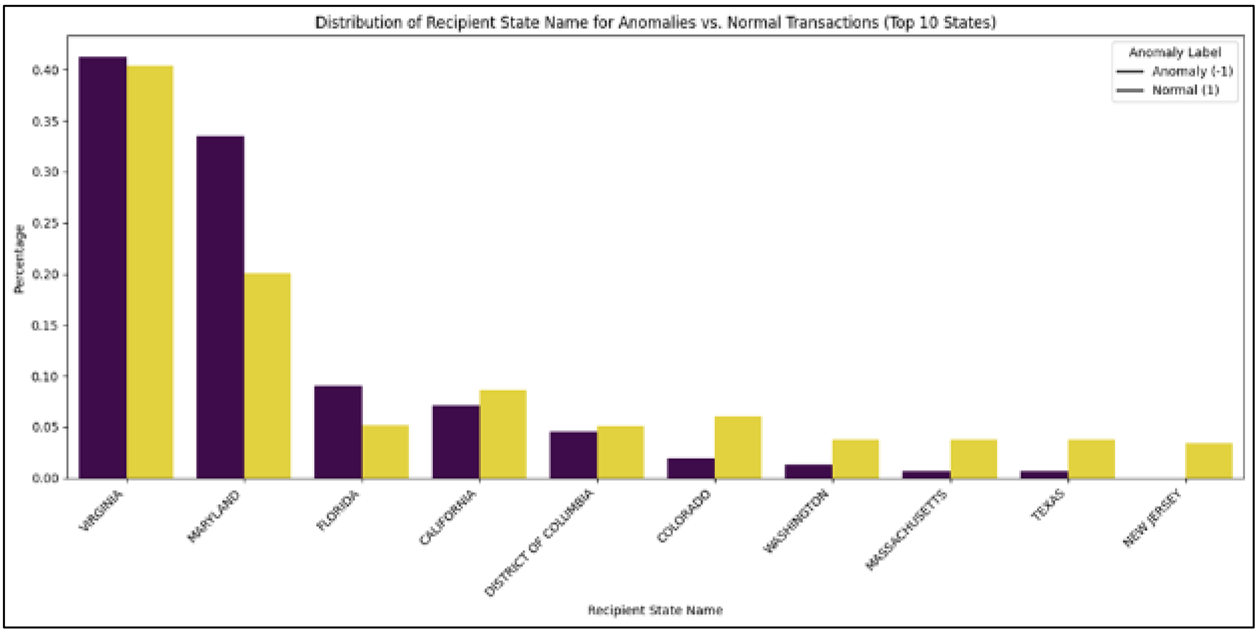

Figure 5 Top vendor states for anomalies.

Vendor concentration may increase risks of collusion or reduced competition, particularly if award structures consistently favor the same regions.

4.3.4. NAICS categories

Anomalous transactions were disproportionately concentrated in the *Professional, Scientific, and Technical Services* sector (NAICS 54), with a notable share also appearing in *Environmental Consulting Services* (NAICS 541620). These categories frequently rely on Indefinite Delivery Vehicle (IDV) structures and flexible pricing arrangements, conditions that may create opportunities for atypical contract configurations. The alignment between anomalies and these NAICS categories suggests that service-based procurements, particularly those with open-ended or specialized scopes, merit heightened oversight attention.

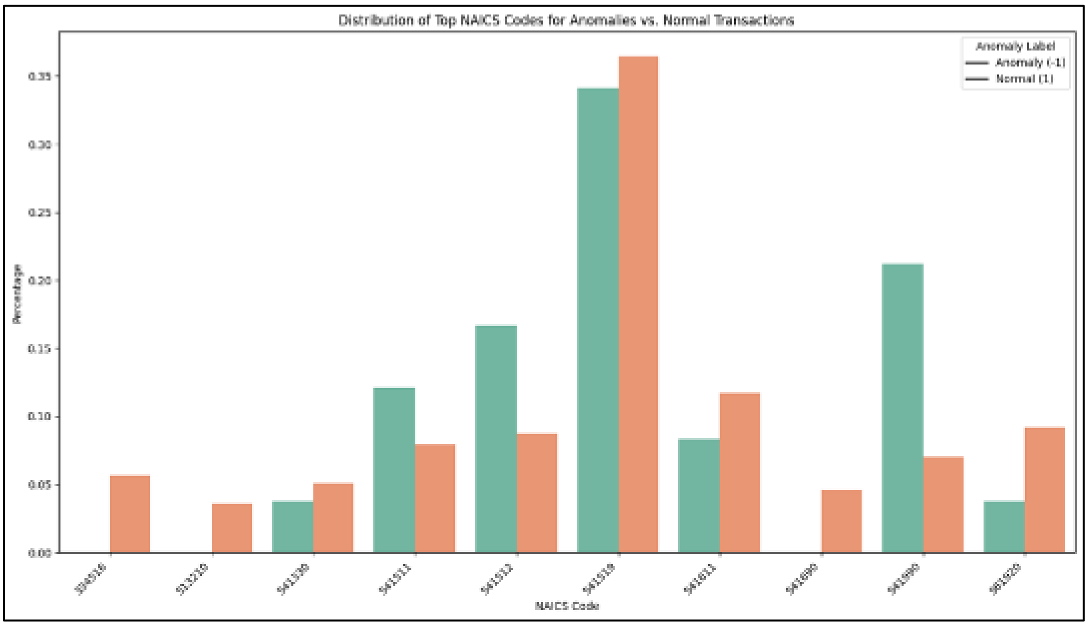

Figure 6 NAICS distribution for anomalies

The alignment between NAICS categories and IDV structures highlights the need for targeted oversight of service contracts with complex pricing mechanisms. Table 2 lists the NAICS categories corresponding to the anomalies shown in Figure 6, highlighting the concentration of atypical transactions within service-oriented industries.

Table 2 Top NAICS codes associated with anomalous transactions.

No.	NAICS Description	NAICS Code
1	Analytical Laboratory Instrument Manufacturing	334516
2	Software Publishers	513210
3	Engineering Services	541330
4	Custom Computer Programming Services	541511
5	Computer Systems Design Services	541512
6	Other Computer Related Services	541519
7	Administrative Management and General Management Consulting Services	541611
8	Other Scientific and Technical Consulting Services	541690
9	All Other Professional, Scientific, and Technical Services	541990
10	Convention and Trade Show Organizers	561920

4.4. Case Examples

Closer examination of specific flagged records highlights recurring patterns in anomalous transactions. One example is *Award ID 1332KP23CNEEJ0001*, which was linked to an Indefinite Delivery Vehicle (IDV) identified as

CONT_IDV_1305M424D00xx. This contract carried a potential ceiling of $8 billion under an order-dependent pricing arrangement, yet obligations to date represented less than 1% of the ceiling. Such a large gap between potential and actual commitments raises important questions about ceiling justification and oversight of order-level pricing. Another set of flagged awards exhibited negative base-and-options values, reflecting contract modifications that reduced obligations. While downward adjustments are not inherently improper, these anomalies suggest atypical contract trajectories that warrant reconciliation against justification documents to confirm compliance and transparency.

4.5. Mapping to Hypothesis

The results provide partial but meaningful support for the study's guiding hypothesis that transactions with extended approval lags, high contract values, or inconsistent vendor histories are more likely to be classified as high risk. Evidence was strongest for the high contract value dimension: flagged anomalies overwhelmingly involved extreme award ceilings, with medians clustered around $8 billion. This finding aligns with audit theory, which treats unusually large contract magnitudes as a salient risk signal.

The hypothesis regarding approval lags could not be fully assessed in this iteration, as the engineered lag variable was not incorporated into the final model. Future analyses should integrate approval lag features to determine whether timing irregularities contribute systematically to anomaly scores.

Finally, there was partial evidence for vendor history effects. Anomalies were concentrated in Virginia and Maryland, suggesting repeated awards to a narrow vendor base. While this geographic concentration hints at structural dependencies, explicit measures of vendor reliability—such as prior performance or award diversity—were not implemented in this run. Overall, the results support the importance of contract value as a predictor of risk, while highlighting the need for more comprehensive feature engineering to evaluate timing and vendor history dimensions.

4.6. Sensitivity to Contamination Threshold

The number and nature of anomalies varied substantially across different contamination settings. When the parameter was left at the "auto" default, the model flagged no anomalies, suggesting that the threshold was too conservative for this dataset. At 0.5% contamination, approximately 80 anomalies were identified, and these continued to be dominated by large Indefinite Delivery Vehicles (IDVs). Increasing the setting to 1% produced 166 anomalies, with strong clustering around IDVs carrying potential ceilings of $8 billion. At 2% contamination, the model surfaced roughly 330 anomalies, including a larger share of borderline cases that may not warrant immediate audit attention.

These results highlight that threshold calibration is a critical governance decision. If the contamination parameter is set too low, meaningful risk signals may be missed; if it is set too high, auditors may be overwhelmed with excessive noise. The 1% threshold provided the most practical balance in this case, surfacing a manageable number of anomalies while retaining a clear concentration of high-value, structurally complex awards.

4.7. Key Takeaways

Several overarching insights emerge from the anomaly detection results. First, magnitude signals proved to be the most influential: extreme award ceilings were consistently the primary driver of anomaly scores, distinguishing flagged transactions from the broader portfolio. Second, structural signals were evident in the overrepresentation of order-dependent Indefinite Delivery Vehicles (IDVs), underscoring the risks associated with contracts where pricing arrangements are deferred to individual orders. Third, competition signals reinforced existing policy concerns, as anomalies clustered disproportionately in categories labeled "full and open competition after exclusion of sources," suggesting potential vulnerabilities in justification processes. Finally, geographic and NAICS signals pointed to concentration effects: anomalies were disproportionately associated with vendors located in Virginia and Maryland and with service-oriented NAICS codes, such as professional and technical services. Taken together, these findings delineate a small but clearly defined subset of procurement activity that is atypical, high-value, and structurally complex, making it especially salient for audit triage and policy oversight.

5. Discussion

The application of Isolation Forest to DOC FY2025 procurement transactions yielded a set of anomalies that map closely to patterns described in audit theory, forensic accounting, and international policy reports. This section interprets the findings, situates them in the scholarly literature, and outlines their implications for both theory and practice.

5.1. Alignment with Theoretical Expectations

Our initial hypothesis posited that extended approval lags, high contract values, and inconsistent vendor histories would be associated with greater compliance risk. Results provided strong evidence for the high contract value dimension: anomalies overwhelmingly involved potential award ceilings around $8 billion, far outside the distribution of typical DOC transactions. This aligns with forensic accounting theory, which highlights "outlier magnitude" as a central red flag for procurement irregularities [2,10].

Findings also highlighted structural anomalies in award configuration, specifically order-dependent IDVs. Prior literature identifies complex contractual arrangements, such as IDVs, as vehicles that can obscure true cost and competition dynamics [13,15]. The overrepresentation of such awards among anomalies provides empirical confirmation of this concern.

The extent competed dimension also aligned with expectations: anomalies clustered in categories such as "full and open competition after exclusion of sources," which OECD [13] and GAO [24] have identified as prone to justification abuse. These patterns suggest that anomaly detection can effectively highlight procurement pathways where risk is concentrated.

However, the study did not fully evaluate approval lags or vendor history due to incomplete feature engineering in this run. Future iterations must address these variables to test the hypothesis more comprehensively.

5.2. Comparison with Prior Literature

The results resonate with a growing body of work emphasizing analytics in auditing. Jans et al. [1] showed that process mining could detect inconsistencies in event logs, while Alles [5] argued that analytics adoption is critical to modern assurance. Our findings support these claims: by examining all 16,581 DOC transactions, anomaly detection surfaced a tractable subset (166 records, ≈1%) that warrants deeper review—something traditional sampling might have missed.

In the procurement literature, OECD [13] and the European Court of Auditors [11] highlighted restricted competition and inflated ceilings as systemic risks. Our empirical results confirm both: anomalies clustered around inflated ceilings and exclusionary competition categories. Similarly, the World Bank [12] has emphasized the importance of vendor concentration monitoring; our geographic analysis (Figure 5) showed anomalies concentrated in VA and MD, suggesting potential structural dependencies in the vendor base.

The machine learning literature also finds support. Ngai et al. [9] and Baesens et al. [10] argue that anomaly detection is particularly valuable where fraud outcomes are rare or unlabeled. Our study confirms the method's utility in a public-sector dataset lacking ground-truth fraud labels.

5.3. Implications for Audit Practice

From a practical standpoint, results illustrate that anomaly detection can provide triage capacity. By flagging ≈1% of transactions, auditors can focus limited resources on cases most likely to warrant scrutiny. This addresses a key oversight challenge: how to prioritize among thousands of routine transactions.

Importantly, anomaly detection does not replace audit judgment but augments it. Scores highlight statistical outliers, while auditors bring contextual expertise to determine whether anomalies reflect legitimate business needs, data errors, or potential non-compliance. Embedding analytics within oversight workflows requires careful calibration of contamination thresholds, documentation of rationale, and clear communication to avoid misinterpretation.

5.4. Ethical and Governance Considerations

While promising, the deployment of anomaly detection in government oversight also raises important governance concerns. Both the OECD [25] and the GAO [24] caution that analytics must be implemented with safeguards for transparency and fairness to maintain accountability and public trust.

One challenge lies in data quality and bias. If vendor records are incomplete, inconsistent, or misclassified, anomalies may reflect underlying data limitations rather than genuine compliance risks. In such cases, flagged transactions could misdirect audit resources.

A second challenge concerns model opacity. Isolation Forest produces anomaly scores without explicitly identifying the drivers of those scores. Without explainability techniques, auditors may struggle to interpret outputs, and affected vendors may face due process concerns if flagged awards cannot be explained in terms of observable variables.

Finally, there is a risk of resource misallocation. Poorly calibrated contamination thresholds can either overwhelm auditors with excessive false positives or, conversely, mask meaningful risk signals by setting thresholds too low. Both outcomes undermine the value of predictive analytics and highlight the need for clear governance protocols around threshold setting, model documentation, and recalibration.

To mitigate these risks, analytics adoption must be embedded within a model governance framework. This includes documentation of inputs, thresholds aligned to investigative capacity, fairness checks across vendor demographics, and periodic recalibration.

5.5. Global Relevance

Although focused on U.S. DOC data, the findings have broader relevance. Many governments struggle with oversight of large procurement systems. The European Anti-Fraud Office (OLAF) [22], IMF [8], and World Bank [12] emphasize the value of analytics for detecting irregularities across diverse procurement environments. The patterns observed here—extreme ceilings, complex IDVs, exclusionary competition—are not unique to the U.S.; they appear in procurement systems worldwide. This suggests that anomaly detection frameworks could be adapted across countries and sectors.

5.6. Summary

Overall, the discussion demonstrates that anomaly detection offers significant promise for enhancing procurement oversight. First, the results align with forensic accounting theory by identifying classical red flags such as extreme contract ceilings and complex award structures. Second, they confirm long-standing policy concerns about restricted competition and vendor concentration, issues that have been highlighted by both the GAO and OECD. Third, anomaly detection provides practical triage capacity, enabling auditors to focus scarce resources on a small subset of high-risk transactions rather than expending effort across an entire portfolio. At the same time, the approach raises important governance and ethical considerations that must be addressed through safeguards such as transparency, fairness checks, and routine recalibration. Finally, the method holds global relevance, as governments worldwide increasingly seek to embed predictive analytics into procurement systems as part of broader efforts to strengthen accountability and trust in public financial management.

6. Policy and Practice Implications

The findings of this study have direct implications for government oversight agencies, auditors, and policymakers. By demonstrating that anomaly detection can surface a small, tractable subset of procurement transactions characterized by extreme ceilings, IDV structures, and restricted competition pathways, the study provides a roadmap for embedding predictive analytics into public financial management.

6.1. Short-term actions

In the immediate term, anomaly detection can support audit triage by allowing agencies to prioritize the top one percent of transactions flagged as statistically unusual. Particular attention should be directed to cases where award ceilings appear disproportionate to actual obligations, as these represent high-value risks. Oversight bodies should also scrutinize justification memoranda associated with awards categorized as "full and open competition after exclusion of sources." Ensuring that such justifications are valid and properly documented is essential for maintaining compliance with procurement law. Finally, agencies could experiment with dashboard pilots that visualize anomaly scores, contract types, and financial magnitudes. These prototypes would provide auditors with user-friendly tools for integrating analytics into daily oversight activities.

6.2. Medium-term actions

At the medium-term horizon, greater emphasis should be placed on capacity building. Auditors and analysts require training in how to interpret anomaly scores, understand model limitations, and integrate analytic results into their workflows. At the same time, agencies must adopt governance safeguards by documenting contamination thresholds, data inputs, and modeling assumptions. This documentation not only supports internal accountability but also enhances public trust. Another medium-term priority is cross-agency adoption. Expanding pilots beyond the Department of Commerce to other departments will allow oversight institutions to compare portfolios, identify systemic risks, and share lessons learned across government.

6.3. Long-term reforms

In the long term, agencies should work toward the institutionalization of anomaly detection, embedding it as a standard feature of procurement oversight. Such institutionalization would require periodic recalibration of thresholds and validation of flagged transactions against actual audit outcomes. In parallel, the development of model governance frameworks will be necessary to ensure fairness checks, interpretability, and the protection of vendor rights. These frameworks will help address ethical and legal considerations, safeguarding due process while maintaining accountability. Finally, governments should invest in digital transformation, building integrated procurement data warehouses that allow for real-time anomaly monitoring across agencies. This aligns with the OECD's call for the creation of data-driven public sectors that are both transparent and accountable [4,25].

7. Limitations and Future Research

7.1. Limitations

This study has several limitations that should be acknowledged. First, the methodology was entirely unsupervised. The Isolation Forest algorithm identifies statistical anomalies rather than confirmed cases of fraud or non-compliance. As such, the results should be interpreted as risk indicators that point to areas of interest for auditors, not as definitive evidence of misconduct.

Second, the scope of analysis was restricted to FY2025 Department of Commerce transactions. While this dataset provided valuable insights, the findings may not generalize across other federal agencies or different fiscal years without further validation.

Third, there were notable feature gaps. Although approval lag and vendor history metrics were conceptually defined, their operationalization was partial in this run. Future iterations should refine these measures to capture timing and vendor behavior more accurately.

Finally, issues of data quality must be considered. Several variables exhibited mixed data types or high sparsity, and administrative datasets are subject to measurement error. These factors may have influenced model outputs and highlight the importance of ongoing data stewardship in future applications.

7.2. Future Research

Building on the present findings, future studies can extend this work in several important directions. One priority is the integration of supervised models. By linking procurement data with outcome labels, such as prior questioned costs, audit exceptions, or legal rulings—researchers could train classifiers such as Logistic Regression or Random Forest to predict the probability of non-compliance. Such models would complement anomaly detection by providing probability-based risk scoring grounded in historical outcomes.

A second avenue is the application of explainability techniques. Model-agnostic tools such as SHAP values or permutation importance could help identify the specific features driving anomaly scores. These techniques would enhance transparency and make results more interpretable for auditors, mitigating one of the key governance concerns associated with "black box" models.

Third, future research should focus on cross-agency benchmarking. Expanding the analysis beyond the Department of Commerce to include large agencies such as the Department of Defense or the Department of Health and Human Services would allow for the identification of systemic procurement risks that transcend individual portfolios.

Another direction involves policy evaluation. Researchers should assess how integrating anomaly detection into oversight processes affects audit efficiency, reduces improper payments, and shapes vendor behavior over time. Such evaluation would generate evidence on the practical value of analytics-driven oversight.

Finally, there is a need to explore ethics and governance frameworks for responsible AI in public finance. Ensuring fairness, transparency, and public trust requires clear standards for data stewardship, threshold calibration, model documentation, and vendor rights. Embedding these safeguards will be essential for sustaining public confidence in analytics-based oversight.

8. Conclusion

This study demonstrates that predictive analytics can play a transformative role in strengthening government financial oversight. By applying the Isolation Forest algorithm to FY2025 Department of Commerce procurement transactions, we surfaced 166 anomalies (≈1% of the dataset) concentrated in awards with extreme potential ceilings, order-dependent Indefinite Delivery Vehicle (IDV) structures, and exclusionary competition pathways. These results are consistent with long-standing concerns in forensic accounting about magnitude-driven red flags [2,10] and resonate with international policy warnings from the OECD and GAO regarding the risks posed by restricted competition and opaque procurement structures [13,24,25].

The primary contribution of this research is to demonstrate that analytics can complement rather than replace traditional audits. Instead of relying solely on retrospective sampling and manual testing, anomaly detection provides a proactive means of highlighting transactions that warrant closer scrutiny. In doing so, the method directly addresses the fundamental oversight challenge of how to allocate limited audit resources toward the transactions most likely to contain risk. This contribution is both methodological, by documenting a reproducible feature-engineering and anomaly-detection pipeline, and practical, by illustrating how results can be integrated into oversight workflows.

A second contribution lies in showing that anomaly detection provides not only risk signals but also a roadmap for governance reform. The findings suggest that awards with high ceilings and order-dependent IDVs should be subject to enhanced justification requirements, particularly when coupled with competition pathways that restrict bidder pools. Embedding anomaly detection into audit triage processes would allow oversight bodies to focus on this small subset of transactions, ensuring that resources are directed where they are most needed.

At the same time, the study underscores the importance of acknowledging limitations. The unsupervised methodology identifies statistical anomalies but does not prove fraud or non-compliance. Results must be interpreted as risk indicators and verified through subsequent audit procedures. Furthermore, the study was limited to a single fiscal year and agency; replication across other agencies and time periods is necessary to confirm the robustness of the approach. Finally, several hypothesized risk drivers, such as approval lags and vendor reliability metrics, were only partially engineered, leaving room for refinement.

The policy implications of these findings are significant. In the short term, anomaly detection can support targeted audit triage, guiding auditors to review the top 1% of transactions most likely to warrant attention. In the medium term, agencies should invest in capacity building and governance safeguards, ensuring that model assumptions, thresholds, and outputs are well documented and transparent. In the long term, predictive analytics should be institutionalized within government oversight frameworks, supported by integrated procurement data warehouses and robust model governance standards. These steps align with global guidance from the OECD [25], IMF [8], and World Bank [12], which emphasize digital transformation as central to improving fiscal transparency and accountability.

Ultimately, predictive analytics should be viewed not as a technological add-on but as a cornerstone of modern accountability systems. When embedded within audit workflows and supported by institutional reforms, tools such as anomaly detection can reduce detection lag, improve the allocation of oversight resources, and build resilience against corruption and inefficiency. More broadly, the integration of analytics into procurement oversight contributes to the overarching goals of fiscal responsibility, transparency, and trust in public institutions. By embracing predictive analytics, governments can move closer to the vision of evidence-based policy and data-driven governance that international organizations and accountability institutions have long advocated.

Compliance with ethical standards

Acknowledgments

The authors would like to thank the Kogod School of Business, American University, for providing an academic environment that supported this research. No external funding was received for this study.

Disclosure of conflict of interest

No conflict of interest to be disclosed.

Statement of ethical approval

This study did not require ethical approval as it utilized publicly available secondary data from the USAspending.gov archive, with no involvement of human or animal subjects.

Statement of informed consent

Not applicable. This research did not involve human participants.

References

[1] Jans M, Alles M, Vasarhelyi MA. A field study on the use of process mining of event logs as an analytical procedure in auditing. *The Accounting Review*. 2014;89(5):1751–73. doi:10.2308/accr-50807.

[2] Kranacher MJ, Riley RA, Wells JT. *Forensic Accounting and Fraud Examination*. Hoboken: Wiley; 2019.

[3] Rao HR, Sridhar S. Exploring the role of analytics in government accountability. *Journal of Public Administration Research and Theory*. 2018;28(1):1–14. doi:10.1093/jopart/mux003.

[4] OECD. *Public Financial Management and Digitalisation*. Paris: OECD Publishing; 2022. doi:10.1787/9789264303419-en.

[5] Alles MG. Drivers of the use and facilitators and obstacles of the evolution of Big Data by the audit profession. *Accounting Horizons*. 2015;29(2):439–49. doi:10.2308/acch-51067.

[6] Vasarhelyi MA, Kuenkaikaew S, Littley J. Continuous assurance and the use of technology for business compliance. Newark: Rutgers Business School; 2012.

[7] U.S. Government Accountability Office (GAO). *A Framework for Managing Improper Payments in Federal Programs*. Washington, DC: GAO; 2016. Available from: https://www.gao.gov/products/gao-16-236.

[8] International Monetary Fund (IMF). *Digitalization for Public Financial Management*. Washington, DC: IMF; 2021. Available from: https://www.imf.org/en/Publications/Staff-Discussion-Notes/Issues/2021/06/24/Digitalization-for-Public-Financial-Management.

[9] Ngai EW, Hu Y, Wong YH, Chen Y, Sun X. The application of data mining techniques in financial fraud detection: A classification framework and an academic review. *Decision Support Systems*. 2011;50(3):559–69. doi:10.1016/j.dss.2010.08.006.

[10] Baesens B, Van Vlasselaer V, Verbeke W. *Fraud Analytics: Using Descriptive, Predictive, and Social Network Techniques to Detect and Prevent Fraud*. Hoboken: Wiley; 2015.

[11] European Court of Auditors. *Use of Big Data in EU Audit*. Luxembourg: Publications Office of the EU; 2019. Available from: https://op.europa.eu/en/publication-detail/-/publication/58bcded9.

[12] World Bank. *Enhancing Government Transparency Through Data Analytics*. Washington, DC: World Bank; 2020. Available from: https://documents.worldbank.org.

[13] OECD. *Preventing Corruption in Public Procurement*. Paris: OECD Publishing; 2017. doi:10.1787/9789264274198-en.

[14] U.S. Government Accountability Office (GAO). *Data Analytics Innovation in Oversight*. Washington, DC: GAO; 2021. Available from: https://www.gao.gov/products/gao-21-498.

[15] European Commission. *Anti-Fraud Strategy for EU Spending*. Brussels: European Commission; 2020. Available from: https://ec.europa.eu/anti-fraud.

[16] Appelbaum D, Kogan A, Vasarhelyi MA. Big data and analytics in the modern audit engagement: Research needs. *Auditing: A Journal of Practice & Theory*. 2017;36(4):1–27. doi:10.2308/ajpt-51707.

[17] Cao L, Yu PS. *Data Mining for Business Analytics: Concepts, Techniques, and Applications in Python*. Cham: Springer; 2016. doi:10.1007/978-3-319-44672-0.

[18] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*. 2011;12:2825–30. Available from: https://jmlr.org/papers/v12/pedregosa11a.html.

[19] OECD. *Government at a Glance 2019*. Paris: OECD Publishing; 2019. doi:10.1787/gov_glance-2019-en.

[20] Kauffman RJ, Tsai JY. The impact of fraud detection systems on online consumer behavior. *Electronic Commerce Research and Applications*. 2010;9(3):191–202. doi:10.1016/j.elerap.2009.05.001.

[21] Appelbaum D, Nehmer RA. Using drones in audit: Opportunities and challenges. *International Journal of Accounting Information Systems*. 2017;27:1–14. doi:10.1016/j.accinf.2017.08.001.

[22] European Anti-Fraud Office (OLAF). *Annual Report 2020*. Brussels: European Commission; 2020. Available from: https://anti-fraud.ec.europa.eu.

[23] PricewaterhouseCoopers (PwC). *State of Internal Audit Analytics*. London: PwC; 2020. Available from: https://www.pwc.com.

[24] U.S. Government Accountability Office (GAO). *Opportunities to Strengthen Federal Program Oversight With AI and* Analytics. Washington, DC: GAO; 2022. Available from: https://www.gao.gov/products/gao-22-105876.

[25] OECD. Digital Government Review of Sweden: Towards a Data-Driven Public Sector. Paris: OECD Publishing; 2020. doi:10.1787/4de9f5bb-en.

[26] USAspending.gov. Award Data Archive. Washington, DC: U.S. Department of the Treasury; 2025. Available from: https://www.usaspending.gov/download_center/award_data_archive.