INTERNATIONAL JOURNAL OF BUSINESS AND ECONOMICS INSIGHTS

Volume: 16; **Issue:** 09
Pages: 01– 26
Accepted: 28 August 2026
Published: 05 September 2026

Data-Driven Financial Management and System Optimization in Public Sector Agencies: A Systematic Literature Review

S M Arif Al Sany[1];

[1]. *Senior Accountant, New Mexico Department of Health (DOH), Santa Fe, New Mexico; USA; M.S. in Management Information Systems, Lamar University, Texas, USA;*
Email: alsany25@gmail.com

DOI: 10.63125/xqd46597
This work was peer-reviewed under the editorial responsibility of the IJBEI, 2026

Abstract

Public agencies increasingly expect analytics and machine learning to improve budgeting, revenue collection, procurement and audit, but the evidence is scattered across disciplines. This systematic literature review, reported with reference to PRISMA 2020, maps peer-reviewed journal articles (2010 to July 2026) on data-driven financial management and financial system optimisation in the public sector. Thirteen searches of the open index OpenAlex returned 7,462 records; 1,157 were screened by title, 84 reports were assessed against their Crossref records and abstracts, and 66 studies were included. Procurement and contracting (21 studies), budgeting, forecasting and fiscal risk (17) and tax administration (13) dominate, followed by audit, accounting and governance (8) and financial management information systems (7); 48% of the studies date from 2025 or 2026. Just over half are predictive modelling studies. Those that state a figure often report high accuracy in detecting collusion, corruption risk and tax non-compliance, but only 15 abstracts compare machine learning with a conventional method: 9 favour machine learning, 4 are mixed and 2 are unfavourable, the last two groups all in budgeting, forecasting and fiscal risk. Financial outcomes such as recoveries or savings, and operational deployment, are rarely reported. Studies of financial management information systems point to integration with treasury operations, continuity of configuration and user involvement, more than software, as the conditions for results, and adoption studies in audit institutions and finance departments point to skills, management support and ethical safeguards. Reporting is incomplete: 37 abstracts omit the sample size and 15 do not name the jurisdiction. We propose an interpretive framework linking data foundations, analytics, decision use and outcomes, and a research agenda centred on outcome evaluation, honest baselines and reporting standards. The review is based on abstracts and bibliographic records screened and coded once, with a language-model tool and without verification by a human reviewer, in one open database, so it is a transparent, extensible map of the literature, not an exhaustive or quality-appraised synthesis.

Keywords

Public financial management; Data analytics; Machine learning; Budgeting and forecasting; Public procurement; Tax administration; Financial management information systems; Public sector audit; Systematic literature review; PRISMA

INTRODUCTION

Public sector agencies raise, allocate, spend and account for money through a chain of functions (budget preparation, revenue collection, procurement, payment, accounting, reporting and audit) that now runs almost entirely on digital systems. Financial management information systems record budgets and transactions, e-procurement portals publish tenders and awards, and tax administrations receive returns and invoices electronically. These systems produce the data on which forecasting, risk scoring and anomaly detection depend, and they have encouraged the expectation that analytics and machine learning can make public financial management more accurate, more efficient and harder to defraud. Statistical fraud detection is long established (Bolton & Hand, 2002), and recent studies have applied machine learning to fiscal distress (Liu et al., 2025; Radebe et al., 2026), revenue forecasting (Chung et al., 2022), procurement collusion (Fazekas et al., 2026; García Rodríguez et al., 2022) and tax compliance (Khreis et al., 2026b; Tuyishimire & Murorunkwere, 2024).

The evidence, however, is scattered across public administration, accounting, economics, computer science and information systems journals, and it is uneven. Some studies report high predictive accuracy; others find that machine learning does no better than traditional forecasting methods (Chung et al., 2022; Larson & Overton, 2024) or fails to predict the event of interest (Afonso & Sanchez Gomez, 2026). Studies of financial information systems describe implementations that are constrained less by technology than by organisation and politics (Gcora & Chigona, 2019; Simpson et al., 2020). Existing reviews are specific to one function, such as fraud detection in procurement (Schneider dos Santos et al., 2025) or in taxation (Belahouaoui & Alm, 2025). A manager or researcher who wants to know what data-driven financial management has been shown to achieve in government, in which functions and under what conditions, has, as far as we could find, no single synthesis to turn to.

This article provides one. It is a systematic literature review, reported with reference to the PRISMA 2020 statement (Page et al., 2021), of peer-reviewed journal articles published from 2010 to July 2026 on data-driven methods and financial system optimisation in public sector financial management. It addresses four research questions:

RQ1. Which public financial management functions have been addressed with data-driven methods, in which settings, and with what study designs and methods?

RQ2. What performance and benefits are reported, and how do machine-learning methods compare with conventional ones where both are tested?

RQ3. What does the evidence show about financial management information systems and the conditions under which they improve financial management?

RQ4. Which organisational, governance and ethical conditions enable or constrain adoption?

The review contributes (1) a documented search log and selection record based on open bibliographic infrastructure; (2) a coded dataset of 66 studies with their settings, designs, methods and stated results; (3) a synthesis across five functions (budgeting, forecasting and fiscal risk; procurement and contracting; tax and revenue administration; financial management information systems; and audit, accounting and governance); and (4) an interpretive framework and research agenda. The review is based on bibliographic records and abstracts, not on full texts, and we state throughout what that does and does not allow us to conclude.

BACKGROUND AND SCOPE

Data-driven financial management

We use data-driven financial management to mean the use of recorded financial and administrative data, analysed with statistical forecasting, machine learning, data or text mining, network analytics or optimisation, to inform or automate decisions in the public financial management cycle. Three kinds of task recur. Forecasting tasks predict a quantity, such as tax revenue or expenditure. Early-warning tasks predict an event, such as fiscal distress of a local government or default on a tax obligation. Screening tasks rank many cases, such as tenders, suppliers or tax returns, so that limited audit or investigative capacity is directed to the cases most likely to matter. Screening tasks are usually highly imbalanced, because the event of interest is rare, which makes the choice of evaluation measure important (He & Garcia, 2009; Saito & Rehmsmeier, 2015); and predicted probabilities are useful for allocating effort only if they are calibrated (Van Calster et al., 2019).

System optimisation

By system optimisation we mean the design, integration and use of the information systems on which financial management runs: integrated financial management information systems (IFMIS or FMIS), enterprise resource planning systems, treasury single accounts, e-procurement and e-invoicing platforms and audit management systems. The term covers both technical configuration (coverage of functions, interoperability) and the organisational process of implementation. We treat these systems as part of the same subject as analytics because they determine which data exist, at what quality, and whether an analytical result can reach a decision.

Explainability and fairness

Financial decisions in government are subject to legal and democratic accountability. A model that scores taxpayers, suppliers or municipalities has consequences for those scored, so interpretable models have been argued to be preferable in high-stakes settings when they perform as well as opaque ones (Rudin, 2019), and post hoc explanation methods such as Shapley-value attributions are widely used with tree-based models (Lundberg et al., 2020; Md Mohsin & Jannatul, 2021; Punnaiah Chowdary & Md, 2020; Punnaiah Chowdary & Md Jamil, 2020). Evidence from other sectors shows that algorithms trained on administrative proxies can reproduce existing disparities (Asheke & Tashriq Ahmed Bin, 2022; Kazi Mohammad Khalid & Sumaiya Sadmeen, 2022; Obermeyer et al., 2019; Punnaiah Chowdary, 2021; Rajkomar et al., 2018). We therefore recorded whether the included studies addressed explanation, oversight and the position of those assessed.

METHODOLOGY

Design and reporting

The review followed the stages of a systematic review in management research (planning, searching, screening, extraction and synthesis) (Tranfield et al., 2003) and is reported with reference to PRISMA 2020 (Page et al., 2021). The research questions, the information source and the main eligibility criteria were set before screening. The search strings were developed iteratively: searches S11 to S13 were added after the first ten showed that information systems, audit and revenue forecasting were poorly covered, and the requirement that an abstract be available was added during eligibility assessment. The protocol was not registered. All searching, screening and coding were done once, on 17 July 2026, with a language-model tool acting as the single reviewer; no human reviewer screened or coded the records, there was no second reviewer and no test of agreement. The review follows the structure of a systematic review but, for the reasons set out under Limitations, does not meet all PRISMA items.

Information source and search strategy

The subscription databases commonly used for reviews (Scopus and Web of Science) were not accessible to us. We searched OpenAlex, an open index of scholarly works, through its public API, and we verified every report assessed for eligibility against its record in Crossref, the DOI registration agency used by most journal publishers. Thirteen searches were run (Table 1). They combined terms for the public sector or a public finance function with terms for data-driven methods or systems, and used three OpenAlex search fields: the title; the title and abstract; and all indexed text, the last restricted to sources that OpenAlex classifies as core. All searches were limited to works of type article. OpenAlex returns results in order of relevance; for searches that returned more than 100 records, the first 100 were screened. Exploratory searches run while developing the strings are not counted.

Table 1: Search log

Search	Field	Query	Filters	Records returned	Screened
S1	Title	budget machine learning	article; English; 2010 to 2026	82	82
S2	Title and abstract	"public sector" "machine learning" budget	article	36	36
S3	Title and abstract	"public sector" "artificial intelligence" budget	article	109	100

Search	Field	Query	Filters	Records returned	Screened
S4	Title and abstract	"public procurement" "machine learning"	article	87	87
S5	Title and abstract	"tax administration" "machine learning"	article	132	100
S6	Title and abstract	"public finance" "machine learning"	article	68	68
S7	All indexed text	"public procurement" "machine learning"	article; core source	1,828	100
S8	All indexed text	"tax administration" "machine learning"	article; core source	530	100
S9	All indexed text	"public sector" budgeting "machine learning"	article; core source	3,894	100
S10	All indexed text	"local government" "fiscal distress" prediction	article; core source	84	84
S11	Title and abstract	"financial management information system" government	article	332	100
S12	Title and abstract	"public sector" audit "data analytics"	article	104	100
S13	Title and abstract	government "revenue forecasting"	article	176	100
Total				7,462	1,157

Eligibility criteria

Table 2 lists the criteria. We included journal articles with a Crossref DOI, published from 2010, in English, that study a public financial management function in a government setting and have a data-driven method or a financial information system as their object. Empirical studies, system demonstrations and literature reviews with a stated method were eligible. Because the review is based on abstracts, a report was eligible only if an abstract was available in its Crossref or OpenAlex record.

Table 2: Eligibility criteria

Criterion	Included	Excluded
Publication type	Article in a scholarly journal with a DOI registered at Crossref	Repository deposits, working papers, reports, book chapters, conference papers, theses
Period and language	Published 2010 to July 2026; English	Earlier publications; other languages
Setting	Government bodies at any level, public agencies, supreme audit institutions, tax and procurement authorities	Private-sector finance; corporate budgeting; household finance
Function	Budgeting and forecasting, fiscal risk, revenue and tax administration, procurement and contract spending, accounting, reporting, audit, financial management information systems	Macroeconomic or policy-evaluation studies that use fiscal variables without a management application
Approach	Data-driven method (statistical forecasting, machine learning, data or text mining, network analytics, optimisation) or a financial information system, as the object of study	Studies in which data or systems are only background
Evidence	Empirical study, system demonstration or literature review with a stated method	Commentaries and conceptual papers without a stated method
Information available	Abstract available in the Crossref or OpenAlex record	No abstract in either source

Selection process

Records were screened by title, one search at a time. No record was kept of the individual title decisions. Records were excluded at this stage if the title showed that the work was outside the public sector or outside financial management, or was not a journal article; repository deposits, working papers and conference papers were excluded, as were records from sources that the reviewer judged not to be established scholarly journals. This last judgement was made from the title and DOI prefix and is a source of subjectivity. Duplicates were removed after title screening, because each search was screened separately; the number of duplicate records was therefore not counted separately from the number of title exclusions. For every retained report we retrieved the Crossref record (title, authors, journal, volume, issue, pages and year) and the abstract. Where Crossref held no abstract, we retrieved the abstract from the OpenAlex record of the same DOI. Eligibility was then assessed from the bibliographic record and the abstract.

Data extraction and coding

For each included study we coded, from the abstract: the function (five categories); the study design (predictive modelling; system or tool design; statistical or econometric analysis; survey; qualitative or mixed methods; review); the methods named; the setting (country or group of countries) and region; whether the sample size or data volume was stated; whether a quantitative performance or effect measure was stated; the main result; and, where the abstract compared machine learning with a conventional or simpler method, the verdict (better; dependent on task, data or horizon; no advantage or failure). Settings were coded only when the abstract named a country, or a city, region or national agency from which the country is evident (for example Seoul, North Carolina, Migori County or the Swiss Federal Railways); we did not infer a country from the authors' affiliations. For one study the only text available was a publisher's impact statement. The verdict was coded only where the abstract states explicitly how a machine-learning method performed against a named conventional or simpler method (regression, an econometric or hedonic model, a naive or rule-based benchmark); this coding involves judgement. Abstracts were retrieved through the two APIs and condensed into structured digests (setting, data, methods, results) by an automated language-model step before coding; the digests were not checked against the publishers' versions of the abstracts. Bibliographic details are those of the Crossref records.

Appraisal of reporting

A formal risk-of-bias assessment requires full texts and was not attempted. Instead, we recorded two indicators of reporting completeness at abstract level, namely whether the abstract states the sample size or data volume and whether it states a quantitative result, and we report them by study design. They describe how much can be learned from an abstract, not the quality of the underlying study.

Synthesis

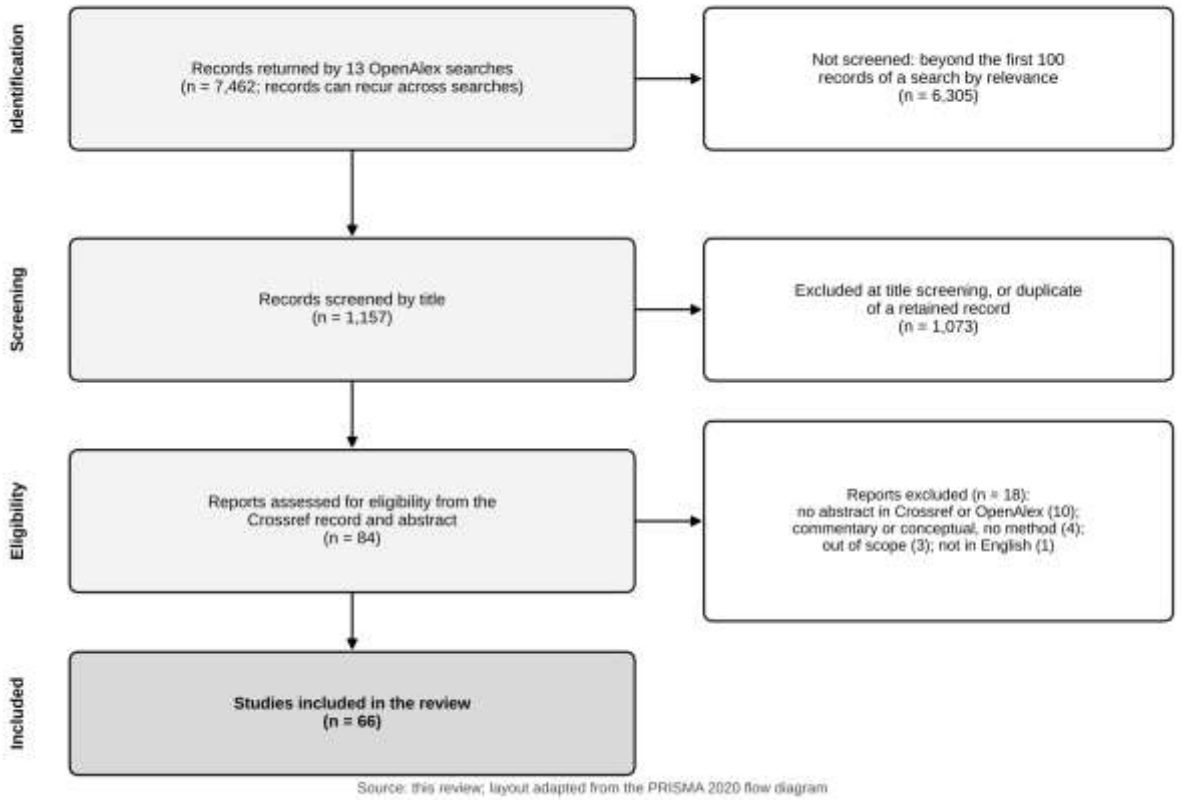
The studies are too heterogeneous in function, outcome and measure for meta-analysis. We combined a descriptive synthesis (counts by function, design, setting and year) with a thematic synthesis (Braun & Clarke, 2006) of the stated results, organised by the four research questions. The coded dataset, the search log and the list of excluded reports with reasons are provided as supplementary material.

RESULTS

Study selection

The 13 searches returned 7,462 records, counting a record once for each search in which it appeared (Figure 1). Of these, 1,157 were screened by title; the remaining 6,305 lay beyond the first 100 results of searches S3, S5, S7 to S9 and S11 to S13 and were not screened. Title screening and removal of duplicates left 84 unique reports, all of which were retrieved from Crossref. 18 were excluded at eligibility assessment: 10 because neither Crossref nor OpenAlex held an abstract, 4 because they were commentaries or conceptual papers without a stated method, 3 because they were out of scope, and 1 because it was not in English (Supplementary Table S1). The review includes 66 studies. The abstract came from the Crossref record for 47 studies and from the OpenAlex record for 19.

Figure 1: Flow of records through the review |



Characteristics of the included studies

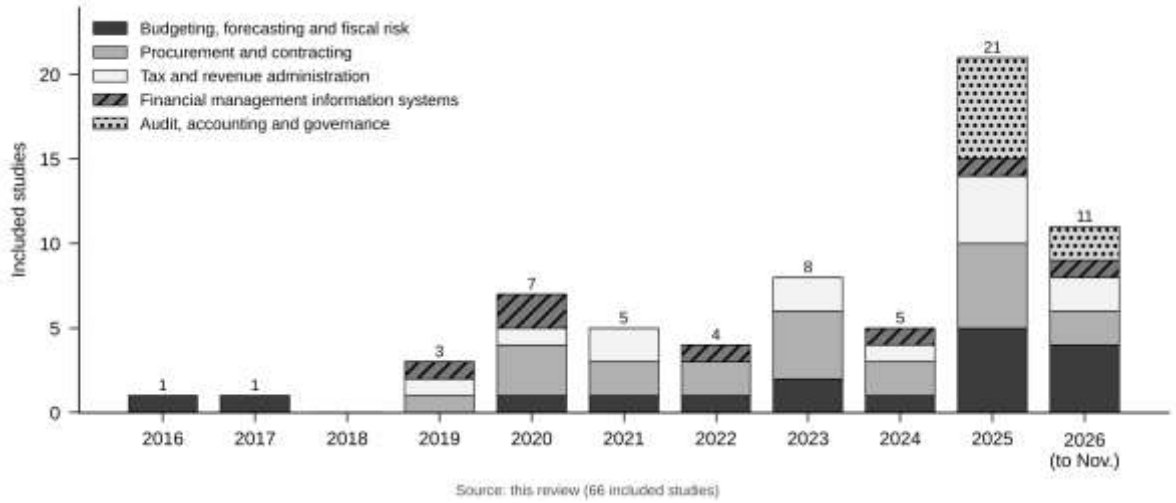
Table 3 summarises the studies. Procurement and contracting was the largest group (21 studies), followed by budgeting, forecasting and fiscal risk (17), tax and revenue administration (13), audit, accounting and governance (8) and financial management information systems (7). The literature is recent: 45 of the 66 studies (68%) were published in 2023 or later and 32 (48%) in 2025 or 2026 (Figure 2). The studies appeared in 50 journals; the most frequent was the Journal of Public Budgeting, Accounting & Financial Management (6 studies), followed by the Journal of Risk and Financial Management (4).

Table 3: Characteristics of the 66 included studies

Characteristic	Studies	% of 66
Function		
Procurement and contracting	21	32
Budgeting, forecasting and fiscal risk	17	26
Tax and revenue administration	13	20
Audit, accounting and governance	8	12
Financial management information systems	7	11
Study design		
Predictive modelling	34	52
System or tool design	9	14
Qualitative or mixed methods	8	12
Statistical or econometric	7	11
Survey	6	9
Review	2	3

Characteristic	Studies	% of 66
Setting stated in the abstract		
Europe	19	29
Africa	6	9
Asia	6	9
Multi-country	5	8
Latin America	5	8
Middle East	5	8
North America	3	5
Not stated	15	23
Not applicable	2	3
Publication year		
2016 to 2019	5	8
2020 to 2022	16	24
2023 to 2024	13	20
2025 to July 2026	32	48
Abstract source		
Crossref record	47	71
OpenAlex record	19	29

Figure 2: Included studies by year of publication and function |



Just over half of the studies (34) were predictive modelling studies and 9 designed or demonstrated a system or tool; the remainder were qualitative or mixed-methods studies (8), statistical or econometric analyses (7), surveys (6) and reviews (2). Design differed sharply by function (Figure 3): procurement, tax and budgeting studies were mostly predictive, whereas all 7 information-system studies and all 8 audit, accounting and governance studies used qualitative, survey or statistical designs. Among the 49 studies whose abstract named a setting, Europe was the most common (19), followed by Africa and Asia (6 each) and multi-country studies (5); only 3 abstracts named a North American setting, and 15 named none (Figure 4).

Figure 3: Number of included studies by function and study design |

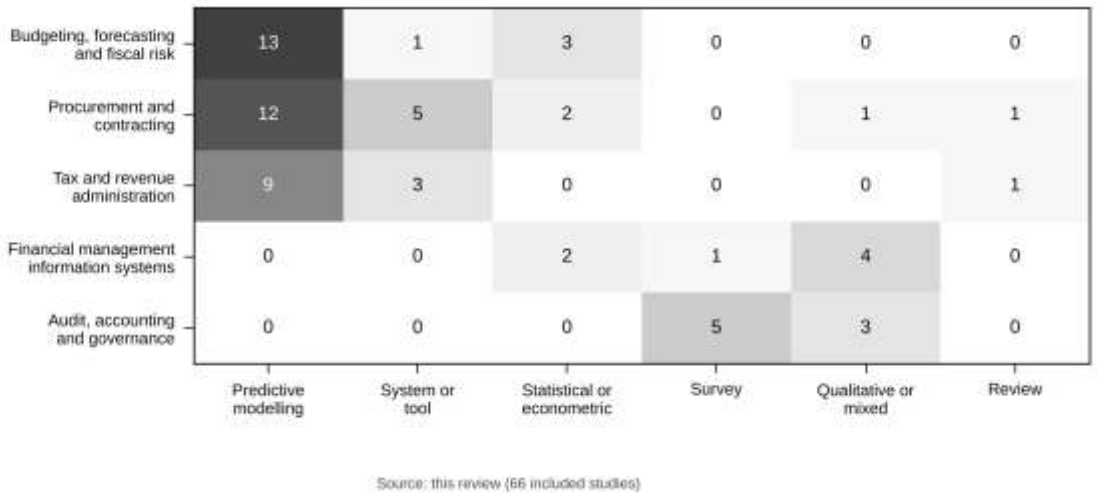
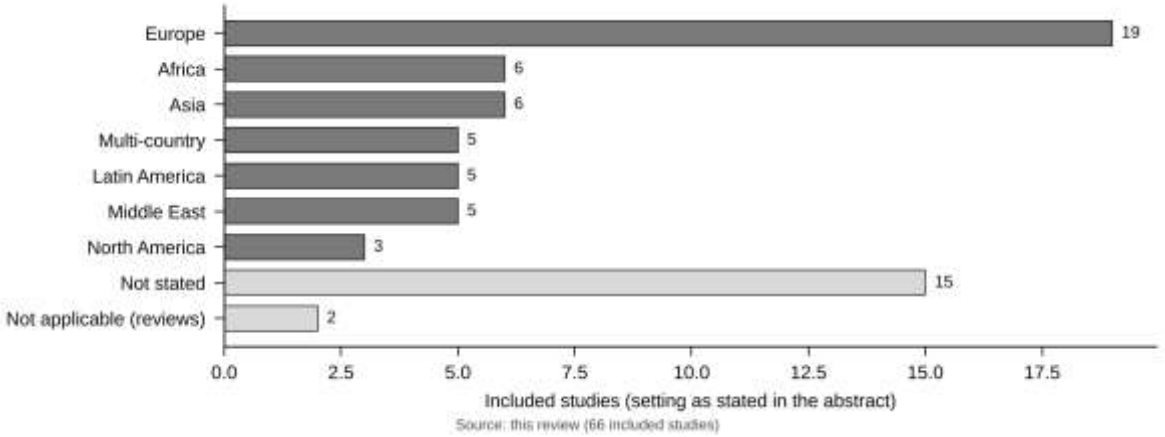


Figure 4: Setting of the included studies as stated in the abstract |



RQ1: Functions, data and methods

Budgeting, forecasting and fiscal risk

The 17 studies in this group fall into three strands (Table 4). The first predicts fiscal distress or deficits of subnational governments. Gorina et al. (2017) built an action-based measure of local fiscal distress and found reserves, debt and revenue composition to be the more important predictors. Later studies applied machine-learning classifiers to local governments in the United States (Liu et al., 2025), Indonesia (Agus et al., 2025; Aminul Islam & Md, 2023; Mashfiqur & Md Jamil, 2023; Md Maksudul & Md Tazul Islam, 2022) and South Africa (Md & Kynatun, 2023; Md Maksudul & Md Takbir Hossen, 2023; Md Mohsin & Mst Kaniz, 2023; Radebe et al., 2026), and to compliance of European Union member states with the deficit limit of the Stability and Growth Pact (Baret et al., 2023; Md. Mamunur & Md, 2023; Sadia & Md. Mosheur, 2023; Sharmin & Mst Kaniz, 2023). Against these, Afonso and Sanchez Gomez (2026) ran almost 500 models on 100 counties and report that none predicted which counties would run deficits in a recession.

The second strand forecasts revenues and expenditures. Williams and Kavanagh (2016) compared forecasting methods on 55 revenue series of 18 local governments and found forecasting software, damped trend and simple exponential smoothing to perform best on monthly and quarterly data; Batóg and Batóg (2021) report for one Polish region that less comprehensive methods can be reasonably accurate. Machine-learning forecasts of local revenue did not, overall, beat traditional statistical methods in Chung et al. (2022), and Larson and Overton (2024) report that pre-processing mattered more than the choice of model for sales tax forecasts. Drăgulin et al. (2025) report better expenditure

forecasts from neural and deep-learning models than from regression, without figures in the abstract, and Chung et al. (2025) found that a large language model produced unreliable revenue forecasts on its own but errors as low as 9.9% with a human in the loop.

The third strand uses analytics to understand or support budget decisions: classification of more than 25,000 citizen proposals in participatory budgeting in Seoul (Gazi Yeasin, 2024b; Sharmin & Nishat, 2023; Shin, 2025; Sumaiya Sadmeen & Kazi Mohammad Khalid, 2023); interpretable models of economies of scale in Brazilian municipal expenditure (Gazi Yeasin, 2024a; Maishar Mostari, 2024; Mashfiqur, 2024; Toffaha et al., 2026); prediction of regional budget execution in Peru (Rocha et al., 2020); a comparison of regression, machine learning and agent computing for linking expenditure programmes to development indicators (Guariso et al., 2023; Masrul & Mashfiqur, 2024; Masrul & Mukut Kanti, 2024; Md Hasinur, 2024); and an AI-assisted fiscal monitoring framework tested on Saudi Arabian budget performance reports (Hashim, 2026; Md Maksudul, 2024; Md Mohsin, 2024; Md Tazul Islam, 2024).

Table 4: Budgeting, forecasting and fiscal risk: included studies

Study	Setting	Design and methods	Main result stated in the abstract
Williams & Kavanagh (2016)	Not stated	Statistical or econometric. Exponential smoothing; damped trend; naive methods	18 local governments, 55 revenue series; simple smoothing methods performed best on monthly and quarterly data
Gorina et al. (2017)	Not stated	Statistical or econometric. Action-based distress measure; regression	Fiscal reserves, debt and revenue composition were the more important predictors of distress
Rocha et al. (2020)	Peru	Predictive modelling. Supervised regression	25 governors; profile indicators explained 72.6% of variance in budget execution
Batóg & Batóg (2021)	Poland	Statistical or econometric. Time-series and causal models	One region, 2000 to 2018; less comprehensive methods gave reasonably accurate estimates
Chung et al. (2022)	Not stated	Predictive modelling. Several ML algorithms; traditional forecasters	Traditional statistical methods outperformed ML overall; k-nearest neighbours did better for property tax
Baret et al. (2023)	European Union	Predictive modelling. Support vector machine; logit benchmark	28 member states (2006 to 2018); eight of 138 predictors; nearly 92% accuracy, above logit
Guariso et al. (2023)	Not stated	Predictive modelling. Regression; machine learning; agent computing	Regression and machine learning struggled to separate sector effects; agent computing was more informative
Larson & Overton (2024)	Not stated	Predictive modelling. ML and traditional sales tax forecasts	ML did not always improve accuracy; data pre-processing made the largest difference
Agus et al. (2025)	Indonesia	Predictive modelling. Seven classifiers by cluster	542 governments (2013 to 2022); logistic regression among the best classifiers; random forest never selected
Chung et al. (2025)	Not stated	Predictive modelling. Large language model; human in the loop	LLM with human review reached errors as low as 9.9% (annual aggregate); LLM alone was unreliable
Drăgulin et al. (2025)	Not stated	Predictive modelling. Neural networks; SVM; deep learning	AI models reported as more accurate than regression and econometric forecasts; no figures in abstract
Liu et al. (2025)	United States	Predictive modelling. Six algorithms; resampling for imbalance	49 states (2015 to 2017); balanced bagging generally best, F1 53.11% to 55.33% against 27% to 58% for logistic regression

Study	Setting	Design and methods	Main result stated in the abstract
Shin (2025)	South Korea	Predictive modelling. Topic modelling; supervised classifier	More than 25,000 citizen proposals (2013 to 2021); 17 topics; selection classifier AUROC 0.73
Afonso & Sanchez Gomez (2026)	United States	Predictive modelling. LASSO; random forest; decision trees	100 counties, more than 900 variables, almost 500 models; none predicted which counties ran recession deficits
Hashim (2026)	Saudi Arabia	System or tool design. AI-assisted fiscal analysis within oversight workflow	81.9% temporal analytical consistency; labels were deterministic risk categories, not verified anomalies
Radebe et al. (2026)	South Africa	Predictive modelling. Penalised logistic regression; random forest; XGBoost	Out-of-time PR-AUC 0.934 to 0.954; top-30 precision 0.967 to 1.000 with recall 0.186 to 0.192
Toffaha et al. (2026)	Brazil	Predictive modelling. Gradient boosting; panel econometrics; SHAP	49,228 municipality-years; boosting best for unseen municipalities, fixed-effects predictor best for later years

Procurement and contracting

Procurement is the most developed application (Table 5). Its data are large and often open: studies used 58,337 and 102,087 Spanish tenders (García Rodríguez et al., 2019, 2020), more than 4 million European contracts (Md. Mainuddin, 2024; Md. Mamunur, 2024; Songeta, 2024; Wachs et al., 2020), almost 4 million United States federal purchase records (Aminul Islam, 2025; Muir & Reich, 2021; Sumaiya, 2024; Tahmedur, 2024) and more than 200,000 pharmaceutical purchases from ten countries and territories (Fazekas et al., 2024). The dominant task is the detection of collusion and corruption risk. García Rodríguez et al. (2022) tested eleven algorithms on data from five countries; Fazekas et al. (2026) combined price and bidding screens in a random forest and report 70% to 84% accuracy across 73 cartels in seven European countries, and 97% when training and testing within a single cartel case; Wallimann and Sticher (2023) adapted machine-learning screens to railway infrastructure tenders in Switzerland; and Lundberg (2025) evaluated five econometric screens on a Swedish asphalt cartel, of which the spatial methods flagged collusion and tests of bid differences did not. Corruption risk has been modelled from the content of Italian calls for tenders (Asheke, 2025; Ashik, 2025; Decarolis & Giorgiantonio, 2022; Kynatun, 2025), from the structure of contracting networks (Wachs et al., 2020), from sanctioned suppliers in Mexico with positive-unlabelled learning (Maishar Mostari, 2025; Masrul, 2025; Md Hasinur, 2025a; Medina-Hernández et al., 2026) and from rating parameters with clustering (Md Hasinur, 2025b; Md Maksudul, 2025; Md Momen Shahriar, 2025; Torres-Berru & López Batista, 2021); signals of collusion have been sought in the stability of supplier groups in Slovakia with causal forests (Demeter et al., 2025). Decarolis and Giorgiantonio (2022) add a caution: the red flags that are most obvious and most scrutinised were uncorrelated, or negatively associated, with corruption risk.

A second set of studies supports the management of contracts: risk of non-compliance with delivery terms in Brazil (Henrique et al., 2020; Md Tazul Islam, 2025; Md. Mamunur, 2025; Sadia, 2025), cost overruns and delays in an oversight tool built on open data (Salazar et al., 2024), procurements likely to end in disputes in Italy (Nai et al., 2023), outcomes predicted from the text of European tender notices (Acikalin et al., 2023; Shabbir, 2025; Sharmin, 2025), and transaction cost measures derived from contract text (Potoski et al., 2025). A third set builds decision-support tools: award price estimation and bidder recommendation (García Rodríguez et al., 2019, 2020; Songeta, 2025; Sumaiya Sadmeen, 2025; Tahmedur, 2025), price prediction for pharmaceuticals (Fazekas et al., 2024), classification of contract records for public reporting, embedded in a tool used by federal staff (Muir & Reich, 2021), and a natural-language system for searching and monitoring tenders (Siciliani et al., 2023). A systematic mapping of 93 papers found that machine learning is used mostly to detect collusion and statistical analysis to detect favouritism, and that real deployments and public datasets are scarce (Schneider dos Santos et al., 2025); interviews in government agencies likewise found low AI maturity in procurement (Andersson et al., 2025).

Table 5: Procurement and contracting: included studies

Study	Setting	Design and methods	Main result stated in the abstract
García Rodríguez et al. (2019)	Spain	Predictive modelling. Random forest regression	58,337 tenders worth EUR 31,426 million (2012 to 2018); award price estimator
García Rodríguez et al. (2020)	Spain	System or tool design. Random forest classifier	102,087 tenders, 1,353,213 companies; winner was in the recommended group in 24% to 38% of tenders
Henrique et al. (2020)	Brazil	Predictive modelling. Five classifiers	Contracts classified by risk of non-compliance; k-nearest neighbours best, above logistic regression and discriminant analysis
Wachs et al. (2020)	European Union	Statistical or econometric. Network science	More than 4 million contracts (2008 to 2016); centralised markets carried higher corruption risk
Muir & Reich (2021)	United States	System or tool design. Hierarchical classifiers; character-level CNN	Almost 4 million purchase records; hierarchical models beat flat ones; embedded in a tool used by federal staff
Torres-Berru & López Batista (2021)	Not stated	Predictive modelling. K-means; SOM; SVM; PCA	Four process clusters; semi-supervised anomaly model reached 95% accuracy
Decarolis & Giorgiantonio (2022)	Italy	Predictive modelling. Standard ML on tender-call content	The most scrutinised red flags were uncorrelated or negatively associated with corruption risk
García Rodríguez et al. (2022)	Brazil, Italy, Japan, Switzerland, United States	Predictive modelling. Eleven ML algorithms	Reasonable collusion detection rates even with minimal information; no figures in abstract
Acikalin et al. (2023)	European Union	Predictive modelling. Multilingual transformers; MLP; k-nearest neighbours	Notices in 22 languages; four outcomes predicted from call text; best model differed by outcome
Nai et al. (2023)	Italy	System or tool design. Classifiers; recommender system	Models to recognise procurements that led to disputes (2007 to 2022); similar-tender recommender
Siciliani et al. (2023)	Not stated	System or tool design. Natural language processing	Semantic search, summarisation, information extraction and dashboards for tender monitoring
Wallimann & Sticher (2023)	Switzerland	Predictive modelling. ML price screens	Screens adapted to railway tenders; tool for category managers
Fazekas et al. (2024)	Ten countries and territories	Predictive modelling. Random forest; OLS	More than 200,000 purchases; random forest R2 0.85, better than OLS
Salazar et al. (2024)	Not stated	System or tool design. ML models plus risk indices on open data	Tool flags cost overruns, delays and process irregularities for oversight agencies
Andersson et al. (2025)	Not stated	Qualitative or mixed methods. Interviews and workshop	18 interviews; generally low AI maturity; organisational, process, technological and data barriers
Demeter et al. (2025)	Slovakia	Predictive modelling. Causal forests	30,349 contracts; stable supplier portfolios achieved 71% lower savings than volatile ones
Lundberg (2025)	Sweden	Statistical or econometric. Five econometric screens	Asphalt cartel, 1995 to 2009; spatial methods flagged collusion in the cartel period; bid-difference tests did not

Study	Setting	Design and methods	Main result stated in the abstract
Potoski et al. (2025)	European Union	Predictive modelling. Text-based ML trained on a manager survey	Preregistered; public database of transaction cost measures for 176 products and services
Schneider dos Santos et al. (2025)	Not applicable	Review. Systematic mapping	93 of more than 6,000 papers; ML mostly for collusion, statistics for favouritism; few deployments, few public datasets
Fazekas et al. (2026)	Seven European countries	Predictive modelling. Random forest on price and bidding screens	73 cartels (2004 to 2021); 70% to 84% accuracy overall, 97% within a single cartel case
Medina-Hernández et al. (2026)	Mexico	Predictive modelling. Positive-unlabelled learning; network features; Shapley values	Outperformed approaches based only on red flags; at least 20% precision in the top 5% of predictions

Tax and revenue administration

The 13 tax studies concentrate on risk scoring for audit and collection (Table 6). They predict under-reporting of value-added tax and the reporting behaviour of electronic billing machines (Murorunkwere et al., 2023; Tuyishimire & Murorunkwere, 2024), corporate tax default in Finland (Abedin, Chi, et al., 2021; Abedin, Hassan, et al., 2021), compliance risk from tax records (Yang, 2025), and payment arrears and audit selection in Lebanon (Khreis et al., 2026a, 2026b). Several are systems built with or for a tax authority: a pipeline that combines a taxpayer network, machine learning and visual analytics, developed with the Italian revenue agency (Didimo et al., 2020); a four-module fraud detection framework tested on Saudi tax returns (Alsadhan, 2023); and a risk indicator for the Serbian tax administration (Atanasijević et al., 2019). Two studies address the administration itself: the efficiency of Indonesian tax offices (Soffan et al., 2025) and the valuation of dwellings for wealth taxation in Norway, where smaller geographic units and machine-learning clustering improved accuracy (Bø et al., 2025). A textometric review of the 2014 to 2024 literature proposes a framework for AI oversight in tax administration (Belahouaoui & Alm, 2025).

One result deserves emphasis because it links prediction to money. Khreis et al. (2026a) report that ranking tax obligations by predicted probability of arrears alone captured 0.4% of realised collection exposure in the top decile, against 92.7% when the ranking was weighted by the amount at stake. A model can therefore be accurate and still direct effort to cases that matter little.

Table 6: Tax and revenue administration: included studies

Study	Setting	Design and methods	Main result stated in the abstract
Atanasijević et al. (2019)	Serbia	System or tool design. Distance-based risk indicator	Risk indicator from the distance between an entity's income distribution and its sector average
Didimo et al. (2020)	Italy	System or tool design. Taxpayer network; ML; information diffusion; visual analytics	Four-step pipeline developed with the national revenue agency on real taxpayer data
Abedin et al. (2021a)	Finland	Predictive modelling. Feature transformation; gradient boosting	Transformations improved prediction; equity, liquidity and debt-to-sales ratios mattered most
Abedin et al. (2021b)	Finland	Predictive modelling. Seven ML approaches; feature transformation	Transformed and untransformed features compared on four datasets of corporate tax payment records

Study	Setting	Design and methods	Main result stated in the abstract
Alsadhan (2023)	Saudi Arabia	System or tool design. Tree model; anomaly scores; compliance score; fraud probability	Four-module framework tested on existing tax returns; no figures in abstract
Murorunkwere et al. (2023)	Not stated	Predictive modelling. Tree-based classifiers	Random forest most robust, 92.3% accuracy in predicting billing-machine reporting
Tuyishimire & Murorunkwere (2024)	Rwanda	Predictive modelling. Supervised classifiers	VAT under-reporting risk higher for cross-border, younger, trade and construction taxpayers
Belahouaoui & Alm (2025)	Not applicable	Review. Textometric systematic review	Literature of 2014 to 2024; proposes an adaptive AI tax oversight framework; governance and income-level differences
Bø et al. (2025)	Norway	Predictive modelling. Hedonic regression; ML clustering	Smaller geographic units and ML clustering substantially improved valuation accuracy for wealth taxation
Soffan et al. (2025)	Indonesia	Predictive modelling. Fuzzy clustering; DEA; neural network	225 of 352 tax offices efficient; optimised network cut mean squared error from 0.0144 to 0.0035
Yang (2025)	Not stated	Predictive modelling. SVM; XGBoost; random forest	3,232 tax records; random forest 92.00% (manufacturing) and 93.39% (services) accuracy
Khreis et al. (2026a)	Lebanon	Predictive modelling. Five classifiers; SHAP	10,625 obligations; XGBoost ROC-AUC 0.784, tree models above logistic regression; ranking by probability alone captured 0.4% of collection exposure against 92.7% when weighted by amount
Khreis et al. (2026b)	Lebanon	Predictive modelling. Ten classifiers; feature selection	4,063 SME and 1,903 large-taxpayer statements; random forest best; not yet deployed

RQ2: Reported performance and comparison with conventional methods

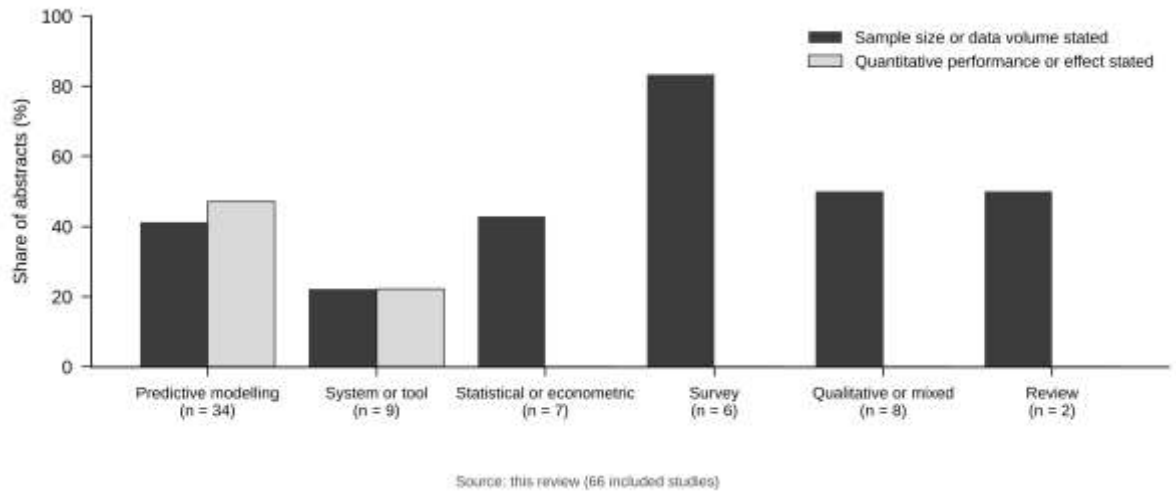
Predictive studies that state a result mostly report classification accuracy or a related measure, and the values are not comparable across studies because tasks, class balance and validation designs differ. Reported accuracies of 92% to 95% (Baret et al., 2023; Murorunkwere et al., 2023; Torres-Berru & López Batista, 2021; Yang, 2025) sit beside F1 scores of about 55% for fiscal distress (Liu et al., 2025) and an area under the ROC curve of 0.73 to 0.78 for proposal selection and tax arrears (Khreis et al., 2026a; Shin, 2025). Few abstracts describe a test on later periods; Radebe et al. (2026) and Toffaha et al. (2026) are examples that do, and the latter found that the best model for unseen municipalities was not the best for later years. Reporting is incomplete (Figure 5): only 29 of 66 abstracts (44%) state the sample size or data volume and 18 (27%) state a quantitative result; among the 43 predictive and system studies the figures are 16 and 18.

Only 15 abstracts state how machine learning compared with a conventional or simpler method (Table 7). In 9 it was reported as better; in 4 the answer depended on the task, the data or the horizon; and in 2 it gave no advantage or failed. The unfavourable and mixed verdicts all fall in the budgeting, forecasting and fiscal risk group; a possible reason, which the abstracts do not test, is that aggregate fiscal series are short and subject to shocks. Outcomes beyond predictive performance, such as revenue recovered, savings achieved or audit time saved, are almost never reported; the 71% lower savings of stable supplier groups in Slovakia (Demeter et al., 2025) and the collection exposure result of Khreis et al. (2026a) are exceptions.

Table 7: Abstracts that compare machine learning with a conventional or simpler method

Verdict in the abstract	Study	Function	Comparison
Machine learning reported as better (9)	Henrique et al. (2020)	Procurement and contracting	Contracts classified by risk of non-compliance; k-nearest neighbours best, above logistic regression and discriminant analysis
	Baret et al. (2023)	Budgeting, forecasting and fiscal risk	28 member states (2006 to 2018); eight of 138 predictors; nearly 92% accuracy, above logit
	Fazekas et al. (2024)	Procurement and contracting	More than 200,000 purchases; random forest R2 0.85, better than OLS
	Bø et al. (2025)	Tax and revenue administration	Smaller geographic units and ML clustering substantially improved valuation accuracy for wealth taxation
	Drăgulin et al. (2025)	Budgeting, forecasting and fiscal risk	AI models reported as more accurate than regression and econometric forecasts; no figures in abstract
	Liu et al. (2025)	Budgeting, forecasting and fiscal risk	49 states (2015 to 2017); balanced bagging generally best, F1 53.11% to 55.33% against 27% to 58% for logistic regression
	Khreis et al. (2026a)	Tax and revenue administration	10,625 obligations; XGBoost ROC-AUC 0.784, tree models above logistic regression; ranking by probability alone captured 0.4% of collection exposure against 92.7% when weighted by amount
	Medina-Hernández et al. (2026)	Procurement and contracting	Outperformed approaches based only on red flags; at least 20% precision in the top 5% of predictions
	Radebe et al. (2026)	Budgeting, forecasting and fiscal risk	Out-of-time PR-AUC 0.934 to 0.954; top-30 precision 0.967 to 1.000 with recall 0.186 to 0.192
Depends on task, data or horizon (4)	Chung et al. (2022)	Budgeting, forecasting and fiscal risk	Traditional statistical methods outperformed ML overall; k-nearest neighbours did better for property tax
	Larson & Overton (2024)	Budgeting, forecasting and fiscal risk	ML did not always improve accuracy; data pre-processing made the largest difference
	Agus et al. (2025)	Budgeting, forecasting and fiscal risk	542 governments (2013 to 2022); logistic regression among the best classifiers; random forest never selected
	Toffaha et al. (2026)	Budgeting, forecasting and fiscal risk	49,228 municipality-years; boosting best for unseen municipalities, fixed-effects predictor best for later years
No advantage, or method failed (2)	Guariso et al. (2023)	Budgeting, forecasting and fiscal risk	Regression and machine learning struggled to separate sector effects; agent computing was more informative
	Afonso & Sanchez Gomez (2026)	Budgeting, forecasting and fiscal risk	100 counties, more than 900 variables, almost 500 models; none predicted which counties ran recession deficits

Figure 5: Reporting completeness of abstracts by study design |



RQ3: Financial management information systems

The seven information-system studies are of a different kind (Table 8). None reports the performance of an algorithm; all examine implementation and effect. Kim and Park (2026), for which only the publisher's impact statement was available, conclude that a financial management information system alone does not ensure a credible expenditure budget and may even increase budget deviations unless it is integrated with treasury single account operations. In a cross-country study, de Aquino and Lino (2025) cluster 105 central governments by the interoperability of their systems and their provision of e-services and find four distinct configurations. The single-country studies, all from Africa and Latin America, describe why systems underperform: in Ghana, use was constrained by power struggles, opportunism and rent-seeking (Simpson et al., 2020), and by technological and socio-cultural barriers at local level (Tetteh et al., 2022); in South African municipalities, top-down and vendor-led implementation without end-user involvement produced resistance (Gcora & Chigona, 2019); in a Kenyan county, training and dependence on outside experts were the key technological factors (Tefera & Otieno, 2024); and in Brazilian local governments, replacing the system, as procurement rules periodically require, risked the loss of accrual accounting policies already configured (de Azevedo et al., 2020).

Table 8: Financial management information systems: included studies

Study	Setting	Design and methods	Main result stated in the abstract
Gcora & Chigona (2019)	South Africa	Qualitative or mixed methods. Case study; interviews	Five municipalities, 16 end-users; top-down, vendor-led implementation fostered resistance
Simpson et al. (2020)	Ghana	Qualitative or mixed methods. Case study; documents and interviews	IFMIS use constrained by power struggles, opportunism and rent-seeking
de Azevedo et al. (2020)	Brazil	Qualitative or mixed methods. Sequential mixed methods	Replacing the FMIS may reduce retention of accrual accounting policies; compulsory rebidding hampers reform
Tetteh et al. (2022)	Ghana	Qualitative or mixed methods. Case study; interviews	86 users; regulatory and organisational factors enabled use; technological and socio-cultural factors were barriers
Tefera & Otieno (2024)	Kenya	Survey. Principal component analysis	255 county employees; training and dependence on experts were the key technological factors

Study	Setting	Design and methods	Main result stated in the abstract
de Aquino & Lino (2025)	105 central governments	Statistical or econometric. Cluster analysis	Four configurations of FMIS interoperability and e-service delivery (GovTech Maturity Index 2022)
Kim & Park (2026)	Not stated	Statistical or econometric. Moderation analysis (from the title)	FMIS alone does not ensure budget credibility and may increase deviations unless integrated with a treasury single account

RQ4: Adoption, governance and ethics

Eight studies examine adoption of analytics and AI in audit and accounting (Table 9). Surveys of Indonesian government auditors find that intention to use big data analytics is associated with expected performance, expected effort and social influence (Saud et al., 2025), that system quality is associated with perceived ease of use and usefulness (Sofyani et al., 2025), and that use is associated with auditor performance as measured in the surveys. A survey (206 responses) of data-mining adoption in supreme audit institutions, where adoption is described as low, identifies enablers and barriers (Bracci et al., 2025); interviews in six European audit institutions describe transformation as institutional and organisational more than technological, with uneven and incremental adoption of AI (Viana et al., 2026); and a Delphi study with 19 experts places management support, infrastructure, budget and continuous training at the start of adoption (Ceki & Moloi, 2025). Among finance officers, those with direct experience of an AI-based accounting platform in Indonesian local governments saw both more benefit and more threat to their work (Maria & Halim, 2025); in Palestine and Jordan, ethical issues mediated the association between AI technologies and accounting task performance (Alslaibi et al., 2026); and a document-based comparison of Estonia, Singapore and Finland lists efficiency and transparency gains alongside bias and privacy risks (Aldemir & Uçma Uysal, 2025).

Table 9: Audit, accounting and governance: included studies

Study	Setting	Design and methods	Main result stated in the abstract
Aldemir & Uçma Uysal (2025)	Estonia, Singapore, Finland	Qualitative or mixed methods. Document-based case study	AI tools reported to improve efficiency and transparency; bias, privacy and ethics remain challenges
Bracci et al. (2025)	Not stated	Survey. SEM (TAM and status quo bias)	206 responses; data-mining adoption in supreme audit institutions is low; enablers and barriers identified
Ceki & Moloi (2025)	Eight countries	Qualitative or mixed methods. Two-round Delphi	19 experts; adoption starts with management support; infrastructure, readiness, budget and training
Maria & Halim (2025)	Indonesia	Survey. CFA; group comparisons	364 officers in 44 local governments; users saw more benefit and more operational and career threat
Saud et al. (2025)	Indonesia	Survey. PLS-SEM (UTAUT)	126 government auditors; intention associated with performance and effort expectancy and social influence, not with facilitating conditions
Sofyani et al. (2025)	Indonesia	Survey. PLS (modified TAM)	System quality associated with perceived ease of use and usefulness; analytics use associated with audit performance and quality
Alslaibi et al. (2026)	Palestine and Jordan	Survey. SEM with bootstrapping	500 valid responses; AI technologies associated with task performance; ethical issues mediated the effects

Study	Setting	Design and methods	Main result stated in the abstract
Viana et al. (2026)	European Union	Qualitative or mixed methods. Comparative interviews	Six audit institutions; transformation mainly institutional rather than technological; AI adoption uneven and incremental

Table 10 brings these results together with the information-system studies and with design features of the predictive studies. Six themes recur: integration before automation; implementation as a social as well as a technical process; skills, training and management support; perceived benefits and threats; explainability and oversight design; and the gap between model and use. Several tools are designed around human review, through visual analytics (Didimo et al., 2020), risk indices (Salazar et al., 2024), Shapley-value explanations (Khreis et al., 2026a; Medina-Hernández et al., 2026; Toffaha et al., 2026) or governance workflows (Hashim, 2026). No included abstract reports an empirical assessment of how errors of a screening model are distributed across the taxpayers, suppliers or jurisdictions that it scores.

Table 10: Enabling and constraining conditions: themes and supporting studies

Theme	What the studies report	Supporting studies
Integration before automation	A financial information system does not by itself improve budget credibility; integration with the treasury single account and with service-facing systems matters	Kim & Park (2026); de Aquino & Lino (2025)
Implementation is social as well as technical	Top-down, vendor-led roll-outs meet resistance; power struggles, rent-seeking and weak user involvement limit use; replacing a system can erase accounting policies	Simpson et al. (2020); Gcora & Chigona (2019); Tetteh et al. (2022); de Azevedo et al. (2020); Tefera & Otieno (2024)
Skills, training and management support	Adoption of analytics by auditors and finance officers depends on perceived usefulness, system quality, training and leadership; adoption in audit institutions is low and uneven	Saud et al. (2025); Sofyani et al. (2025); Bracci et al. (2025); Viana et al. (2026); Ceki & Moloi (2025); Andersson et al. (2025)
Perceived benefits and threats	Users report benefits for prediction, reporting and task performance, together with operational, ethical and career concerns	Maria & Halim (2025); Alslaibi et al. (2026); Aldemir & Uçma Uysal (2025)
Explainability and oversight design	Tools are built around analyst review: explanations, visual analytics, risk indices and human-in-the-loop forecasting	Didimo et al. (2020); Salazar et al. (2024); Hashim (2026); Khreis et al. (2026a); Medina-Hernández et al. (2026); Toffaha et al. (2026); Chung et al. (2025)
From model to use	Few abstracts report operational deployment; one mapping study notes few real deployments and few public datasets	Schneider dos Santos et al. (2025); Muir & Reich (2021); Khreis et al. (2026b)

DISCUSSION AND CONCLUSIONS

Summary of evidence

This review assembled 66 journal articles on data-driven financial management and financial system optimisation in the public sector. Four findings stand out. First, the sample is recent, with almost half of the included studies published in 2025 or 2026, and unevenly spread: procurement and tax, where data are abundant and the task is a well-defined screening problem, account for 34 of the 66 studies, while budget execution, cash management, payroll, grants and asset accounting hardly appear. Our search strings did not target these functions by name, so their absence may partly reflect the search. Second, predictive studies that state a figure often report good accuracy, but the abstracts rarely state a comparison with a simple baseline, and where they do the result is far from uniformly favourable: of 15 such comparisons, 9 favoured machine learning, 4 were mixed and 2 were negative. Third, the

studies of financial information systems, though few, point in the same direction: integration, user involvement and institutional conditions, more than software, shape whether a system improves financial management. Fourth, evidence of use and of financial outcomes is thin: few abstracts mention deployment, and almost none reports money saved, recovered or better allocated.

Where data-driven methods help

The pattern of results suggests when analytics is most likely to add value. It is strongest where many similar cases must be ranked with limited review capacity and where records are numerous and structured, as in tenders, contracts and tax returns (Fazekas et al., 2026; Khreis et al., 2026b; Wachs et al., 2020). In these settings the purpose of the model is to direct attention, and its value depends on what reviewers find in the cases it puts first. It is weakest where the quantity to be predicted is a short aggregate series subject to shocks, as in local revenue and deficit forecasting, where traditional methods were not beaten overall (Chung et al., 2022; Larson & Overton, 2024) and one large modelling exercise could not predict recession deficits at all (Afonso & Sanchez Gomez, 2026). The studies of Lebanese tax arrears and of Italian red flags add two cautions that apply to any screening use: a ranking by probability alone can miss almost all of the money at stake (Khreis et al., 2026a), and indicators that are obvious and closely scrutinised can be ineffective as predictors, which the authors of that study attribute to their predictability (Decarolis & Giorgiantonio, 2022).

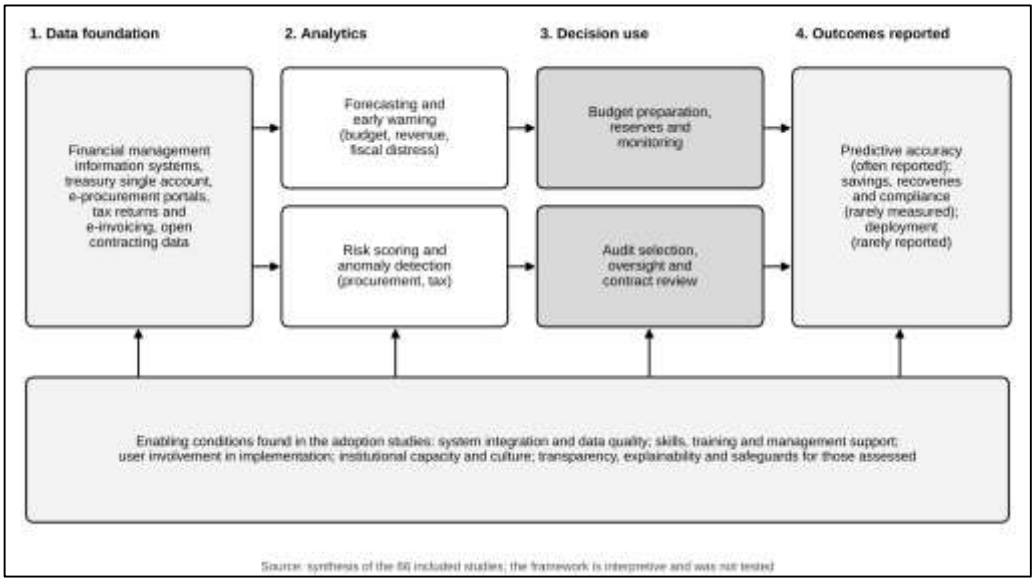
System optimisation as the precondition

The information-system studies and the analytics studies are largely separate literatures, but they bear on each other. Predictive models need data that are complete, consistent over time and linked across functions, and their output must reach a budget officer, auditor or contract manager inside a workflow. The system studies show that these conditions cannot be assumed: systems are replaced and configurations lost (de Azevedo et al., 2020), implementation is contested (Gcora & Chigona, 2019; Simpson et al., 2020; Tetteh et al., 2022), and automation without treasury integration may not improve, and may worsen, budget credibility (Kim & Park, 2026). Each of these points rests on one or a few abstracts, and none of the system studies examined analytics directly. With that caution, the practical implication we draw is one of sequence: investment in interoperability, data quality and user capability plausibly comes before, and conditions the return on, investment in advanced analytics.

An interpretive framework

Figure 6 arranges the findings as a chain from data foundation to analytics, decision use and outcomes, resting on the enabling conditions identified by the adoption studies. The framework is a summary of where the reviewed evidence lies, not a tested model. Most studies sit in the second column; the first is covered by the system studies; the third and fourth columns, which concern how results are used and what they change, are where evidence is scarcest.

Figure 6: Interpretive framework of data-driven public financial management, with the stages at which the included studies provide evidence |



Implications for practice

For finance ministries, audit institutions, tax authorities and procurement agencies, the review supports five practical points. (1) Start from a decision with limited capacity, such as which returns to audit or which contracts to review, and judge a model by the yield of the cases it selects, weighted by the amounts involved. (2) Keep a simple baseline in production alongside any machine-learning model, and test both on later periods; in forecasting, attend to data preparation before model choice (Larson & Overton, 2024). (3) Treat the financial information system as the analytical platform: integrate it with treasury and service systems, and protect configurations and data continuity when systems are replaced. (4) Involve the users: the adoption studies link use to perceived usefulness and system quality (Saud et al., 2025; Sofyani et al., 2025) and to training and management support (Ceki & Moloi, 2025; Tefera & Otieno, 2024). (5) Build in explanation, human review and a route of appeal for those who are scored, because interpretable approaches are preferable where they perform as well (Rudin, 2019) and algorithms built on administrative data can reproduce disparities (Obermeyer et al., 2019; Rajkomar et al., 2018).

Research agenda

Table 11 sets out the gaps observed and the priorities that follow. The most important is evaluation of outcomes: until studies report what happened when a model was used, in recoveries, savings, forecast error in budget terms or time, the case for data-driven financial management rests on predictive accuracy alone. Closely related are routine comparison with baselines, publication of negative results such as those of Afonso and Sanchez Gomez (2026), and minimum reporting standards, since 15 of 66 abstracts do not name the jurisdiction studied and 37 do not state the sample size. Calibration and measures suited to rare events deserve more attention than accuracy (He & Garcia, 2009; Saito & Rehmsmeier, 2015; Van Calster et al., 2019).

Table 11: Gaps in the reviewed literature and priorities for research and practice

Gap observed in the included studies	Research and practice priority
Results reported as predictive accuracy, rarely as money saved, recovered or better allocated	Evaluate outcomes that matter to finance ministries and auditors: yield per case reviewed, forecast error in budget terms, recoveries, time saved
Comparisons with simple baselines are uncommon and, where made, often mixed or unfavourable	Always report a conventional baseline and an out-of-time test; publish negative results
15 of 66 abstracts do not state the jurisdiction; 37 do not state the sample size	Adopt minimum reporting items for public-finance analytics (setting, data source and period, sample, validation design, baseline)
Procurement and tax dominate; budget execution, cash management, payroll, grants and asset accounting are rarely studied	Extend analytics to under-studied functions using data already held in financial management information systems
System studies and analytics studies are separate literatures	Study analytics as part of the information system: data quality, interoperability, treasury integration and user workflow
Fairness, due process and the position of those scored are seldom examined empirically	Audit false-positive burdens across taxpayer, supplier and jurisdiction groups; test explanation and appeal mechanisms
Evidence on generative AI is limited to single exploratory studies	Benchmark large language models against statistical forecasts and human experts with pre-specified protocols

Limitations

This review has substantial limitations, and its conclusions should be read with them in mind. (1) It is based on bibliographic records and abstracts. Full texts were not read, so results, designs and limitations that authors did not put in the abstract are not reflected, and no risk-of-bias assessment was

possible. The abstracts were condensed by an automated step before coding and were not checked against the publishers' versions. (2) Screening and coding were done once; the reviewer was a language-model tool, no human screened or coded the records, there was no second reviewer, and the judgements of which sources were established journals and of the verdict on machine learning were subjective. Individual title-screening decisions were not recorded, and the retrieval log is incomplete for some included studies. (3) The search used a single open index, OpenAlex, because subscription databases were not available. For eight of the thirteen searches only the first 100 results by relevance were screened, so 6,305 returned records were never examined, and relevant studies have certainly been missed. Because duplicates were not counted, the number of unique records returned and screened is unknown. Year and language limits were applied as search filters only in S1 and otherwise at screening. The search strings emphasised machine learning and may under-represent work described in other terms, such as business intelligence, optimisation or performance budgeting. (4) Ten otherwise relevant reports were excluded only because no abstract was available in open metadata, which removes some well-cited studies and biases the sample towards publishers that deposit abstracts. (5) Coding from abstracts makes the counts conservative in particular ways: a setting, sample size or comparison was recorded only if the abstract stated it. (6) The search was run on one day, the protocol was not registered, search strings and one eligibility criterion were added during the review, and the review covers English-language journal articles only. For these reasons the review should be regarded as a structured, transparent mapping of the literature that can be extended, not as an exhaustive or quality-appraised synthesis.

Conclusion

Data-driven methods have been applied across the public financial management cycle, most often to rank procurement and tax cases for scrutiny and to anticipate fiscal distress, and many studies report accurate predictions. The reviewed evidence does not yet show that these methods routinely outperform simpler ones, particularly in revenue and budget forecasting, nor what they deliver in financial terms once in use. The few studies of financial management information systems suggest that integration, continuity and user involvement condition what such systems, and any analytics built on them, can deliver. Progress now depends less on new algorithms than on evaluation of outcomes, honest baselines, better reporting and attention to the systems and institutions in which the models must work.

DATA AVAILABILITY

The search log, the list of the 84 reports assessed with the reasons for exclusion, the coded dataset of the 66 included studies, the code that produced the tables and figures, and the reference list in EndNote XML and RIS formats are supplied as supplementary material. Bibliographic records were retrieved from the public APIs of OpenAlex (api.openalex.org) and Crossref (api.crossref.org) on 17 July 2026.

REFERENCES

[1]. Abedin, M. Z., Chi, G., Uddin, M. M., Satu, M. S., Khan, M. I., & Hajek, P. (2021). Tax default prediction using feature transformation-based machine learning. *IEEE Access*, *9*, 19864-19881. https://doi.org/10.1109/access.2020.3048018

[2]. Abedin, M. Z., Hassan, M. K., Khan, I., & Julio, I. F. (2021). Feature transformation for corporate tax default prediction: Application of machine learning approaches. *Asia-Pacific Journal of Operational Research*, *39*(04), 2140017. https://doi.org/10.1142/s0217595921400170

[3]. Acikalin, U. U., Gorgun, M. K., Kutlu, M., & Tas, B. K. O. (2023). How you describe procurement calls matters: Predicting outcome of public procurement using call descriptions. *Natural Language Engineering*, *30*(6), 1255-1276. https://doi.org/10.1017/s135132492300030x

[4]. Afonso, W., & Sanchez Gomez, J. (2026). Machine learning to predict local government deficits during economic downturns. *Journal of Public Budgeting, Accounting & Financial Management*, 1-22. https://doi.org/10.1108/jpbafm-09-2025-0261

[5]. Agus, P., Ridha, M. H., Pramusetiyo, F., & Hidayat, R. (2025). Machine learning implementation to predict financial distress in local governments of Indonesia. *Pacific Accounting Review*, 1-22. https://doi.org/10.1108/par-11-2024-0317

[6]. Aldemir, C., & Uçma Uysal, T. (2025). Artificial intelligence for financial accountability and governance in the public sector: Strategic opportunities and challenges. *Administrative Sciences*, *15*(2), 58. https://doi.org/10.3390/admsci15020058

[7]. Alsadhan, N. (2023). A multi-module machine learning approach to detect tax fraud. *Computer Systems Science and Engineering*, *46*(1), 241-253. https://doi.org/10.32604/csse.2023.033375

[8]. Alslaibi, N., Qawasmeh, R., Samara, H., Hussein, A., & Hussein, W. (2026). Artificial intelligence and accounting task performance in public sector institutions through the mediating role of ethical issues in Palestine and Jordan. *Discover Global Society*, *4*(1), 241. https://doi.org/10.1007/s44282-026-00592-3

[9]. Aminul Islam, M. (2025). Calibrated Machine Learning for Reducing Decision Risk and Improving Forecast Reliability in U.S. Enterprise Analytics. *American Journal of Interdisciplinary Studies*, *15*(12), 01-36. https://doi.org/10.63125/869qdj55

[10]. Aminul Islam, M., & Md, A. (2023). Adaptive Class-Imbalance Learning for High-Sensitivity Fraud Detection in Large-Scale Digital Transaction Networks. *International Journal of Business and Economics Insights*, *13*(09), 01-33. https://doi.org/10.63125/bmxwv741

[11]. Andersson, P. E., Arbin, K., & Rosenqvist, C. (2025). Assessing the value of artificial intelligence (AI) in governmental public procurement. *Journal of Public Procurement*, *25*(1), 120-139. https://doi.org/10.1108/jopp-05-2024-0057

[12]. Asheke, M. (2025). Artificial Intelligence Driven Supply Chain Risk Forecasting and Optimization for Resilient Industrial Production Networks. *ASRC Procedia: Global Perspectives in Science and Scholarship*, *1*(01), 2923–2960. https://doi.org/10.63125/rapn5860

[13]. Asheke, M., & Tashriq Ahmed Bin, A. (2022). Federated Learning and Differential Privacy for Cross-Firm Construction-Materials Supply-Chain Disruption Forecasting: A Quantitative Analysis of Privacy-Utility Trade-Offs. *American Journal of Advanced Technology and Engineering Solutions*, *12*(09), 01-40. https://doi.org/10.63125/g4at4237

[14]. Ashik, M. (2025). PRISMA-Based Systematic Literature Review of Reliability-Centered Turnaround Planning for Reducing Downtime in Power Generation Facilities. *American Journal of Scholarly Research and Innovation*, *4*(01), 1036–1080. https://doi.org/10.63125/2878n718

[15]. Atanasijević, J., Jakovetić, D., Krejić, N., Krklec-Jerinkić, N., & Marković, D. (2019). Using big data analytics to improve efficiency of tax collection in the tax administration of the Republic of Serbia. *Ekonomika preduzeca*, *67*(1-2), 115-130. https://doi.org/10.5937/ekopre1808115a

[16]. Baret, K., Barbier-Gauchard, A., & Papadimitriou, T. (2023). Forecasting stability and growth pact compliance using machine learning. *The World Economy*, *47*(1), 188-216. https://doi.org/10.1111/twec.13518

[17]. Batóg, B., & Batóg, J. (2021). Regional government revenue forecasting: Risk factors of investment financing. *Risks*, *9*(12), 210. https://doi.org/10.3390/risks9120210

[18]. Belahouaoui, R., & Alm, J. (2025). Tax fraud detection using artificial intelligence-based technologies: Trends and implications. *Journal of Risk and Financial Management*, *18*(9), 502. https://doi.org/10.3390/jrfm18090502

[19]. Bø, E. E., Nygård, O. E., & Thoresen, T. O. (2025). Improving housing valuation for taxation: The Norwegian case. *International Tax and Public Finance*, *33*(5), 1807-1827. https://doi.org/10.1007/s10797-025-09928-2

[20]. Bolton, R. J., & Hand, D. J. (2002). Statistical fraud detection: A review. *Statistical Science*, *17*(3), 235-255. https://doi.org/10.1214/ss/1042727940

[21]. Bracci, E., Tallaki, M., & Otia, J. E. (2025). Propensity factors of artificial intelligence technology adoption by public sector auditors. *Journal of Public Budgeting, Accounting & Financial Management*, *38*(1), 59-86. https://doi.org/10.1108/jpbafm-09-2024-0189

[22]. Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

[23]. Ceki, B., & Moloi, T. (2025). Technology adoption framework for supreme audit institutions within the hybrid TAM and TOE model. *Journal of Risk and Financial Management*, *18*(8), 409. https://doi.org/10.3390/jrfm18080409

[24]. Chung, I. H., Kara, B., McShea, M., Pathak, R., & Williams, D. (2025). Using large language models to forecast local government revenue. *Public Finance Journal*, *2*(2), 85-98. https://doi.org/10.59469/pfj.2025.46

[25]. Chung, I. H., Williams, D. W., & Do, M. R. (2022). For better or worse? Revenue forecasting with machine learning approaches. *Public Performance & Management Review*, *45*(5), 1133-1154. https://doi.org/10.1080/15309576.2022.2073551

[26]. de Aquino, A. C. B., & Lino, A. F. (2025). Aligning back-end systems with front-end services: Configurations of financial management information systems and e-services provision. *Journal of Public Budgeting, Accounting & Financial Management*, *38*(1), 87-110. https://doi.org/10.1108/jpbafm-12-2024-0265

[27]. de Azevedo, R. R., Lino, A. F., de Aquino, A. C. B., & Machado-Martins, T. C. P. (2020). Financial management information systems and accounting policies retention in Brazil. *International Journal of Public Sector Management*, *33*(2/3), 207-227. https://doi.org/10.1108/ijpsm-01-2019-0027

[28]. Decarolis, F., & Giorgiantonio, C. (2022). Corruption red flags in public procurement: New evidence from Italian calls for tenders. *EPJ Data Science*, *11*(1), 16. https://doi.org/10.1140/epjds/s13688-022-00325-x

[29]. Demeter, G., Delina, R., Urminsky, J., & Zajarosova, M. (2025). Assessing the impact of supplier group stability on public procurement performance in Slovakia: A causal forest analysis of collusion signals. *IEEE Access*, *13*, 41607-41624. https://doi.org/10.1109/access.2025.3547688

[30]. Didimo, W., Grilli, L., Liotta, G., Menconi, L., Montecchiani, F., & Pagliuca, D. (2020). Combining network visualization and data mining for tax risk assessment. *IEEE Access*, *8*, 16073-16086. https://doi.org/10.1109/access.2020.2967974

[31]. Drăgulin, B., Ştefan, V., Tăbîrcă, A.-I., Şcheau, M.-C., Radu, V., & Munteanu, V. (2025). AI-driven models for forecasting public expenditures in the digital era. *Electronics*, *14*(20), 4047. https://doi.org/10.3390/electronics14204047

[32]. Fazekas, M., Tóth, B., Wachs, J., & Abdou, A. (2026). Public procurement cartels: A large-sample testing of screens using machine learning. *International Journal of Industrial Organization*, *104*, 103228. https://doi.org/10.1016/j.ijindorg.2025.103228

[33]. Fazekas, M., Veljanov, Z., & de Oliveira, A. B. (2024). Predicting pharmaceutical prices. Advances based on purchase-level data and machine learning. *BMC Public Health*, *24*, 1888. https://doi.org/10.1186/s12889-024-19171-9

[34]. García Rodríguez, M. J., Rodríguez-Montequín, V., Ballesteros-Pérez, P., Love, P. E. D., & Signor, R. (2022). Collusion detection in public procurement auctions with machine learning algorithms. *Automation in Construction*, *133*, 104047. https://doi.org/10.1016/j.autcon.2021.104047

[35]. García Rodríguez, M. J., Rodríguez Montequín, V., Ortega Fernández, F., & Villanueva Balsera, J. M. (2019). Public procurement announcements in Spain: Regulations, data analysis, and award price estimator using machine learning. *Complexity*, *2019*(1), 2360610. https://doi.org/10.1155/2019/2360610

[36]. García Rodríguez, M. J., Rodríguez Montequín, V., Ortega Fernández, F., & Villanueva Balsera, J. M. (2020). Bidders recommender for public procurement auctions using machine learning: Data analysis, algorithm, and case study with tenders from Spain. *Complexity*, *2020*, 1-20. https://doi.org/10.1155/2020/8858258

[37]. Gazi Yeasin, A. (2024a). Designing Trustworthy AI Output Interfaces: A UX Framework for Graceful Handling of Generative Model Uncertainty in Consumer Applications. *American Journal of Scholarly Research and Innovation*, *14*(07), 01–40. https://doi.org/10.63125/4m5fjb12

[38]. Gazi Yeasin, A. (2024b). A Quantitative Evaluation of Behavioral Design Patterns on User Engagement and Activation in Large-Scale Campus Software Products. *American Journal of Interdisciplinary Studies*, *14*(06), 01-40. https://doi.org/10.63125/x3d5k517

[39]. Gcora, N., & Chigona, W. (2019). Post-implementation evaluation and challenges of integrated financial management information systems for municipalities in South Africa. *SA Journal of Information Management*, *21*(1). https://doi.org/10.4102/sajim.v21i1.1066

[40]. Gorina, E., Maher, C., & Joffe, M. (2017). Local fiscal distress: Measurement and prediction. *Public Budgeting & Finance*, *38*(1), 72-94. https://doi.org/10.1111/pbaf.12165

[41]. Guariso, D., Castañeda, G., & Guerrero, O. A. (2023). Budgeting for SDGs: Quantitative methods to assess the potential impacts of public expenditure. *Development Engineering*, *8*, 100113. https://doi.org/10.1016/j.deveng.2023.100113

[42]. Hashim, H. A. (2026). Governance-centered AI framework for public-sector budgetary decision-making and financial risk management. *Electronics*, *15*(17), 3786. https://doi.org/10.3390/electronics15173786

[43]. He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, *21*(9), 1263-1284. https://doi.org/10.1109/tkde.2008.239

[44]. Henrique, B. M., Sobreiro, V. A., & Kimura, H. (2020). Contracting in Brazilian public administration: A machine learning approach. *Expert Systems*, *37*(5), e12550. https://doi.org/10.1111/exsy.12550

[45]. Kazi Mohammad Khalid, A., & Sumaiya Sadmeen, M. (2022). Geospatial Artificial Intelligence for Urban Flood Susceptibility and Drainage Resilience: A Prisma 2020 Systematic Review and Bibliometric Meta-Synthesis. *Review of Applied Science and Technology*, *12*(02), 01–51. https://doi.org/10.63125/epqdvy23

[46]. Khreis, M., Harb, H., & Dia, S. (2026a). A machine learning framework for predicting tax payment arrears: Comparative model evaluation and Shapley-value interpretability. *Journal of Risk and Financial Management*, *19*(9), 723. https://doi.org/10.3390/jrfm19090723

[47]. Khreis, M., Harb, H., & Dia, S. (2026b). VERITAS: A verified-data machine learning approach to segment-specific tax audit planning. *Journal of Risk and Financial Management*, *19*(9), 718. https://doi.org/10.3390/jrfm19090718

[48]. Kim, M., & Park, W. (2026). Financial management information systems, treasury single accounts, and expenditure budget credibility: A moderation-effect approach. *Public Money & Management*, 1-13. https://doi.org/10.1080/09540962.2026.2641194

[49]. Kynatun, M. (2025). Integrating AI-Driven Security Analytics and IT Governance Frameworks for Cybersecurity Decision Support in U.S. Financial Information Systems. *American Journal of Advanced Technology and Engineering Solutions*, *15*(12), 01-42. https://doi.org/10.63125/kzcdyz73

[50]. Larson, S., & Overton, M. (2024). Modeling approach matters, but not as much as preprocessing. *Public Finance Journal*, *1*(1), 29-48. https://doi.org/10.59469/pfj.2024.8

[51]. Liu, R., Li, H., Yoon, K., & Vasarhelyi, M. A. (2025). Using machine learning algorithms to improve fiscal distress prediction models: The case of U.S. local governments. *Journal of Information Systems*, *39*(3), 131-155. https://doi.org/10.2308/isys-2023-049

[52]. Lundberg, J. (2025). Cartel detection in public procurement: Evaluation of five econometric methods. *Journal of Public Procurement*, *26*(1), 1-29. https://doi.org/10.1108/jopp-03-2024-0028

[53]. Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, *2*(1), 56-67. https://doi.org/10.1038/s42256-019-0138-9

[54]. Maishar Mostari, Z. (2024). Uncertainty-Calibrated Explainable AI for Cybersecurity Alert Prioritization: Improving Detection Reliability and Human Decision-Making. *American Journal of Scholarly Research and Innovation*, *14*(12), 01–40. https://doi.org/10.63125/an0pcq39

[55]. Maishar Mostari, Z. (2025). Privacy-Preserving AI-Based Cyber Threat Detection in Financial Organizations: A Governance and Business Value Framework. *American Journal of Interdisciplinary Studies*, *15*(04), 01-51. https://doi.org/10.63125/p8fct530

[56]. Maria, E., & Halim, A. (2025). Institutional readiness and resistance to AI-based accounting platforms: Evidence from Indonesian local governments. *Transforming Government: People, Process and Policy*, *20*(2), 248-266. https://doi.org/10.1108/tg-06-2025-0176

[57]. Mashfiqur, R. (2024). Anomaly Detection and Process Mining for Identifying Financial Irregularities and Strengthening Procurement Controls. *American Journal of Advanced Technology and Engineering Solutions*, *14*(09), 91-140. https://doi.org/10.63125/hmg5ta84

[58]. Mashfiqur, R., & Md Jamil, A. (2023). Machine Learning-Based Cost Variance Prediction for Early Detection of Budget Overruns in Procurement-Intensive Operations. *Review of Applied Science and Technology*, *13*(12), 142–199. https://doi.org/10.63125/7g92wd03

[59]. Masrul, H. (2025). Explainable Machine Learning for Real-Time Financial Fraud Detection and Risk-Based Transaction Monitoring in U.S. Banking and Digital Payment Systems. *American Journal of Data Science and Analytics*, *6*(10), 70-108. https://doi.org/10.63125/4h9v1876

[60]. Masrul, H., & Mashfiqur, R. (2024). Unified Analytical Modeling for Cryptocurrency Price Volatility Forecasting and Market-Manipulation Detection in U.S. Digital Asset Markets. *American Journal of Advanced Technology and Engineering Solutions*, *14*(11), 01-40. https://doi.org/10.63125/nqgh7934

[61]. Masrul, H., & Mukut Kanti, B. (2024). AI-Driven Project Portfolio and Business Analytics for Forecasting Cost Overruns, Schedule Delays, Resource Bottlenecks, and Enterprise Performance Across U.S. Industries. *Review of Applied Science and Technology*, *14*(12), 14–54. https://doi.org/10.63125/x7te4644

[62]. Md, F., & Kynatun, M. (2023). A Quantitative Assessment of Cybersecurity Risk Scoring Models for Enterprise Database Systems in Cloud-Native Environments. *Review of Applied Science and Technology*, *13*(12), 50–103. https://doi.org/10.63125/gk2krm47

[63]. Md Hasinur, R. (2024). Human in the Loop Artificial Intelligence for Loan Document Verification and Credit Risk Reporting: A Bibliometric and Science Mapping Analysis. *American Journal of Interdisciplinary Studies*, *14*(06), 41-71. https://doi.org/10.63125/8x359j95

[64]. Md Hasinur, R. (2025a). Machine Learning for Nonperforming Corporate Loan Recovery: A Systematic Review and Meta-Analysis of Predictive Performance. *American Journal of Scholarly Research and Innovation*, *4*(01), 1174–1202. https://doi.org/10.63125/a7v6f533

[65]. Md Hasinur, R. (2025b). Present and Future of Digital Banking Services in Bangladesh. *Review of Applied Science and Technology*, *4*(02), 725-749. https://doi.org/10.63125/k7tgr183

[66]. Md Maksudul, H. (2024). Supplier-Risk Scoring and Procurement Resilience for Distributed Athletic-Apparel and Equipment Supply Networks. *American Journal of Advanced Technology and Engineering Solutions*, *14*(09), 44-90. https://doi.org/10.63125/a9fmgj50

[67]. Md Maksudul, H. (2025). Data-Driven Fulfillment Optimization for Time-Sensitive Teamwear and Sporting Goods Distribution. *American Journal of Scholarly Research and Innovation*, *4*(01), 1120–1173. https://doi.org/10.63125/rpezfz61

[68]. Md Maksudul, H., & Md Takbir Hossen, S. (2023). Predictive Inventory Optimization for Reducing Stockouts and Excess Inventory in U.S. Sports and Apparel Operations. *Review of Applied Science and Technology*, *13*(12), 104–141. https://doi.org/10.63125/8djn1j48

[69]. Md Maksudul, H., & Md Tazul Islam, C. (2022). Machine Learning-Based Cash-Flow Forecasting for Improving Financial Planning in U.S. Small and Mid-Sized Businesses. *International Journal of Scientific Interdisciplinary Research*, *12*(12), 01–38. https://doi.org/10.63125/w3pgjx97

[70]. Md Mohsin, B. (2024). Artificial Intelligence for Medication Safety and Nursing Documentation in Emergency Departments: A Bibliometric Analysis. *American Journal of Interdisciplinary Studies*, *14*(10), 01-34. https://doi.org/10.63125/2krdb618

[71]. Md Mohsin, B., & Jannatul, F. (2021). Emergency Nursing Triage and Timely Escalation of Care: A Quantitative Clinical Audit. *International Journal of Scientific Interdisciplinary Research*, *11*(07), 01–33. https://ijsir.org/index.php/IJSIR/article/view/138

[72]. Md Mohsin, B., & Mst Kaniz, F. (2023). AI-Enabled Clinical Decision Support for Standardized Nurse-Led Emergency Care Pathways: A Multisite Mixed-Methods Implementation Study. *Review of Applied Science and Technology*, *13*(08), 01–33. https://doi.org/10.63125/ac09sa31

[73]. Md Momen Shahriar, J. (2025). Viscoelastic Creep Characterization of LDPE Under Multilevel Stress and Temperature Using Time-Temperature Superposition and Constitutive Modeling. *American Journal of Scholarly Research and Innovation*, *15*(11), 01–38. https://doi.org/10.63125/2edstd86

[74]. Md Tazul Islam, C. (2024). Explainable AI For Financial and Operational Decision Support: Improving Transparency and Managerial Trust. *American Journal of Advanced Technology and Engineering Solutions*, *14*(12), 01-38. https://doi.org/10.63125/kwfvzw68

[75]. Md Tazul Islam, C. (2025). Business Intelligence and Data-Quality Frameworks for Improving Evidence-Based Decisions across Multi-Source Enterprise Data. *American Journal of Scholarly Research and Innovation*, *15*(07), 01–41. https://doi.org/10.63125/qhh6ka59

[76]. Md. Mainuddin, F. (2024). Quantitative Structural Retrofit Assessment Models for Strengthening Existing Steel Buildings Under Increased Load Demands. *Review of Applied Science and Technology*, *3*(04), 325–366. https://doi.org/10.63125/yyqnte84

[77]. Md. Mamunur, R. (2024). Predictive Analytics for Cloud Service Reliability A Comparative Evaluation of System Performance and Disaster-Recovery Metrics. *Review of Applied Science and Technology*, *14*(12), 55–91. https://rast-journal.org/index.php/RAST/article/view/171

[78]. Md. Mamunur, R. (2025). AI-Driven Quantitative Cybersecurity Risk Assessment of Identity and Access Management in Hybrid Cloud Environments. *American Journal of Interdisciplinary Studies*, *15*(10), 01-37. https://doi.org/10.63125/1yc7ee20

[79]. Md. Mamunur, R., & Md, A. (2023). Cloud Migration Readiness and Operational Risk in U.S. Organizations: A Quantitative Analysis of Infrastructure and Business Continuity Indicators. *American Journal of Scholarly Research and Innovation*, *13*(07), 01-27. https://doi.org/10.63125/28jkj206

[80]. Medina-Hernández, M., Kertész, J., & Fazekas, M. (2026). Learning from sanctioned government suppliers: A machine learning and network science approach to detecting fraud and corruption in Mexico. *Scientific Reports*, *16*(1), 22382. https://doi.org/10.1038/s41598-026-48873-w

[81]. Muir, W. A., & Reich, D. (2021). Using machine learning to improve public reporting on U.S. government contracts. *INFORMS Journal on Applied Analytics*, *51*(6), 463-479. https://doi.org/10.1287/inte.2021.1098

[82]. Murorunkwere, B. F., Ihirwe, J. F., Kayijuka, I., Nzabanita, J., & Haughton, D. (2023). Comparison of tree-based machine learning algorithms to predict reporting behavior of electronic billing machines. *Information*, *14*(3), 140. https://doi.org/10.3390/info14030140

[83]. Nai, R., Meo, R., Morina, G., & Pasteris, P. (2023). Public tenders, complaints, machine learning and recommender systems: A case study in public administration. *Computer Law & Security Review*, *51*, 105887. https://doi.org/10.1016/j.clsr.2023.105887

[84]. Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, *366*(6464), 447-453. https://doi.org/10.1126/science.aax2342

[85]. Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., . . . Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, n71. https://doi.org/10.1136/bmj.n71

[86]. Potoski, M., Lund-Sørensen, B., & Petersen, O. H. (2025). Measuring transaction costs in public sector contracting through machine learning and contract text. *Public Administration Review*, *86*(1), 199-216. https://doi.org/10.1111/puar.13947

[87]. Punnaiah Chowdary, R. (2021). GC-MS Strategies for Identification of Trace Impurities in Active Pharmaceutical Ingredients. *American Journal of Advanced Technology and Engineering Solutions*, *11*(05), 01-35. https://ajates-scholarly.com/index.php/ajates/article/view/131

[88]. Punnaiah Chowdary, R., & Md, A. (2020). Practical HPLC Method Validation Under ICH Q2(R2) and Q14: An Analytical Quality Framework. *International Journal of Business and Economics Insights*, *10*(03), 01–36. https://ijbei-journal.org/index.php/ijbei/article/view/73

[89]. Punnaiah Chowdary, R., & Md Jamil, A. (2020). Risk-Based Lifecycle Management of Chromatographic Methods for Generic API Quality: A Systematic Review. *Review of Applied Science and Technology*, *10*(09), 01–41. https://doi.org/10.63125/r0xnar90

[90]. Radebe, N. E., Nomlala, B. C., & Matenda, F. R. (2026). Forecasting municipal financial distress in South Africa: A machine learning approach. *Forecasting*, *8*(1), 18. https://doi.org/10.3390/forecast8010018

[91]. Rajkomar, A., Hardt, M., Howell, M. D., Corrado, G., & Chin, M. H. (2018). Ensuring fairness in machine learning to advance health equity. *Annals of Internal Medicine*, *169*(12), 866-872. https://doi.org/10.7326/m18-1990

[92]. Rocha, J. L. M., Aurelio, M., Irazema, N., Teofilo, J., Quispe, D., & Oscar, J. (2020). Level of budget execution according to the professional profile of regional governors applying machine learning models. *International Journal of Advanced Computer Science and Applications*, *11*(11). https://doi.org/10.14569/ijacsa.2020.0111139

[93]. Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, *1*(5), 206-215. https://doi.org/10.1038/s42256-019-0048-x

[94]. Sadia, A. (2025). Privacy-Aware Personalization: Machine Learning, Model Validation, And Consumer Trust in AI-Mediated Marketing. *American Journal of Scholarly Research and Innovation*, *4*(01), 1081–1119. https://doi.org/10.63125/j7tfr242

[95]. Sadia, A., & Md. Mosheur, R. (2023). Consumer Trust, Perceived Authenticity, and Personalized Digital Marketing: A Quantitative Decision-Science Study. *International Journal of Scientific Interdisciplinary Research*, *4*(2), 252–295. https://doi.org/10.63125/cajkqt60

[96]. Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLOS ONE*, *10*(3), e0118432. https://doi.org/10.1371/journal.pone.0118432

[97]. Salazar, A., Pérez, J. F., & Gallego, J. (2024). VigIA: Prioritizing public procurement oversight with machine learning models and risk indices. *Data & Policy*, *6*, e75. https://doi.org/10.1017/dap.2024.83

[98]. Saud, I. M., Sofyani, H., Utami, T. P., Haq, M. M., & Fathmaningrum, E. S. (2025). Big data analytics-based auditing adoption in public sector: Indonesian evidence. *Cogent Business & Management*, *12*(1), 2454320. https://doi.org/10.1080/23311975.2025.2454320

[99]. Schneider dos Santos, E., Machado dos Santos, M., Castro, M., & Tyska Carvalho, J. (2025). Detection of fraud in public procurement using data-driven methods: A systematic mapping study. *EPJ Data Science*, *14*(1), 52. https://doi.org/10.1140/epjds/s13688-025-00569-3

[100]. Shabbir, A. (2025). Physics-Based Failure Prognostics for Preventing Pressure-System Degradation and Unplanned Industrial Shutdowns. *American Journal of Scholarly Research and Innovation*, *14*(09), 01–41. https://doi.org/10.63125/3asd0053

[101]. Sharmin, A. (2025). Transferable Patient-Centered Communication Models: A Multisite Implementation Study. *American Journal of Scholarly Research and Innovation*, *15*(09), 01–32. https://doi.org/10.63125/j5p8bm56

[102]. Sharmin, A., & Mst Kaniz, F. (2023). "My Pain Was Not Taken Seriously" Medical Gaslighting As an Expectancy Violation in Women's Healthcare. *American Journal of Health and Medical Sciences*, *13*(06), 55–85. https://ajhmsjournal.com/index.php/ajhms/article/view/20

[103]. Sharmin, A., & Nishat, J. (2023). Communication Accommodation and Patient Trust in Healthcare and Chronic Illness Care. *Review of Applied Science and Technology*, *13*(08), 34–72. https://doi.org/10.63125/6p0xwb79

[104]. Shin, B. (2025). Exploring the potential of machine learning to reduce administrative burden in participatory budgeting: A case study of Seoul. *Journal of Public Budgeting, Accounting & Financial Management*, *38*(1), 237-264. https://doi.org/10.1108/jpbafm-09-2024-0188

[105]. Siciliani, L., Taccardi, V., Basile, P., Di Ciano, M., & Lops, P. (2023). AI-based decision support system for public procurement. *Information Systems*, *119*, 102284. https://doi.org/10.1016/j.is.2023.102284

[106]. Simpson, S. N. Y., Tetteh, L. A., & Agyenim-Boateng, C. (2020). Exploring the socio-cultural factors in the implementation of public financial management information system in Ghana. *Journal of Accounting & Organizational Change*, *16*(3), 349-368. https://doi.org/10.1108/jaoc-10-2018-0100

[107]. Soffan, S., Bramantoro, A., & Alzahrani, A. A. (2025). Combination of machine learning and data envelopment analysis to measure the efficiency of the tax service office. *PeerJ Computer Science*, *11*, e2672. https://doi.org/10.7717/peerj-cs.2672

[108]. Sofyani, H., Amalia, R., Abu Hasan, H., & Saleh, Z. (2025). Big data analytics in enhancing public sector auditing: Drivers, benefits and the moderating role of auditor certification. *Journal of Accounting in Emerging Economies*, *16*(1), 31-50. https://doi.org/10.1108/jaee-12-2024-0538

[109]. Songeta, D. (2024). Human-In-The-Loop Artificial Intelligence for Account Reconciliation and Financial-Reporting Transparency: A Bibliometric and Science-Mapping Analysis. *American Journal of Scholarly Research and Innovation*, *14*(12), 41–72. https://doi.org/10.63125/jamadx91

[110]. Songeta, D. (2025). Machine Learning for Retail Sales and KPI Forecasting: A Systematic Review and Meta-Analysis of Predictive Performance. *American Journal of Advanced Technology and Engineering Solutions*, *15*(08), 01-36. https://doi.org/10.63125/xan6v087

[111]. Sumaiya, M. (2024). Hospital Wastewater as A Reservoir of Antimicrobial Resistance Genes: Implications for Public Health and Genomic Surveillance in The United States. *Review of Applied Science and Technology*, *3*(04), 448–508. https://doi.org/10.63125/ahvas976

[112]. Sumaiya Sadmeen, M. (2025). Spatiotemporal Graph Neural Networks and Digital Twins for Urban Drainage Networks: Predictive Inundation, Critical-Node Detection, And Infrastructure Prioritization. *American Journal of Scholarly Research and Innovation*, *15*(12), 01-40. https://doi.org/10.63125/awvjew66

[113]. Sumaiya Sadmeen, M., & Kazi Mohammad Khalid, A. (2023). Multi-Source Remote Sensing and Dem-Derived Hydrogeomorphic Indicators for High-Resolution Urban Flood Susceptibility Mapping: A Spatial Cross-Validation Framework. *Journal of Sustainable Development and Policy*, *13*(07), 01-46. https://doi.org/10.63125/z4xt7k38

[114]. Tahmedur, R. (2024). Machine-Learning-Assisted Project Cost Estimation for Technology-Intensive Engineering Programs. *International Journal of Scientific Interdisciplinary Research*, *14*(12), 01–54. https://doi.org/10.63125/7gfgm030

[115]. Tahmedur, R. (2025). Development of a Decision-Support Model for Agile Hybrid Engineering Project Delivery in Technology Driven Organizations. *Review of Applied Science and Technology*, *4*(04), 339–381. https://doi.org/10.63125/v7jjv448

[116]. Tefera, O., & Otieno, J. O. (2024). Factors influencing implementation of the integrated financial management information system for managing public finances. *Journal of Governance and Regulation*, *13*(2), 145. https://doi.org/10.22495/jgrv13i2art14

[117]. Tetteh, L. A., Muda, P., Susuawu, D., Sunu, P., & Aneyire, T. K. (2022). Exploring the use of an integrated financial management information system (IFMIS) in the local government institutions in Ghana. *Information Resources Management Journal*, *35*(1), 1-25. https://doi.org/10.4018/irmj.298973

[118]. Toffaha, K., Bernardelli, L. V., Kortt, M. A., Wijeweera, A., & Simsekler, M. C. E. (2026). Predicting economies of scale in Brazilian municipal expenditure: Panel econometrics and interpretable machine learning. *Environment and Planning B: Urban Analytics and City Science*. https://doi.org/10.1177/23998083261502609

[119]. Torres-Berru, Y., & López Batista, V. F. (2021). Data mining to identify anomalies in public procurement rating parameters. *Electronics*, *10*(22), 2873. https://doi.org/10.3390/electronics10222873

[120]. Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Management*, *14*(3), 207-222. https://doi.org/10.1111/1467-8551.00375

[121]. Tuyishimire, O., & Murorunkwere, B. F. (2024). Applications of big data analytics in tax compliance monitoring: A case study of Rwanda's value-added tax. *CESifo Economic Studies*, *70*(4), 578-587. https://doi.org/10.1093/cesifo/ifae027

[122]. Van Calster, B., McLernon, D. J., van Smeden, M., Wynants, L., & Steyerberg, E. W. (2019). Calibration: The Achilles heel of predictive analytics. *BMC Medicine*, *17*(1). https://doi.org/10.1186/s12916-019-1466-7

[123]. Viana, L. C., Romão, A. L., & Fernandes, G. (2026). Digital transformation in supreme audit institutions: An exploratory analysis of the Napoleonic model in the European Union. *Journal of Public Budgeting, Accounting & Financial Management*, 1-23. https://doi.org/10.1108/jpbafm-11-2025-0312

[124]. Wachs, J., Fazekas, M., & Kertész, J. (2020). Corruption risk in contracting markets: A network science perspective. *International Journal of Data Science and Analytics*, *12*(1), 45-60. https://doi.org/10.1007/s41060-019-00204-1

[125]. Wallimann, H., & Sticher, S. (2023). On suspicious tracks: Machine-learning based approaches to detect cartels in railway-infrastructure procurement. *Transport Policy*, *143*, 121-131. https://doi.org/10.1016/j.tranpol.2023.09.010

[126]. Williams, D. W., & Kavanagh, S. C. (2016). Local government revenue forecasting methods: Competition and comparison. *Journal of Public Budgeting, Accounting & Financial Management*, *28*(4), 488-526. https://doi.org/10.1108/jpbafm-28-04-2016-b004

[127]. Yang, L. (2025). Predictive modeling of tax compliance risks: A comparative study of machine learning approaches. *PLOS One*, *20*(9), e0331715. https://doi.org/10.1371/journal.pone.0331715