INTERNATIONAL JOURNAL OF PUBLIC FINANCE AND ACCOUNTABILITY

NAWALA

Volume 3, Number 1, 2024

The Use of Big Data Analytics in Fraud Detection within Public Financial Management

Arisman[1*]

[1] Institut Teknologi Yogyakarta, Yogyakarta, Indonesia

Abstract

Article history:
Received: January 31, 2024
Revised: February 11, 2024
Accepted: March 29, 2024
Published: June 30, 2024

Keywords:
Big Data Analytics, Fraud Detection, Public Financial Management, Transparency.

Identifier:
Nawala
Page: 1-13
https://nawala.io/index.php/ijpfa

Fraud in public financial management poses a persistent threat to fiscal integrity and public trust, particularly as financial transactions become increasingly complex and data-driven. This study examines how big data analytics can be utilized to strengthen fraud detection processes in the public sector. Using a systematic literature review approach, the article synthesizes recent empirical and conceptual research to evaluate the effectiveness, opportunities, and challenges of implementing big data technologies. The analysis reveals that advanced analytics significantly improve the detection of anomalies and suspicious patterns by enabling real-time monitoring and predictive modeling. The discussion integrates evidence from multiple jurisdictions, comparing technological capabilities with institutional readiness, and addressing barriers such as data privacy, skills gaps, and interoperability issues. Findings suggest that big data analytics, when embedded within robust governance frameworks, can enhance transparency, improve audit efficiency, and support proactive fraud prevention strategies in public financial systems.

*Corresponding author:
(Arisman)

1. Introduction

Public financial management (PFM) systems handle vast volumes of transactions across procurement, payroll, tax, customs, and social transfers—domains that are both high-value and vulnerability-prone. As governments digitize workflows and consolidate financial data, the opportunities for analytics-driven oversight have expanded markedly. According to Appelbaum et al. (2017), big data analytics (BDA) offers the potential for earlier detection of anomalous patterns, improved targeting of audits, and continuous monitoring of controls, shifting fraud detection from episodic, sample-based review to near real-time surveillance. Yet realizing this promise in the public sector requires adapting methods proven in financial statement and payment fraud contexts to the unique complexity, heterogeneity, and governance constraints of PFM data environments (Pamisetty, 2021).

The academic literature documents a broad toolkit, supervised and unsupervised learning, network analysis, anomaly detection, text mining, and process mining, that can elevate fraud risk assessment (Cao et al., 2015; Yoon et al., 2015). In corporate contexts, Achakzai and Peng (2023) found that logistic regression, random forests, and support vector machines often outperform traditional red-flag methods in detecting irregularities. Translating these advances to PFM means leveraging full-population testing across ledgers, commitments, and vendor payments; profiling entities across procurement lots; and linking master data (vendors, bank accounts, officers) to uncover collusive structures typical of bid-rigging and conflict-of-interest schemes (Appelbaum et al., 2017).

Still, several implementation challenges recur. First, data quality and integration are chronic obstacles in government platforms, where legacy systems, inconsistent coding, and "off-system" transactions degrade signal-to-noise ratios. Cao et al. (2015) note that class imbalance is another critical issue: verified fraud labels are scarce, while the majority of transactions are legitimate, conditions that bias standard classifiers unless techniques such as cost-sensitive learning or anomaly detection are applied. Furthermore, Yoon et al. (2015) argue that explainability and auditability matter more in the public sector, as models must produce transparent rationales to meet due-process expectations and enable corrective action. Organizational capacity and governance, clear ownership between internal audit, supreme audit institutions, treasury, procurement, and anti-corruption bodies, ultimately determine whether analytic insights lead to timely investigations and sanctions (Mohammadi et al., 2020).

The literature also highlights the value of combining transactional analytics with process perspectives. Process mining can reconstruct actual P2P and budget execution flows, revealing control bypasses such as split purchases below tender thresholds that a pure anomaly model may miss (Cao et al., 2015). Network and entity resolution techniques can map relationships between vendors, bank accounts, and civil servants to surface shell entities and collusive cliques, which are common in procurement and payroll fraud (Appelbaum et al., 2017). However, Albashrawi (2016) caution that analytics is not a silver bullet: its effectiveness depends on iterative model governance, feedback loops with investigators, and integration with whistleblower and case-management systems.

Against this backdrop, a systematic literature review on BDA for fraud detection in PFM clarifies which analytical approaches consistently add value, the data prerequisites for deployment, and the governance frameworks that turn detections into deterrence. By synthesizing peer-reviewed evidence across accounting analytics, information systems, and audit research, this article scopes the state of the art, identifies implementation pitfalls, and distills practical recommendations for building credible, explainable, and sustainable fraud analytics in government finance.

2. Literature Review

Early work framed fraud detection as a classification and anomaly-detection problem, cataloging supervised, unsupervised, and hybrid techniques and highlighting recurring obstacles such as severe class imbalance, concept drift, and the scarcity of verified labels (Albashrawi, 2016; Fernández et al., 2021). Subsequent surveys consolidate these insights for financial contexts, underscoring the shift toward ensemble learning, feature construction at scale, and the integration of network information (Akoglu et al., 2015; West & Bhattacharya, 2016). In public financial management (PFM), these challenges are amplified by heterogeneous administrative systems, evolving fraud schemes, and policy-driven reporting changes that induce drift in data-generating processes (Jans et al., 2014; Yoon et al., 2015).

Network-aware approaches show strong promise when fraudsters act in collusive structures. Graph analytics and community detection have improved the identification of organized schemes beyond transaction-level red flags (Akoglu et al.,

2015). In payment streams analogous to public disbursements, extending feature spaces with relational signals (shared identifiers, devices, or vendors) has yielded meaningful gains over standalone classifiers (Van Vlasselaer et al., 2015). At the same time, hybrid pipelines that combine unsupervised outlier scoring for candidate selection with delayed supervised feedback help cope with class imbalance and label latency, conditions common in audit and investigative workflows (Carcillo et al., 2018).

From an assurance perspective, process-analytic methods leverage event logs to detect control deviations and segregation-of-duties violations, supporting continuous auditing in complex procurement-to-pay and benefits-payment cycles (Jans et al., 2014). Big-data evidence, logs, metadata, and digital exhaust, can complement traditional vouchers, but raises issues of provenance, explainability, and auditability; empirical work in auditing stresses the need for interpretable features and traceable model decisions to sustain evidentiary value (Yoon et al., 2015). Comprehensive surveys of fraud systems emphasize operational risks, data drift, feature brittleness, and adversarial adaptation, and recommend ongoing model monitoring and periodic recalibration in production environments (Abdallah et al., 2016; West & Bhattacharya, 2016).

Overall, the literature points to three converging directions for PFM fraud analytics: (i) hybrid architectures that blend anomaly detection with supervised learning under imbalance; (ii) graph-enriched representations to surface collusion; and (iii) process mining and explainable modeling to align detection with audit standards and accountability requirements (Jans et al., 2014; Akoglu et al., 2015).

These strands collectively inform designs that are technically robust yet auditable within the governance constraints of the public sector.

3. Methods

This study employed a systematic literature review (SLR) approach to synthesize existing research on the application of big data analytics in fraud detection within public financial management (PFM). The review focused on peer-reviewed journal articles and high-quality conference papers published in 2014-2023 that discuss analytical techniques, implementation contexts, and performance outcomes relevant to fraud detection in the public sector.

Relevant studies were identified through targeted searches in major academic databases, including Scopus, Web of Science, and Google Scholar, using combinations of keywords such as "big data analytics", "fraud detection", "public financial management", and "public sector auditing". Boolean operators and filters were applied to refine results, ensuring that only empirical and theoretical studies of sufficient quality were considered.

The inclusion criteria focused on studies that directly addressed the use of advanced analytics for fraud detection in public finance or closely related domains such as governmental auditing and public-sector procurement. Exclusion criteria removed duplicates, non-peer-reviewed sources, and works without substantial analytical content. Selected studies were reviewed in full, with data extracted on the analytical methods applied, fraud detection performance, contextual challenges, and

governance considerations. A thematic synthesis approach was then used to organize findings into coherent categories aligned with the research objectives.

4. Results and Discussion

The systematic review reveals that big data analytics has significantly enhanced fraud detection capabilities in public financial management by enabling governments to process large volumes of heterogeneous financial data in real time. Several studies indicate that the integration of advanced analytical tools such as machine learning, artificial intelligence (AI), and predictive modeling into public sector auditing allows auditors and oversight bodies to identify anomalous transactions that traditional auditing methods might overlook (Bierstaker et al., 2014; Omar et al., 2014). By leveraging structured and unstructured data from diverse sources such as procurement records, tax filings, and expenditure logs. These systems can detect irregular spending patterns that could indicate fraudulent activities.

The findings also suggest that the predictive and prescriptive capabilities of big data analytics are particularly valuable in the prevention stage of fraud management. Instead of focusing solely on post-fraud investigations, analytics tools enable proactive monitoring, reducing the financial and reputational losses associated with fraud incidents (Vasarhelyi et al., 2015). For example, anomaly detection algorithms applied in real time can flag transactions that deviate from established norms, prompting immediate investigation by auditors. This represents a shift from reactive to preventive control frameworks in PFM.

Despite these benefits, the review identifies significant implementation challenges. One recurring theme is the issue of data quality and interoperability between government departments. Many public agencies operate legacy financial management systems that are not fully compatible with modern analytics tools, resulting in fragmented and incomplete datasets (Mouzakitis et al., 2017). Without comprehensive and clean datasets, even sophisticated algorithms may produce false positives or fail to detect subtle fraudulent activities. Additionally, concerns around data privacy and security emerge as critical governance issues. Ensuring compliance with data protection regulations while maintaining analytical capability requires robust legal and technical safeguards (Al-Htaybat & von Alberti-Alhtaybat, 2017).

Another important finding relates to the skill gaps in the public sector. While private sector organizations often have dedicated data science teams, many government audit offices lack staff with the technical expertise needed to deploy and interpret advanced analytics tools effectively (Appelbaum et al., 2017). Capacity-building initiatives, such as specialized training in data analytics for auditors and collaboration with external analytics experts, are essential to bridging this gap. Moreover, political will and institutional support are necessary to ensure the adoption and sustained use of analytics-based fraud detection systems in PFM.

The review also reveals that the successful application of big data analytics in fraud detection requires strong governance frameworks that clearly define roles, responsibilities, and accountability mechanisms. Integrating analytics into existing audit and anti-corruption frameworks without proper policy alignment can lead to operational inefficiencies and underutilization of technological investments (Krahel

& Vasarhelyi, 2014). As such, governance reforms and process re-engineering are often required to maximize the benefits of analytics-driven oversight.

Interestingly, several studies highlight that big data analytics not only improves the detection of financial fraud but also enhances overall fiscal transparency. By making certain analytical outputs publicly available such as summary reports on anomalies or risk-prone sectors governments can foster citizen trust and encourage public participation in fiscal oversight (Meijer, 2014). This creates a feedback loop where increased transparency discourages fraudulent practices, thus reinforcing the effectiveness of PFM systems.

In summary, the results indicate that big data analytics holds considerable promise in strengthening fraud detection within public financial management. However, technological, organizational, and regulatory challenges must be addressed to realize its full potential. Addressing these barriers requires a holistic approach that combines investment in technology infrastructure, capacity development for audit personnel, inter-agency data integration, and the creation of robust governance frameworks. The discussion underscores that while the adoption of big data analytics is a powerful step toward more transparent and accountable PFM, its effectiveness ultimately depends on the broader institutional and political context in which it is implemented.

5. Conclusion

This study highlights the transformative role of big data analytics in enhancing fraud detection within public financial management. By enabling the processing of

large and diverse datasets in real time, big data technologies significantly improve the identification of irregularities that may signal fraudulent activities. The shift from reactive detection to proactive prevention represents a major advancement in public sector oversight, offering the potential to reduce both financial losses and reputational risks.

However, the findings underscore that technology alone cannot ensure effective fraud detection. Issues such as data fragmentation, lack of interoperability, insufficient technical skills, and concerns over data privacy present significant barriers to implementation. Overcoming these challenges requires strategic investments in infrastructure, the development of technical capacity among audit professionals, and the establishment of robust governance frameworks that balance analytical capability with regulatory compliance.

Ultimately, the adoption of big data analytics must be supported by political will, inter-agency collaboration, and a culture of transparency to deliver its full benefits. When integrated into broader fiscal governance systems, these tools not only enhance fraud detection but also strengthen public trust in financial management processes. The results of this review suggest that big data analytics, if properly implemented, can serve as a cornerstone for more accountable, transparent, and resilient public financial systems.

References

Abdallah, A., Maarof, M. A., & Zainal, A. (2016). Fraud detection system: A survey. Journal of Network and Computer Applications, 68, 90-113.

Achakzai, M. A. K., & Peng, J. (2023). Detecting financial statement fraud using dynamic ensemble machine learning. International Review of Financial Analysis, 89, 102827.

Akoglu, L., Tong, H., & Koutra, D. (2015). Graph based anomaly detection and description: a survey. Data mining and knowledge discovery, 29(3), 626-688.

Albashrawi, M. (2016). Detecting financial fraud using data mining techniques: A decade review from 2004 to 2015. Journal of Data Science, 14(3), 553-569.

Al-Htaybat, K., & von Alberti-Alhtaybat, L. (2017). Big Data and corporate reporting: impacts and paradoxes. Accounting, auditing & accountability journal, 30(4), 850-873.

Appelbaum, D., Kogan, A., & Vasarhelyi, M. A. (2017). Big data and analytics in the modern audit engagement: Research needs. Auditing: A Journal of Practice & Theory, 36(4), 1-27.

Bierstaker, J., Janvrin, D., & Lowe, D. J. (2014). What factors influence auditors' use of computer-assisted audit techniques?. Advances in Accounting, 30(1), 67-74.

Cao, M., Chychyla, R., & Stewart, T. (2015). Big data analytics in financial statement audits. Accounting horizons, 29(2), 423-429.

Carcillo, F., Dal Pozzolo, A., Le Borgne, Y. A., Caelen, O., Mazzer, Y., & Bontempi, G. (2018). Scarff: a scalable framework for streaming credit card fraud detection with spark. Information fusion, 41, 182-194.

Fernández, A., García, S., Galar, M., Prati, R. C., Krawczyk, B., & Herrera, F. (2018). Learning from imbalanced data sets (Vol. 10, No. 2018, p. 4). Cham: Springer.

Jans, M., Alles, M. G., & Vasarhelyi, M. A. (2014). A field study on the use of process mining of event logs as an analytical procedure in auditing. The Accounting Review, 89(5), 1751-1773.

Krahel, J. P., & Vasarhelyi, M. A. (2014). AIS as a facilitator of accounting change: Technology, practice, and education. Journal of Information Systems, 28(2), 1-15.

Meijer, A. (2014). Transparency.

Mohammadi, M., Yazdani, S., Khanmohammadi, M. H., & Maham, K. (2020). Financial reporting fraud detection: An analysis of data mining algorithms. International Journal of Finance & Managerial Accounting, 4(16), 1-12.

Mouzakitis, S., Papaspyros, D., Petychakis, M., Koussouris, S., Zafeiropoulos, A., Fotopoulou, E., ... & Psarras, J. (2017). Challenges and opportunities in renovating public sector information by enabling linked data and analytics. Information Systems Frontiers, 19(2), 321-336.

Omar, N., Koya, R. K., Sanusi, Z. M., & Shafie, N. A. (2014). Financial statement fraud: A case examination using Beneish Model and ratio analysis. International Journal of Trade, Economics and Finance, 5(2), 184.

Pamisetty, V. (2021). Integrating Predictive Analytics and IT Infrastructure for Advanced Government Financial Management and Fraud Detection. Available at SSRN 5275676.

Van Vlasselaer, V., Bravo, C., Caelen, O., Eliassi-Rad, T., Akoglu, L., Snoeck, M., & Baesens, B. (2015). APATE: A novel approach for automated credit card

transaction fraud detection using network-based extensions. Decision support systems, 75, 38-48.

Vasarhelyi, M. A., Kogan, A., & Tuttle, B. M. (2015). Big data in accounting: An overview. Accounting Horizons, 29(2), 381-396.

West, J., & Bhattacharya, M. (2016). Intelligent financial fraud detection: a comprehensive review. Computers & security, 57, 47-66.

Yoon, K., Hoogduin, L., & Zhang, L. (2015). Big data as complementary audit evidence. Accounting Horizons, 29(2), 431-438.