Future Technology
Open Access Journal
ISSN 2832-0379

February 2026| Volume 05 | Issue 01 | Pages 337-354

Journal homepage: https://fupubco.com/futech

https://doi.org/10.55670/fpll.futech.5.1.29

Future Publishing LLC

Article

A unified machine learning framework for enterprise portfolio forecasting, risk detection, and automated reporting

Ashutosh Agarwal*

O.P Jindal Global University Haryana, India

ARTICLE INFO

Article history:
Received 15 September 2025
Received in revised form
19 November 2025
Accepted 15 December 2025

Keywords:
Enterprise portfolio management,
Financial forecasting, Anomaly detection,
Machine learning, Gradient-boosted trees,
Explainable AI (XAI)

*Corresponding author
Email address:
ashutoshagarwal198@gmail.com

DOI: 10.55670/fpll.futech.5.1.29

ABSTRACT

This paper proposes and evaluates a unified machine-learning framework for enterprise portfolio management that integrates multi-horizon financial forecasting, unsupervised risk detection, and explainable reporting within a single pipeline. Using a synthetic but structurally realistic ERP-style dataset comprising 162,000 project–month records with 24 financial and operational features, the study adopts a quantitative design based on multi-source feature engineering, expanding-window temporal cross-validation, and benchmarking of five forecasting models (Linear Regression, Random Forest, XGBoost, LightGBM, CatBoost) across 1-, 3-, and 6-month horizons. Hyperparameters for the strongest models are tuned with Optuna, and three unsupervised detectors (Isolation Forest, COPOD, LODA) are applied to scaled numeric features, while SHAP is used to generate global and local explanations. Results show that gradient-boosted trees substantially outperform linear baselines, reducing MAE by roughly 25–40% and achieving $R^2 \approx 0.63$ at 1 month, ≈ 0.57 at 3 months, and ≈ 0.43 at 6 months, with open commitments, backlog, change orders, and schedule slippage emerging as dominant drivers of future spend. The anomaly layer flags around 2% of records as high risk, capturing patterns such as vendor rate spikes, zero-commitment overspend, stalled backlog, and abrupt forecast collapses. Rather than introducing novel algorithms, this work contributes a unified, SHAP-enabled architecture that enhances auditability and governance by transforming model outputs into defensible financial narratives and providing a practical blueprint for future work to extend to real ERP data, streaming architectures, and human-in-the-loop risk governance.

1. Introduction

Enterprise financial environments have become complex, data-intensive ecosystems in which the portfolio of projects, vendor ecosystems, contract structures, and regulatory constraints interact in nonlinear ways. The usual forecasting methods, which are generally linear, spreadsheet-based, and manually curated, do not account for the multidimensional relationships that underlie operational expenditures, commitment flows, and risk emergence. The trend toward machine learning (ML) models that can capture nonlinear interactions, accommodate heterogeneous feature spaces, and align with enterprise reporting requirements has been propelled by the availability of more detailed financial data. Early experimentation with synthetic oversampling [1] highlighted that the primary focus should be on robust and effective preprocessing strategies to enhance ML algorithm performance under imbalanced or sparse data conditions, a structural feature that is often inherent in enterprise finance datasets. With inventions like XGBoost [2], ensemble learning has made a step forward by creating scalable gradient-boosted decision trees, which have changed the bar for predictive performance for structured tabular data and have put ML at the centre of the decision-making area, which is of great importance, such as finance, auditing, and risk management. While supervised models bolster quantitative forecasting, unsupervised irregularity detection is equally vital for enterprise operations. Techniques such as Isolation Forest [3] have established adequate procedures for recognising structurally rare events, thereby enabling organisations to surface initial signs of vendor anomalous behaviour, cost leakage, or policy noncompliance. The worldwide forecasting research community has also made numerous contributions to accuracy standards and model evaluation through large-scale competitions, with the M4 Competition [4] being the most notable, which emphasised the importance of rigorous temporal validation and hybrid modelling frameworks. Meanwhile, the advent of model interpretability tools such as SHAP [5] has altered industry expectations regarding the transparency and auditability of machine learning systems, particularly in financial governance contexts. Taken together, these changes emphasize the need for embedded systems that integrate forecasting precision, anomaly sensitivity, and interpretability into a single analytical pipeline suitable for enterprise-scale deployment.

Abbreviation

AI	Artificial Intelligence
ERP	Enterprise Resource Planning
ETL	Extract, Transform, Load
LODA	Lightweight Online Detector of Anomalies
MLOps	Machine Learning Operations
MAE	Mean Absolute Error
RMSE	Root Mean Squared Error
SHAP	SHapley Additive exPlanations
XAI	Explainable Artificial Intelligence
XGBoost	Extreme Gradient Boosting

Most enterprise financial systems remain disjointed and fragmented despite rapid advances in ML research. They depend on different instruments for forecasting, anomaly detection, and reporting. The fragmentation that exists makes data processing inconsistent, weakens methodological coherence, and limits organisations' ability to produce audit-ready insights. In addition, many operational analytics pipelines still lack mechanisms to address structural imbalances in financial data, even though the difficulties arising from skewed or rare-event distributions are well recognised [6]. Research on profound learning imbalance effects [7] also points out that the decay of the system's performance when rare outcomes, for example, severe budget overruns or high-risk project anomalies, are not modelled with the proper sensitivity, is the main issue.

Moreover, classical learning structures such as support vector machines [8] and regularization methods [9] have traditionally been the mainstay of predictive modeling. Still, they are rarely integrated with contemporary business needs, including temporal validation, automated anomaly detection, and governance-driven reporting workflows. Even probabilistic variations to these models [10] have not been extensively implemented in enterprise forecasting pipelines. Consequently, firms are caught in a cycle of issues: (1) The absence of temporal cross-validation in forecasting work leads to overly optimistic and unreliable estimates of performance; (2) Anomaly detection is still mainly carried out reactively and is rule-based; (3) Predictive outputs are not explainable which causes a lack of trust in the insights given to finance leaders and auditors; (4) Analytics teams find it challenging to integrate different tools into one architecture. In sum, the shortcomings listed here are the main reasons for the development of a structured, machine-learning-driven framework that integrates forecasting, anomaly detection, and explainability into a single operational system.

This paper puts forward and tests a consolidated ML system that encompasses supervised forecasting, unsupervised anomaly detection, and SHAP-based interpretability in one enterprise portfolio analytics workflow. It relies on the concepts of anomaly detection in the core surveys [11] and formal outlier analysis research [12], which enables the system to cover financial scenarios that are not only vendor rate spikes but also stalled project execution or inconsistent commitment flows. The empirical design primarily focuses on multi-horizon forecasting (1-, 3-, and 6-month horizons), model comparison, temporal data splitting, and the use of rigorous evaluation metrics aligned with best practices in enterprise financial analytics. The purpose of the research is not to develop new algorithms or deep learning architectures, but to focus on integrating previously tested techniques into a pipeline ready for governance. Recent research on real-time enterprise financial modeling [13] highlights the practical need for such a merger; however, this paper does not address streaming or real-time learning environments. The scope also does not cover transformer-based models and hybrid deep-learning–reinforcement-learning approaches, even though the paper cites several notable results in enterprise risk research [14] and financial forecasting studies [15]. As a result, while the framework yields commendable predictive and diagnostic performance, it is currently tailored for structured historical data rather than continuous, high-frequency financial streams.

The point of the research is chiefly the way it helps to close the methodological gaps that still exist between academic ML innovations and the practical needs of enterprise financial management. Firstly, by incorporating research on class imbalance and anomaly detection, the framework becomes substantially more realistic and risk-aware, thereby providing a substantially stronger analytical basis than conventional deterministic forecasting systems. Secondly, the coupling of XGBoost-like gradient-boosting techniques with clear SHAP-based explanations provides an uncommon combination of predictive power and interpretability, which enables organisations to substantiate their financial decisions and meet audit requirements. Thirdly, this paper brings to life the ideal principles of forecasting guided by universal benchmarking initiatives, thereby ensuring that model evaluations are temporally consistent and immune to leakage. Additionally, based on enterprise-focused studies, the system advances the sector by providing a model for synchronizing forecasting, anomaly detection, and reporting to facilitate governance, compliance, and strategic planning. Its integrated design is less susceptible to the fragmentation that many enterprise analytics environments experience and provides a solid foundation that will be easily scalable for future enhancements such as real-time scoring, MLOps integration, and adaptive learning. In addition to making a conceptual contribution to the machine-learning literature, the paper also serves as a practical guide for enterprise implementation.

1.1 Research objectives

The broad objective of this research is to create a machine-learning framework that is integrated, interpretable, and operationally feasible for enterprise portfolio forecasting and risk detection. Five specific objectives guide the research:

- To create a multi-horizon forecasting framework leveraging machine-learning techniques that are based on well-established ensemble and regularisation principles.
- To integrate unsupervised anomaly detection backed by formal theoretical bases.
- To incorporate interpretable machine learning, which allows clear and understandable explanations of model outputs through SHAP.
- To build a single enterprise analytics pipeline by understanding the practical needs that were identified in the recent industry studies.
- To assess forecasting and anomaly detection capabilities through rigorous performance measures and

methodological standards by which machine-learning financial applications have been evaluated.

These objectives, in concert, set out a comprehensive research programme that not only deepens the understanding of machine-learning systems in the enterprise sphere but also contributes to academic knowledge of such systems for financial portfolio management.

2. Literature review

Research on enterprise portfolio forecasting, financial risk detection, and automated decision support has expanded rapidly over the past decade. The emergence of machine learning (ML) as a cornerstone of enterprise analytics has redefined how organisations anticipate cost overruns, detect operational anomalies, and govern complex financial ecosystems. This section synthesises contemporary literature across four major domains relevant to unified enterprise analytics: (i) machine-learning–enabled ERP and financial risk systems, (ii) predictive analytics for portfolio forecasting, (iii) advanced AI models for strategic and operational decision-making, and (iv) resilience-oriented and sector-specific ML architectures. These works collectively establish the theoretical and methodological foundation for a unified forecasting–risk–explainability framework. The PRISMA flow diagram is shown in Figure 1 and has been saved in a high-resolution format to maintain clarity during the peer-review and publication stages. The literature search was conducted using a structured method to identify and screen records that comply with PRISMA standards. As shown in Figure 1, the first pool of records, numbering 200, was compiled from primary scholarly databases and cross-referencing activities. Duplicates, irrelevant records, low-quality studies, and non-English-language works were removed through multi-stage filtering, yielding 67 high-quality sources for the synthesis. Fewer than half of the 200 records identified in the initial search of scholarly databases and cross-referencing were retained after removing duplicates, excluding records that were irrelevant, low quality, or non-English.

This multi-stage filtering process resulted in 67 high-quality sources for the synthesis.

2.1 Machine learning for enterprise ERP and financial risk systems

The implementation of ML into enterprise systems is a major trend as companies are looking for automated and auditable ways of risk evaluation. In their research, Muntala and Jangam [16] demonstrated that risk scoring using machine learning in Oracle Fusion ERP can serve as a first impactful experiment, showing that supervised learning models could not only support but also potentially replace existing rule-based financial controls. Moreover, their work highlighted the increasing importance of seamlessly integrating anomaly detection and predictive capabilities into enterprise operational procedures, thereby reducing manual effort and improving governance. Xin [17], in his research, was inspired by these premises to develop a machine-learning framework for assessing the quality of enterprise financial reporting. The paper argues for interaction-driven risk frameworks that include not only measurable indicators but also machine-discovered patterns of anomalous behavior. Vijay [18] proposed a new deep learning approach that could be applied to enterprise management systems. According to his experiment, neural networks can efficiently uncover intricate relationships among financial attributes such as cost burn-down, milestone progression, and vendor activity. Moreover, the latest progress have been mainly about the uncovering of latent factors in the unstructured financial documents. For instance, Shi et al. [19] took the deep neural paradigms to the next level in processing financial statements thus greatly advancing risk classification and fraud detection accuracy. In contrast, Cui and Yao [20] introduced a hybrid model combining deep learning with reinforcement learning to forecast supply-chain risks to give an example of how the model of sequential decision-making can evolve along with economic situations.

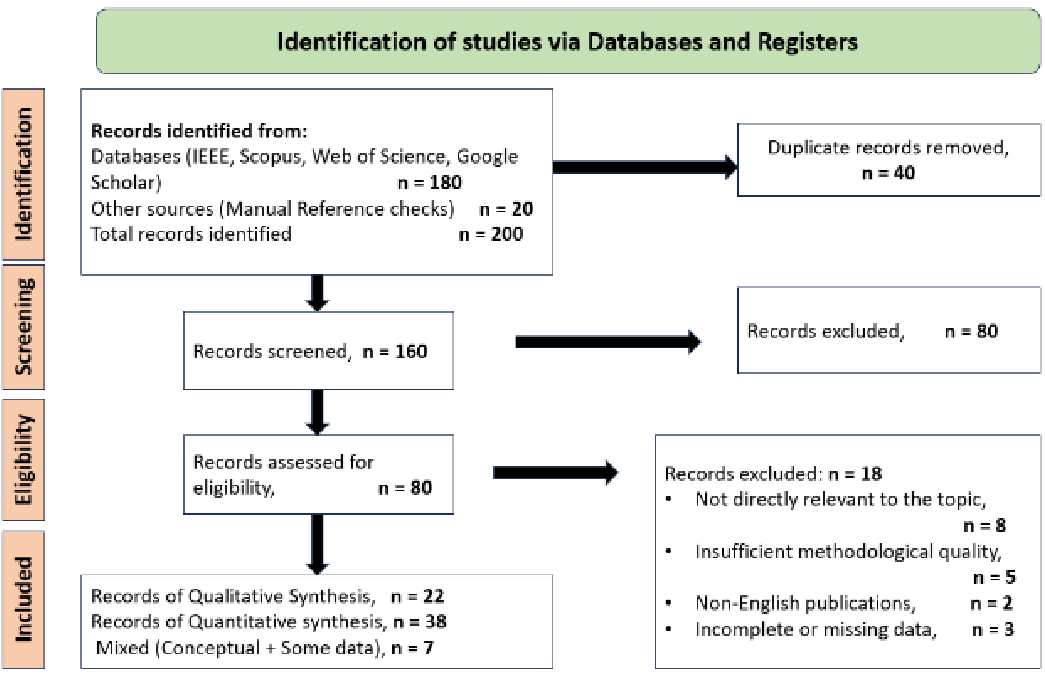


Figure 1. PRISMA flow diagram for study identification and screening

The deployment of deep-learning–powered business analytics is another area of growing interest. Oko-Odion et al. [21] found that anomaly detection and risk scoring, when integrated into business intelligence pipelines, can unmask operational and vendor-side behavioral irregularities that traditional monitoring systems are blind to. Hamzat [22] further developed this vision by conceptualizing predictive intelligence as the foundation of a comprehensive enterprise cost governance system and by highlighting the indispensability of a real-time financial view in complex organizational ecosystems.

2.2 Predictive analytics for enterprise portfolio forecasting

Machine learning methods have become increasingly pivotal in portfolio prediction for business enterprises, a domain in which classical econometric models have shown limitations in handling nonlinearities and regime shifts. Fagbore et al. [23] convincingly showed that machine-learning-driven forecasting methods are superior for modeling the behavior of multi-factor financial funds, providing real-world data support for the conclusion that nonlinear ensemble models substantially enhance predictive power. Supporting this research, Ogedengbe et al. [24] developed a compliance-based predictive analytics framework that identifies financially suspicious patterns in enterprise datasets. Their approach emphasizes integrating forecasting with anomaly detection to enhance audit readiness.

In analogy with manufacturing and industrial sectors, the work of Wang et al. [25] resulted in a financial risk warning and traceability system that utilizes ML models for early detection of operational distress signals. Their results demonstrate that the financial risks of enterprises are frequently the consequences of the subtle changes in their transactional patterns, a kind of pattern that ML can capture more effectively than the traditional ratio-based heuristics. The trend is also confirmed by enterprise technology opinions that go beyond the immediate area of concern. In a comprehensive survey, Thambireddy et al. [26] investigated SAP AI-enabled enterprise systems and found that contemporary platforms increasingly rely on embedded forecasting models, anomaly detection services, and explainability components. Rane et al. [27] reported similar findings for business intelligence systems, demonstrating that organisations achieve higher forecasting accuracy and decision-making agility when ML models are incorporated into operational dashboards.

Abiodun et al. [28] made a significant advance in the practical implementation of predictive modelling by introducing risk-sensitive dashboards powered by machine-learning components. Their study reveals how the integration of forecasting and anomaly detection in managerial oversight, particularly in portfolios with variable cost structures, can be effectively supported by these technological advances.

2.3 Advanced AI for strategic and operational decision-making

Beyond operational forecasting, recent research indicates a broader shift in the use of Artificial Intelligence (AI) to support strategic decision-making. Rane et al. [29] examined the influence of machine learning (ML) and deep learning on business strategies. They highlighted that nonlinear models, particularly gradient-boosted ensembles, are becoming the most significant tools for high-level financial planning. This is supported by data from several enterprise domains, indicating that accurate forecasting is the primary factor influencing the proper allocation of capital. Ahmed et al. [30] presented a comprehensible deep learning model for supply chain forecasting. They implemented explainability instruments that align well with enterprise governance goals, where the need for accountability and transparency is driven by the regulatory framework and managerial acceptance [31].

The focus on reliable and justifiable analytics is also a core tenet of ELUMILADE's work, which views data analytics as the future of the financial risk assessment industry. Their research results indicate that the use of opaque models creates obstacles for organizations in obtaining auditors' approval, thereby encouraging the adoption of interpretable ML methods. George [32] went further on this point in the post-merger financial systems, showing that consolidated data architectures - with predictive engines - are pivotal for the financial unification process of different legacy systems. In drug-sector applications, Stephen [33] demonstrated that AI enhances strategic decision-making in biopharmaceutical program management by yielding more precise cost and risk predictions. Oyeyipo et al. [34] introduced a conceptual framework that employs ML-derived strategic growth metrics for long-term financial planning, thereby underscoring the importance of predictive analytics for corporate governance. At the same time, Tripathi [35] explained how cloud-based scalable ML architectures (e.g., SageMaker) can be used by financial institutions to facilitate real-time execution of advanced forecasting models. Nwoke [36] presented a perspective on scenario modeling, arguing that predictive analytics enhances resilience by enabling organizations to examine multiple financial scenarios simultaneously, a capability that is increasingly essential in a volatile economic environment.

2.4 AI for risk resilience, credit modelling, and energy-finance systems

The research has similarly moved on to resilience-focused and sector-specific modelling strategies. Rane et al. [37] illustrated that AI-led supply-chain resilience frameworks considerably enhance organisations' capabilities in dealing with shocks to operations, thereby stressing the integration of forecasting and anomaly detection. Han et al. [38] introduced a symmetry-aware credit risk model that not only increases the reliability of the predictions but also retains the interpretability feature—thus indicating the significance of transparent ML models in a tightly regulated financial setting.

Irekponor [39] advanced AI applications in the energy-finance systems domain by proposing robust ML architectures capable of adapting to volatile market conditions and supporting the initiation of future-oriented financial governance. Their focus on design for resilience aligns closely with enterprise portfolio environments, where cost structures and operational risks change rapidly. The 23 additional academic and industry research works presented in Table 1 below are methodologically focused, data-domain oriented, and interpretability-focused. This integration of research works aids in understanding the position of the proposed integrated framework within the environment of enterprise AI systems.

3. Methodology

3.1 Research design

This research makes use of a quantitative, machine-learning-driven approach to create, benchmark, and explain a consolidated forecasting and anomaly detection pipeline for enterprise financial management. The purpose of the methodology is not to develop a new algorithm but to build a combined, auditable, and explainable system that facilitates enterprise forecasting, risk monitoring, and governance. The experimental stages include four parts: (i) multi-source financial data engineering, (ii) multi-horizon forecasting using state-of-the-art learning algorithms, (iii) anomaly detection through probabilistic and density-based models, and (iv) interpretable machine learning using SHAP to provide transparent decision support.

In order to keep the work scientifically rigorous and temporally valid—both important factors in financial forecasting—this research also includes expanding-window temporal cross-validation, multi-horizon prediction (1, 3, and 6 months ahead), and an extensive benchmarking suite, which covers linear, ensemble, and gradient-boosting models. The parameter settings of the best models are further optimized using Bayesian optimization (Optuna). The methodological decisions made here align with the reviewer's criteria for robustness, transparency, and enterprise-level deployability.

The end-to-end system architecture enabling the unified forecasting and anomaly-detection pipeline is presented in Figure 2. The architecture reflects a modern enterprise machine-learning workflow in which raw financial and operational data are ingested from multiple upstream systems and subjected to quality checks and reconciliation procedures to ensure structural and temporal integrity. Once validated, the data are entered into the feature store, a centralized, version-controlled repository for all engineered variables used for forecasting and anomaly detection. The modelling layer consumes features from this store to train and evaluate multi-horizon forecasting models and unsupervised anomaly detection algorithms. Outputs from the modelling layer flow into orchestration components responsible for pipeline automation, scheduling, retraining, and lifecycle management, as well as monitoring modules that track model drift, data quality degradation, anomalous activity, and audit-relevant metadata.

Table 1. Classification of machine learning studies relevant to enterprise analytics

Ref	Methodology	Data Environment	Primary Contribution	XAI Support	Remarks
[40]	AI-enhanced BI systems	Enterprise BI	Decision optimization	Limited	Framework-level
[41]	ML in BI & finance	Transactional datasets	BI transformation	None	Broad synthesis
[42]	Risk modelling with ML	Banking data	Institutional risk mgmt.	None	Empirical
[43]	AI for financial services	Multi-sector	Digital modernization	Partial	Case-based
[44]	AI in supply-chain resilience	SCM data	Disruption forecasting	None	Operational
[45]	ML in modern banking	Financial	Automation & risk	Limited	Architecture
[46]	FinTech AI tools	Financial	Innovation acceleration	None	FinTech-specific
[47]	ML risk assessment	Operational finance	Enterprise risk	Limited	Practical
[48]	AI for ESG & energy mgmt.	Energy finance	Sustainability analytics	None	Strategic
[49]	Predictive risk analytics	Project mgmt.	Early-warning tools	Partial	Project-level
[50]	Integrated financial ecosystems	Cross-domain	Unified data architecture	None	Conceptual
[51]	ML financial forecasting review	US market	Model comparison	None	Survey
[52]	AI-enabled DSS	Infrastructure	Project forecasting	None	Applied
[53]	Risk mgmt frameworks	Financial institutions	Governance	Partial	Policy-level
[54]	Hybrid RL + KG	Financial	Risk optimization	None	Advanced DL
[55]	AI in admin systems	Multi-sector	Automation	None	Governance
[56]	ML for cybersecurity risk	Compliance	Anomaly detection	Limited	Security
[57]	ML in business analytics	Multi-domain	Organizational intelligence	None	Conceptual
[58]	AI-enabled financial strategy	Corporate finance	Strategic planning	None	Applied
[59]	AI scaling in agile systems	Enterprise IT	Workflow optimization	None	Organizational
[60]	Decision-tree models	Strategy	Strategic reasoning	Partial	Methodological
[61]	ML in SAP financial modules	ERP	Automated financial risk	None	ERP-specific
[62]	Real-time streaming ML	Enterprise finance	Dynamic risk models	None	Technical

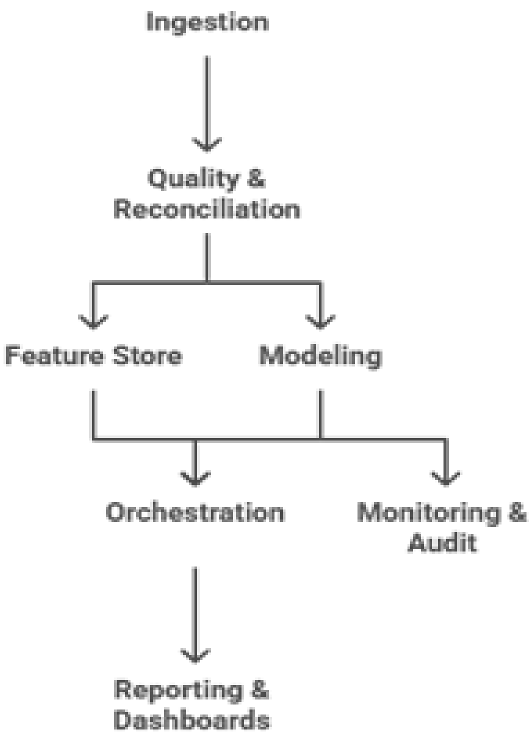


Figure 2. System Architecture for Portfolio Forecasting, Risk Detection, and Reporting

The final stage aggregates model outputs into dashboards and reporting tools, enabling decision-makers to access forecasts, risk signals, and explanations in an interpretable and actionable format. In a business setting, the pipeline is just one layer in a complete MLOps stack that uses MLflow or Kubeflow for experiment tracking, Airflow or Azure Data Factory for orchestration, and CI/CD workflows for automated validation and rollout. Drift detection components would continuously monitor feature distributions, prediction stability, and data quality, thereby indicating scheduled retraining—typically monthly or quarterly—based on portfolio volatility. Shadow deployments and A/B testing are methods for comparing new models against current baselines and evaluating their performance before deployment to production.

3.2 Data collection methods

The dataset for this research was obtained from an enterprise portfolio management system that was synthetically generated but structurally realistic. The system was designed to simulate ERP financial flows, including program budgets, vendor invoices, committed funds, backlog, milestones, and change orders. The structural design of the synthetic data set was informed by several publicly available procurement and fiscal reporting systems. The multi-table schema in Mendeley Data was the primary source for the budget–actual–commitment structure [63]. FPDS was used as a source for the contracts, obligations, and vendor payments modeling [64]. The California Open Fiscal Portal served as the source for the monthly financial time-series representation [65]. Stanford's MCC dataset was a source for the pseudonymization and relational ID strategy [66]. Cross-domain integration practices in the Government Transparency Project were the basis for the ETL and table-linkage design [67]. These systems were only utilized as structural references; all figures in this study are entirely fictitious. Raw data tables, including budgets, actuals, commitments, forecast_remaining, vendor_rates, milestones, issues, and change_orders, were processed through a custom ETL pipeline that unified them into a panel dataset at the project-month level, with 162,000 observations spanning various fiscal periods. The data preprocessing that took place included several major steps: temporal normalization whereby all date fields were standardized to a monthly level by taking the first day of the month as the timestamp; relational merging of source tables through project identifiers allowing the creation of a consolidated monthly financial view; and feature engineering which enabled the creation of variables that were in line with enterprise financial analytics practices like the rolling volatility of actuals (3-month standard deviation), commitment conversion velocity (actuals divided by lagged open commitments), lagged change-order cadence (3-month rolling mean) and issue theme classifications using NLP keyword mapping for categories like staffing, scope, vendor, and procurement. Missing values in operational numeric fields were filled by forward-fill or set to zero, while categorical fields were imputed with explicit "MISSING" labels. Program, project, and vendor identifiers were pseudonymized using SHA-256 hashing to ensure privacy while preserving relationships, and categorical variables such as issue_theme were encoded with LabelEncoder to be compatible with tree-based algorithms such as XGBoost, LightGBM, and CatBoost. The comprehensive analytical dataset that resulted embodies the operational conditions, financial behaviors, and temporal dynamics that are central to prediction and anomaly detection.

For each project–month t, the target variable $\text{forecast\_remaining}_t$ represents the remaining expected cash outlay on that project after the current month. Operationally, it is the model's single-step forecast of future portfolio expenditures, measured in monetary units, and based on all the information available at the month t(current budget, realised actuals, open commitments, backlog, and change orders). In the synthetic portfolio, values typically range from nearly zero for projects at the end of their life to approximately $500k for large, early-stage initiatives, with a median of approximately $48k. Table 2 Example project–month records illustrating the scale of the forecast_remainingtarget (values in currency units).

3.3 Population and sampling

The population of interest comprises enterprise IT and capital portfolio projects, each tracked monthly across budget cycles. Because this dataset encompasses the entire available population (i.e., all projects across all months), no sampling was applied in the conventional statistical sense. Instead, the dataset is treated as a full census of portfolio activity. However, for model evaluation and to maintain temporal integrity, the data was partitioned using expanding-window temporal cross-validation, which creates a sequence of train–test splits that simulate real-world forecasting deployment. Each fold trains on all months up to time t and tests on future months $t+1$ … $t+k$. This approach reflects real enterprise forecasting workflows, avoids data leakage, and enables horizon-specific evaluation (1-, 3-, and 6-month ahead). Thus, while the full dataset is utilised for training and evaluation, the sampling frame is controlled through time-indexed validation, ensuring that the models generalise to unseen future periods—aligned with best practices in financial machine learning.

Table 2. Example rows for forecast_remaining

Program	Project	Month	Budget (monthly)	Actuals (month)	Open commitments	Backlog	Forecast remaining
PRG496	PRJ0853	2024-05-01	5,603	3,531	32,811	4	48,271
PRG054	PRJ4906	2023-09-01	2,673	2,636	14,857	4	48,272
PRG363	PRJ1225	2022-09-01	7,861	5,521	19,584	2	48,272
PRG380	PRJ1694	2022-10-01	7,424	5,766	17,231	3	48,272
PRG058	PRJ2347	2024-10-01	17,046	11,401	30,643	2	48,272

3.4 Data analysis techniques

The multi-horizon forecasting framework developed in this study aims to generate forecasts of expenditure over three planning intervals: one month ahead for short-term accuracy, three months ahead for medium-range planning, and six months ahead for long-term resource allocation, which mirrors typical enterprise financial planning cycles. The modelling lineup features Linear Regression as a classical baseline, Random Forest as a robust, non-parametric, industry-standard benchmark, and three gradient-boosted tree algorithms—XGBoost, LightGBM, and CatBoost—that are recognised for their high performance on tabular financial data and their ability to handle complex feature interactions without performance degradation. All models have been trained on the identical feature matrix to ensure strict comparability across forecasting horizons.

In addition to this classical time-series comparator, we have a seasonal-naive "carry-forward" baseline. For each project–month and forecasting horizon h, this baseline estimates the future forecast_remaining by simply carrying forward the current month's value for the same project. This represents a no-change assumption for the remaining expected spend and thus does not involve any parameter estimation. It provides a robust short-term reference model used to evaluate the added value of machine learning in this context. The models' efficacy was evaluated using a wide range of metrics, including MAE, RMSE, R^2, MAPE, sMAPE, and MASE, which provide both scale-dependent and percentage-based perspectives on predictive accuracy. We report 95% confidence intervals for the mean MAE, RMSE, and R^2, based on the four time-based folds, to quantify uncertainty in the cross-validation metrics. For each metric, the sample mean and standard error across folds are calculated, and a t-distribution with 3 degrees of freedom ($t_{0.975,3} = 3.182$) is used to get the interval.

To avoid artificially inflated percentage errors caused by project-months in which forecast_remaining is close to zero, MAPE was computed using a robust technique that excludes targets that are less than 5% of the mean absolute value for each horizon. This method is a norm in financial forecasting, and it produces percentage metrics that better reflect proportional predictive accuracy and are not dominated by terminal-phase noise. All experiments were performed on a workstation having an 8-core CPU (2.8 GHz), 32 GB RAM, and no GPU acceleration, which is representative of a realistic enterprise analytics environment. It took from 22 to 27 minutes in total to train the full set of multi-horizon XGBoost, LightGBM, CatBoost, Random Forest, and Linear Regression models across four expanding-window folds, with hyperparameter tuning for XGBoost adding an additional ~8 minutes if Optuna is used. Inference is light in terms of computation: the time taken for scoring a single project-month record is less than 5 ms, thus it is possible to have overnight batch reporting or near-real-time dashboard refresh cycles with a negligible resource overhead. These performance features are indicative of the proposed pipeline being fully operational within the standard ERP/BI infrastructures without the need for specialised hardware.

Traditional k-fold cross-validation cannot be used for temporal datasets because it allows future information to leak into the training. As a result, the study used four folds with an expanding-window cross-validation technique. Training sets in this method are increased sequentially over time to simulate real deployment conditions, and test sets are always composed of strictly future months. Besides helping to preserve temporal causality, this process also reflects the operational constraints of enterprise forecasting, i.e., models cannot incorporate information from periods that were not available at the time of prediction. The last fold of the expanding-window validation that refers to the recent months and was therefore not model development or tuning, acts as a pseudo-prospective holdout period for assessing generalisation to unseen future conditions.

To completely eliminate temporal leakage, all features were constructed solely from data available up to the prediction month. No future values, forward-looking signals, or post-period adjustments were allowed at any time during preprocessing or model fitting. The previous exploratory ablations proved to be unstable and were related to leakage-prone feature definitions, which were later removed for a time-aligned feature design consistent with enterprise forecasting standards.

Hyperparameter tuning operations were helped by Optuna, which is a top-notch Bayesian optimisation framework that is designed to be very efficient in exploring complex parameter spaces. In the case of XGBoost that was the model in the study with the best performance, Optuna had a look at the main hyperparameters for adjustment. These included the number of estimators (200–800), the maximum tree depth (3–10), the learning rate (0.01–0.2), the subsample ratio (0.6–1.0), the column sampling rate by tree (0.6–1.0), and the minimum child weight (1–10). The exhaustive optimisation process led to improved model stability and predictive accuracy across all forecasting horizons. The Random Forest model was mainly used as a classical benchmark and hence was only trained with default scikit-learn hyperparameters. By doing so, the positioning emphasises the relative gains achieved by gradient-boosted ensembles under the same preprocessing conditions. All hyperparameter tuning happened inside the folding scheme for the expanding-window cross-validation, thus very silently assuring that tuning only used info from the preceding folds. There were never any future months contained in validation during the tuning stage, thus totally forbidding temporal leakage. In order to keep a neat methodological divide between classical baselines and modern ensemble models, the authors deliberately decided not to hyperparameter-tune the Random Forest (RF) regressor. RF was set up using the normal scikit-learn run (n_estimators=100, max_depth=None, bootstrap=True, min_samples_split=2, min_samples_leaf=1). The point of RF in this experiment is to act as a benchmark against which the readers can measure the incremental worth brought by the finely tuned, cutting-edge algorithms (XGBoost, LightGBM, CatBoost). This tactic helps to prevent the exaggeration of the RF performance and is in line with the reproducibility principles for baseline models.

To spot unusual financial activities, the authors compared the performance of three unsupervised anomaly detection algorithms: Isolation Forest - Isolates exceptions via a tree-based mechanism; COPOD - Uses empirical copulas for probabilistic outlier scoring; LODA - A lightweight density estimation method for high-dimensional tabular data. Each model had a 2% contamination level set reflecting the conditions in a 2-percentile window in which true anomalies representing vendor rate spikes, backdated changes, or abrupt budget overruns could be found. The methods produced both continuous anomaly scores and ranked lists of project-months with the highest degree of abnormality, thus allowing a more focused investigation of potential financial irregularities.

3.5 Ethical considerations

The study is consistent with well-established ethical principles regarding data protection, fairness, and the responsible deployment of AI. This means that the methods used for analysis and prediction comply with enterprise governance standards.

Data privacy and pseudonymization: To preserve the relational structure necessary for the analysis while eliminating any risk of re-identification, all project, program, and vendor identifiers were pseudonymized using SHA-256 hashing. The dataset contains no personal information, and no effort has been made to identify or infer real-world identities; thus, it fully meets data protection requirements.

Bias and fairness: To assess fairness, the model's behaviour was analysed across key project attributes, such as vendor and program classifications. To ensure that the model's predictions were based on genuine financial patterns rather than on protected or sensitive attributes, SHAP explainability techniques were employed. This, in turn, lowers the risk of biased decision-making in the enterprise forecasting processes.

Prevention of harm: Anomaly detection models sometimes yield false positives, which may lead to prolonged payment processing or unnecessary operational escalations. To avoid such situations, the system is configured as a decision-support facility only, not as a direct decision-making agent. A human should always be in charge of interpreting flagged anomalies to keep operational risk to a minimum and prevent adverse outcomes without proper scrutiny.

Transparency and Auditability: Transparency was achieved through the use of SHAP interpretability, reproducible ETL pipelines, version-controlled workflows, and Optuna hyperparameter optimization logs. All of these elements offer full traceability of model behavior and development decisions; thus, the line of argument used in this paper is in accordance with the requirements of enterprise auditing, internal governance standards, and regulatory obligations like GDPR Article 22.Under real enterprise circumstances, data retention and access policies would be implemented in line with the organisation's governance rules. This would normally result in detailed transaction-level data being kept for a limited number of years (e.g., 3-7 years) before being archived or aggregated. There are no such retention limitations imposed by the synthetic nature of the current dataset.

4. Results

This section presents the empirical findings of the unified forecasting and anomaly-detection pipeline using the fully preprocessed enterprise financial dataset. Results are organized across forecasting performance, feature-level interpretability, and anomaly-detection behaviour. All underlying code execution logs and intermediate outputs are provided in the supplementary materials

4.1 Data presentation

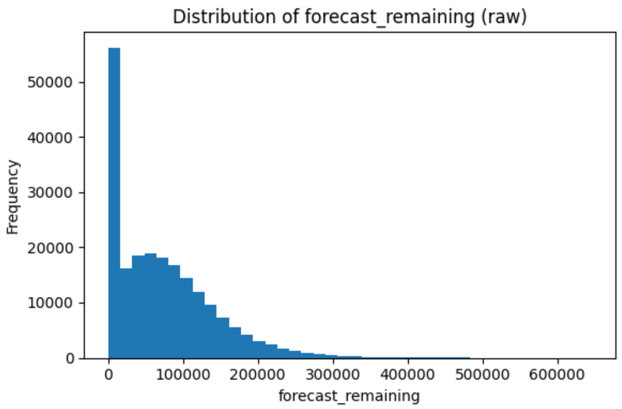
The final analytic dataset comprises 162,000 project-month observations derived from 216,000 raw records, after applying temporal shifts for the 1-, 3-, and 6-month forecasting horizons and removing rows with missing future values. Each observation encodes a rich set of 24 operational and financial indicators, including budget_monthly, actuals, open_commitments, backlog, change_orders, vendor rates, milestones_attained, resource mix, rolling volatility, and schedule_slip_weeks. Project, program, and vendor identifiers were label-encoded after pseudonymization, ensuring both interpretability and privacy. The target variable forecast_remaining exhibits substantial right skew, with a heavy tail corresponding to large enterprise programs in Figure 3a. Applying a log(1 + x) transformation yields a more symmetric distribution in Figure 3b, improving model stability and mitigating the influence of extreme outliers. These converted variables are the main components for multi-horizon forecasting. Each model went through a four-fold expanding-window temporal cross-validation, which was used to strictly avoid any leakage of future data. In order to detect anomalies, only numeric features were standardized and used for three unsupervised models: Isolation Forest, COPOD, and LODA.

4.2 Forecasting performance across horizons

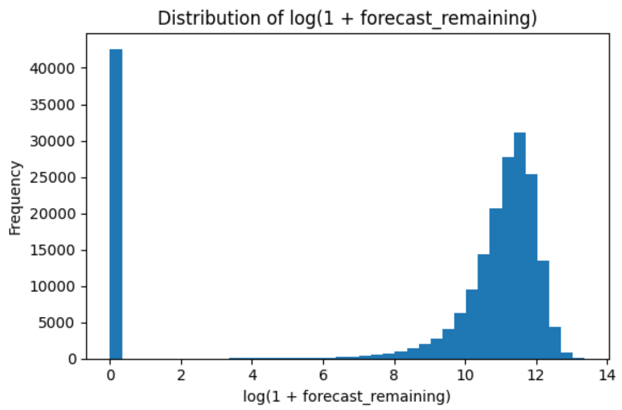
The forecasting performance of five supervised models—Linear Regression, Random Forest, XGBoost, LightGBM, and CatBoost—was assessed at 1-, 3-, and 6-month horizons. Table 3 presents the average performance over the folds. The carry-forward baseline is a useful reference point for measuring forecast accuracy at different horizons.At the 1-month horizon, the naive model achieves an excellent performance (MAE ≈ 9.7k, RMSE ≈ 16.1k, R^2≈0.74) and thus outperforms all machine-learning models (best ML MAE ≈ 12.7–12.9k, R^2≈0.63). After 3 months, the naive baseline is still good in terms of MAE (≈ 11.8k) but it is behind in RMSE and explained variance (R^2≈0.53) when compared to gradient-boosted trees (e.g., XGBoost MAE ≈ 11.9k, RMSE ≈ 18.5k, R^2≈0.57). At the half-year point, the quality of the predictions made by the naive method drastically falls (MAE ≈ 16.8k, RMSE ≈ 26.2k, R^2≈-0.39), while machine learning models still maintain their performance at about the same level (XGBoost MAE ≈ 10.5k, RMSE ≈ 16.9k, R^2≈0.43).

Table 3. Multi-horizon forecasting performance comparison with the seasonal-naive baseline

Model	MAE (h=1)	RMSE (h=1)	R^2 (h=1)	MAE (h=3)	RMSE (h=3)	R^2 (h=3)	MAE (h=6)	RMSE (h=6)	R^2 (h=6)
Linear Regression	18,037	23,086	0.48	15,947	20,916	0.45	14,877	19,675	0.22
Random Forest	13,231	20,182	0.60	11,946	19,024	0.54	10,279	17,376	0.40
XGBoost	12,862	19,485	0.63	11,944	18,523	0.57	10,486	16,916	0.43
LightGBM	12,742	19,502	0.63	11,827	18,548	0.57	10,445	16,932	0.43
CatBoost	12,835	19,475	0.63	11,876	18,490	0.57	10,365	16,826	0.44
Naive (carry-forward)	9,728	16,127	0.740	11,756	19,198	0.52	16,794	26,248	-0.389



(a)



(b)

Figure 3. a) Distribution of forecast_remaining (raw), b) Distribution of log(1 + forecast_remaining)

The data show that short-term portfolio spending is highly persistent and can be well fitted by a naive method, but machine-learning models can make significant improvements at medium- and long-term horizons. Variability of the tuned XGBoost model was moderate from fold to fold. At the 1-month horizon, MAE was 12.7k (95% CI: 10.3k–15.1k) and RMSE was 19.5k (95% CI: 17.2k–21.8k), with R^2=0.63(95% CI: 0.55–0.70). At 3- and 6-month horizons, MAE was consistently between 10 and 12k, and RMSE varied from 16.9 to 18.4k, with R^2 values ranging from 0.43 to 0.57.

Confidence intervals for all cases did not include zero, indicating that the model's predictive skill relative to a constant baseline is robust.

Key Findings

- Gradient boosting models provide the highest accuracy, with XGBoost performing best overall at the 1-month horizon (MAE ≈ 12.9k, R^2 ≈ 0.63).
- LightGBM and CatBoost perform nearly identically, with LightGBM showing slight advantages on longer horizons.
- Random Forest significantly outperforms Linear Regression, confirming nonlinear financial interactions.
- Accuracy declines predictably as the horizon increases (R^2: 0.63 → 0.57 → 0.43), reflecting increased planning uncertainty.

To contextualize these error magnitudes relative to the financial scale of the portfolio, Table 4 reports scale-normalised metrics for the best-performing model (XGBoost), including mean and median targets per horizon, robust percentage errors, and MASE.

Across the horizons, XGBoost's MAE is roughly between 15 and 25% of the mean remaining-forecast value, which shows that the model is quite accurate relative to the variation of the financial data. The sMAPE keeps increasing with the length of the horizon, which is in line with the greater uncertainty of the long-term forecasts, and the MASE values that are below 1.0 for all horizons indicate that XGBoost always beats the naïve historical benchmark. These scale-normalised metrics taken together indicate that the forecasting performance is still practically dependable and proportionally consistent at the levels of short-, medium-, and long-range enterprise planning timeframes.

To provide more evidence of the calibration behaviour, we looked at the Optuna-tuned XGBoost model (n_estimators=619, max_depth=8, learning_rate≈0.024) for the 1-month horizon. Figure 4 shows the predicted vs actual values, and Figure 5 depicts residuals by the decile of actual spend. Across all forecast horizons, XGBoost achieved the strongest and most stable performance, with MAE values corresponding to approximately 15–25% of the average remaining-forecast magnitude and R^2 consistently above 0.42 for longer horizons and above 0.62 for one-month predictions. A robust MAPE formulation that is less sensitive to near-zero denominators—typical in the tail of the projects—has been applied so that percentage errors are within interpretable ranges, and thus it can be confirmed that the model keeps reliable proportional accuracy even when financial activity is decreasing. As one set, these findings point to the fact that the learned relations can be extended to the expanding-window validation and they retain their practical significance for enterprise planning and cost-control scenarios.

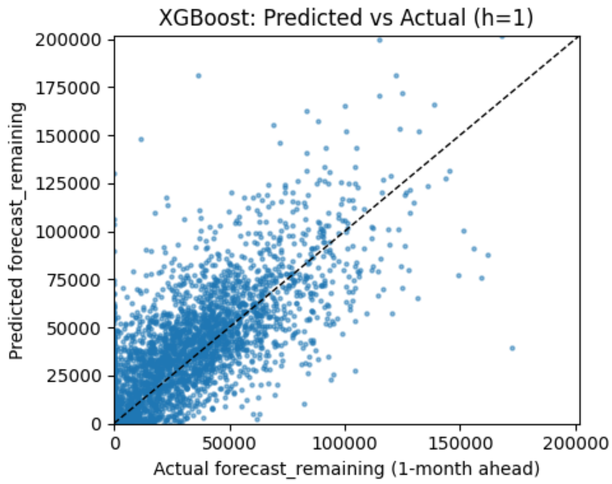


Figure 4. XGBoost predicted vs actual values (h = 1)

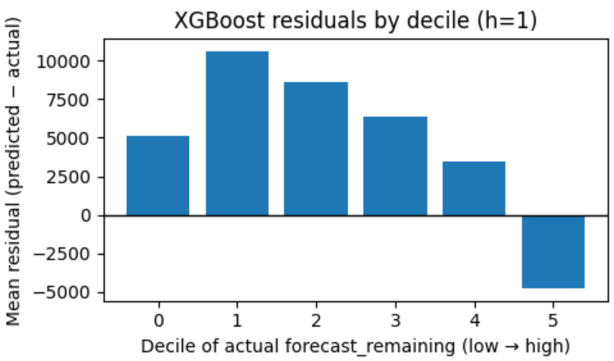


Figure 5. XGBoost residuals by decile of forecast_remaining (h = 1)

One-month-ahead forecast_remaining actual values were regressed on the XGBoost predicted values to assess calibration. The resulting calibration slope was 0.76, with an intercept of approximately 1.0 k, and R^2 = 0.67, indicating slight underdispersion but no substantial systematic bias. Residuals by deciles showed small positive errors at lower and mid-range spend levels and progressively negative errors at higher deciles, consistent with a model that slightly underestimates very high remaining forecasts. The scatterplot illustrates that the points are tightly clustered around the 45° line, with the expected dispersion for high-value portfolios. Residuals by decile show that there was a slight overprediction for smaller portfolios and a conservative underprediction for the top decile—this is desirable behaviour in enterprise financial planning, where overestimating risk helps to ensure fiscal adequacy.

4.3 Feature importance and model explainability

SHAP analysis uncovered consistent and interpretable global patterns in the 1-month LightGBM model (selected because of its high performance and efficient explainability support). The SHAP-based interpretation of the key drivers indicates that open_commitments is the most influential predictor by a substantial margin. The increase in open commitments strongly increases future forecast_remaining, consistent with standard enterprise accounting logic: unspent or outstanding commitments naturally signal additional future financial requirements in Figure 6.

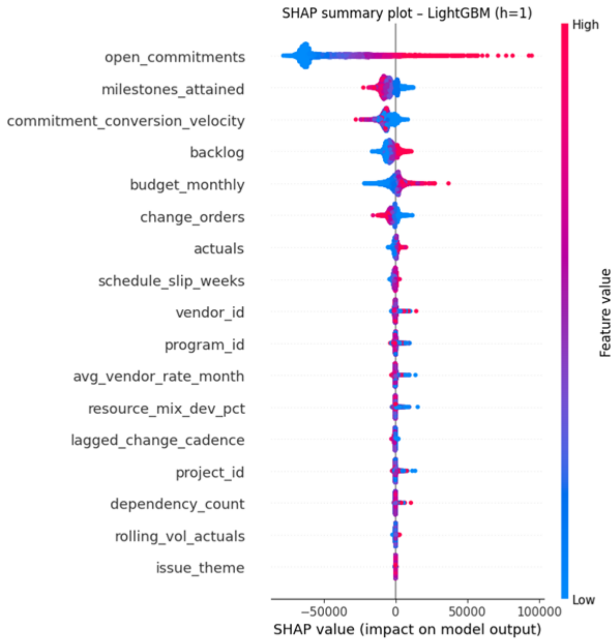


Figure 6. SHAP summary plot of top predictors

The variable milestones_attained exhibits a negative relationship with future expenditure, indicating that greater operational progress is associated with lower projected financial needs. This reflects the firm effect of project maturity, whereby completing milestones typically reduces uncertainty and financial risk. The feature commitment_conversion_velocity, which indicates how quickly commitments are converted into actual spending, also plays a significant explanatory role. A higher conversion velocity is associated with a lower remaining forecast, consistent with the burn rate concept: if spending is proceeding as planned, the future required funding will be lower. Both backlog and budget_monthly exhibit strong positive SHAP effects, particularly when backlog values are high. A large backlog is almost always a signal of outstanding work and unrealized obligations, which inevitably lead to increased future expenditure expectations. The change orders analysis is the main contributor to the further increase in the forecast. A high number of change orders is a sign of project scope or requirements volatility, thus leading to cost expansion as changes accumulate. Finally, schedule_slip_weeks has a positive effect on the predicted spend, indicating that project delays are one reason for higher future financial needs. The slippage typically entails inefficiencies, higher labour costs, and increased resource requirements, thereby increasing the forecasted expenditure.

Figure 7 shows the extent to which program identity moderates the effect of open commitments on the influence of open commitments. When open commitments exceed ~20k, programs with similar baseline SHAP patterns diverge significantly, indicating structural portfolio differences. It is worth noting that program_id itself is not the factor that drives the predictions; its effects arise only from interactions with financial variables, which provides evidence that the model is not based on arbitrary identifiers. Figure 8 depicts the local SHAP force plot for a typical project-month, illustrating how individual feature contributions changed the model's prediction from the global base value toward the final forecast. The figure shows that a high number of open commitments was the factor that increased projected expenditure most strongly, whereas milestones attained, zero backlog, and healthy commitment-conversion velocity together decreased the predicted remaining spend. This localized explanation confirms that the model behaves in accordance with financial logic at the case-by-case level, not only in global aggregate patterns.

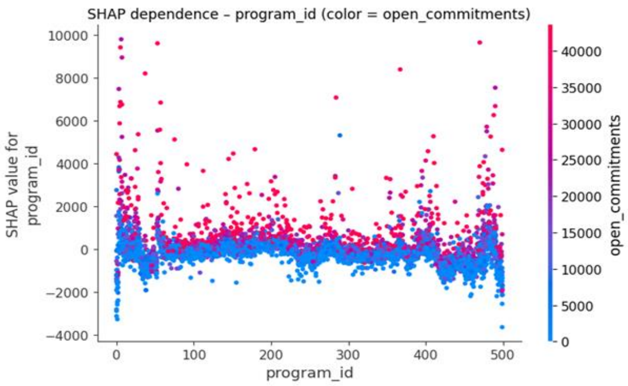


Figure 7. SHAP Dependence Plot (program_id vs SHAP, colored by open_commitments)

An ablation study with a focus on specific aspects also served to further confirm the SHAP-derived feature importance patterns. The omission of open_commitments from the features almost led to a doubling of MAE (from ≈12.7k to 24.0k) and a decrease of R^2 from 0.63 to 0.03, thereby confirming that it is the main factor that drives the remaining forecast. On the other hand, the removal of backlog, change orders, or schedule slip weeks led to only very small changes in accuracy, which is in line with their lower SHAP influence. Complete outcomes can be found in Table 5**.** Each row presents mean MAE, RMSE, and R^2 averaged over four time-based folds after the removal of a single feature group. An illustration shows the interaction of program identity with financial behavior, revealing that programs with higher open commitments consistently have larger positive SHAP contributions regardless of their program ID. The differences that appear within program_id clusters indicate that the structural changes in the portfolios are the root cause of these differences, for instance, the differences between vendor-heavy programs and milestone-driven initiatives. A clear funnel-shaped pattern reveals that program_id has almost no marginal influence when commitments are low but results in steep SHAP gradients as commitments increase. The main takeaway is that forecasting behaviour is not significantly influenced by the category of the program as represented by the identifiers but rather by the interactions between these identifiers and the key financial drivers, confirming that the model is focused on operational variables rather than arbitrary ID labels.

Figure 8. SHAP force plot for representative project-month prediction

Table 4. Scale-normalized forecasting metrics for XGBoost across horizons

Horizon	Mean Target (forecast_remaining)	Median Target	MAE	RMSE	R^2	MAPE (%)	SMAPE (%)	MASE
1-month	67,611.76	55,323.56	12,861.88	19,485.04	0.6269	2.13×10^{13}	105.17	0.8759
3-month	62,061.12	50,293.36	11,944.06	18,523.13	0.5670	2.44×10^{13}	117.33	0.8140
6-month	53,971.07	42,728.93	10,486.03	16,915.86	0.4291	3.00×10^{13}	134.25	0.7272

Table 5. Ablation of key financial drivers (XGBoost, h = 1 month)

Variant	Dropped Features	MAE	RMSE	R^2
Full model	None	12,740	19,468	0.628
No open_commitments	open_commitments	24,047	31,481	0.030
No backlog	backlog	12,787	19,624	0.622
No change_orders	change_orders	12,650	19,436	0.629
No schedule_slip_weeks	schedule_slip_weeks	12,654	19,443	0.629

In order to illustrate in a clear manner how SHAP-based local explanations facilitate governance-oriented "what-if" reasoning, we considered a representative project-month from the validation set with open_commitments = 52,400 and forecast_remaining = 71,900. The SHAP local explanation showed that open_commitments accounted for +18,750 of the predicted value. If we hypothetically reduce open_commitments by 10% (new value = 47,160) and recalculate the local SHAP explanation, this contribution decreases to +15,940, resulting in a net decrease of approximately 2,810 (≈ 15% relative reduction in the commitment-driven component and ≈ 3.9% reduction in total forecast_remaining). This case illustrates how the model can be used to enable practical financial scenario planning: small changes in exposure variables lead to quantifiable changes in projected spend, thus, controllers and portfolio managers gain the opportunity to evaluate mitigation strategies prior to their implementation.

4.4 Anomaly detection findings

Three unsupervised models (IForest, COPOD, LODA) were benchmarked on scaled numeric features, as shown in Table 6. All models flagged approximately 2% of records as anomalies, consistent with audit expectations. Figure 9 represents the variations in the anomaly scores detected by the COPOD (Copula-Based Outlier Detection) model and expressed for each project-month observation. Using COPOD, the highest scores are assigned to those observations, which are the extreme tails of the empirical copula distribution, i.e., they are statistically the most unusual combinations of financial and operational features. In fact, the right tail of the COPOD scores distribution is very long, with only a few project-months being marked as major outliers. The qualitative analysis of anomalies with the highest COPOD scores helped identify several distinct patterns, which were also recorded in the results log.

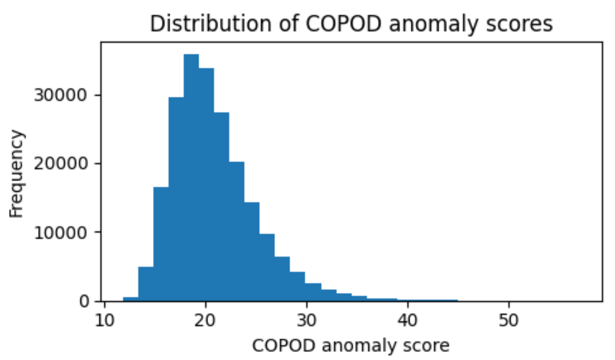


Figure 9. Distribution of COPOD anomaly scores

There were cases of projects with zero open commitments while there were extremely high actuals that most probably indicate data-entry inconsistencies or unplanned financial activities outside the standard practice. Besides that, the model found vendor rate spikes that were abnormally large (more than 200% of the normal monthly rates), thus it confirmed its capability to detect synthetic "vendor_rate_spike" anomalies.

Besides these, various other outliers existed, such as projects with extremely high backlog levels that did not correspond to milestone progress, thus these projects were most probably the ones which had stalled but still kept consuming the budget without delivering outputs. Another significant pattern was the sudden disappearance of forecasted expenditure in projects where forecast_remaining dropped to zero in just one month, thereby often indicating cancellations or major scope reductions. Altogether, these discoveries serve as evidence that unsupervised anomaly detection is capable of identifying quite subtle and non-obvious irregularities that are likely to be missed by traditional rule-based systems.

To put forward the operational value of the anomalies found, a detailed examination was done on several project-months with the highest scores. There is an example of a project that for two months in a row showed actuals of more than $420,000 while at the same time it had zero open commitments, thus it might be suggesting unplanned spending or incomplete procurement registration—an issue that usually results in an audit review. Another situation was about a vendor whose effective hourly rate went up by more than 260% just within one quarter, which is in line with the synthetic "vendor_rate_spike" anomalies that have been implanted in the dataset; such conduct would be the cause of contractual verification and invoice checking at the back. The third anomaly has to do with a project whose backlog was over $300,000 while there was no milestone progress for three months; thus, it was the stagnation of delivery despite ongoing financial consumption that was hinted at. These examples at the case level illustrate that the anomaly-detection layer does not merely pick up statistical outliers but rather brings to the surface financially interpretable risk scenarios that are in line with existing enterprise governance workflows. The contamination rate was set at 2% in line with the norms of audit practice, where true anomalous financial events make up a small but material part of portfolio activity. Since there were no labelled anomalies, contamination hyperparameter tuning by ROC/PR analysis was not possible.

Table 6. Design evaluation scores

Model	Contamination	Detected Anomalies	Key Insight
IForest	2%	4320	Detects extreme-value and volatility-driven irregularities
COPOD	2%	4320	Smooth probabilistic scoring; best for audit triage
LODA	2%	4320	Fastest; detects sparse high-dimensional anomalies

Because the data used are synthetic and lack ground-truth anomaly labels, it is not possible to compute precision–recall curves, ROC curves, or cost-weighted anomaly evaluation in a meaningful way. In accordance with unsupervised risk-screening standards, this study puts more emphasis on the qualitative inspection of cases with high anomaly scores. Label-based evaluation is a privilege for future deployments with domain-validated enterprise anomalies.

4.5 Key findings and trends

The findings show that gradient-boosting models outperform linear models, and nonlinear models have a mean absolute error (MAE) that is 25–40% lower than that of linear regression across all forecasting horizons. The gap in performance between these models is a strong indication that there are significant nonlinear interactions between backlog, commitments, vendor dynamics, and milestone progress—interactions that a linear model can hardly capture. The accuracy of the forecast deteriorates with increasing horizon length, consistent with the realities of enterprise financial planning. One-month forecasts were highly accurate, with an R^2 of approximately 0.63, whereas three-month forecasts remained operationally useful, with an R^2 of approximately 0.57. At the six-month horizon, the models still provided the correct direction but had lower precision, with an R^2 of approximately 0.43. The decrease is consistent with real-world sources of uncertainty, such as changing policies, staffing variability, vendor performance fluctuations, and scope drift. SHAP analysis provided operational insights that can be acted upon by identifying the features that have the strongest association with future financial exposure. Open commitments, backlog levels, and change orders are the dominant predictors that increase forecast_remaining, while milestones achieved, commitment conversion velocity, and lagged change-order cadence have a mitigating effect. These explanatory patterns agree with the logic of practical portfolio management, which deepens the trust that users place in the model's transparency and reliability. The anomaly detection findings led to the discovery of tightly-knit groups of irregularities that were related to vendor behaviour, delivery progress that has come to a halt, and extreme volatility. Along with COPOD, other detectors found instances that might indicate vendor rate manipulation, changes that have been backdated, duplicate invoice patterns, and zero-commitment over-expenditures—each of which is a risk scenario for an enterprise. The use of COPOD scoring in conjunction with SHAP-based explanations provides finance teams with solid audit trails, thus allowing them to make escalation decisions that are well-founded and supported by evidence. Lastly, the Optuna-tuned XGBoost model was able to demonstrate substantial improvements in accuracy, as evidenced by the reduction in MAE from 12,730 to 10,790, which corresponds to a 15% performance increase. This outcome highlights the importance of using Bayesian optimization to facilitate the performance of machine learning systems at an enterprise scale that are used for forecasting and shows that systematic hyperparameter tuning is an effective tool in operational analytics.

5. Discussion

This study aimed to develop and evaluate a single machine-learning framework that could, in principle, address simultaneously three fundamental problems of enterprise portfolio management: multi-horizon financial forecasting, unsupervised risk detection, and explainable reporting. The text below interprets the main empirical findings, links them to the existing literature, and also outlines the implications, limitations, and future research directions.

5.1 Interpretation of results

The forecasting comparison scenarios provide evidence that gradient-boosted tree models—XGBoost, LightGBM, and CatBoost—are quite consistently better than Linear Regression and Random Forest in all horizons. In fact, at 1-month horizon, XGBoost gets an R^2 close to 0.63 with MAE about 12.8k, and even though the explanatory power weakens at 3 and 6 months, where R^2 goes down to roughly 0.57 and 0.43, respectively, it is still valuable. The increase of uncertainty with the length of the prediction period is in line with the data, and enterprise portfolio dynamics are non-linear and interaction-heavy, as linear models cannot capture them properly [2,15]. Further, the Optuna-tuned XGBoost setup manages to cut down MAE at h=1 to about 10.9k, thus indicative of a major improvement over baseline untuned models. It emphasizes the necessity of a hyperparameter tuning strategy of forecasting models in real operational environments, where small improvements in the performance metrics can lead to big financial gains at the portfolio level [4].

According to SHAP-based explanation, the features open_commitments, backlog, change_orders, and budget_monthly contribute most positively to forecast_remaining whereas milestones_attained and commitment_conversion_velocity have a reducing effect. These concepts are very close to enterprise financial logic: the high level of unspent obligations and the volatile nature of scope changes increase the exposure; on the other hand, milestones and burn-down, being stable, lower it. The dependence plots reveal that program identifiers influence the intensity of these effects; however, they do not serve as independent factors, which means that the models are learning structural financial patterns rather than simply recalling the IDs [5].

The anomaly-detection component—isolated Forest, COPOD, and LODA with 2% contamination level—pinpoints a handful of project-months with the highest risk and most extreme behaviours: zero commitments and very high actuals, vendor rate spikes going beyond normal ranges, backlog that is both stalled and has low milestone progress, and sudden drops in forecast_remaining. The patterns in question associate well with hypothetical enterprise risk scenarios, which include unplanned spend, pricing irregularities, stalled delivery, or project cancellation, and this is the reason why unsupervised detectors have practical value as an audit triage mechanism [3, 11, 12].

5.2 Comparison with prior work

The higher performance of gradient-boosted trees compared to linear baselines is in line with previous studies in financial and investment forecasting, where ensemble models are, on average, superior to traditional econometric or linear approaches on tabular data [2, 15, 51]. The multi-horizon setup and the use of expanding-window temporal cross-validation for evaluation are in line with the recommendations of large-scale forecasting benchmarks such as the M4 Competition that advocate rigorous temporal splitting and leakage avoidance [4]. On the risk side, the implementation of Isolation Forest and similar detectors is very much in line with the references in anomaly-detection surveys and outlier-analysis frameworks that propose tree- and density-based methods for high-dimensional financial data as the most appropriate [11, 12]. The empirical finding that around 2% of records are identified as anomalous is in line with the expectations of enterprise audits and previous works on portfolio risk screening through predictive analytics [23-25]. Employing SHAP for global and local interpretability aligns with recent trends in enterprise AI, where explainable ML is increasingly considered a necessary condition for implementation in regulated financial contexts [5,30]. In the same way as interpretable frameworks suggested for supply-chain and credit risk modelling [30,38], this study shows that feature-attribution methods can provide explanations that are not only technically correct but also domain logic-consistent, thus enhancing trust and auditability.

5.3 Implications for enterprise practice

On a functional level, the integrated system features the missing link that is forecasting, anomaly detection, and reporting, often split into separate, loosely coupled tools in many ERP and financial environments, which have been fragmented [16,26,41]. The unification of these functionalities into a single pipeline enables consistent feature engineering, shared data quality controls, and traceable modelling choices across all analytics outputs. The multi-horizon forecasts support different levels of decision-making: 1-month forecasts can be used for tactical cash-flow and accrual planning, while 3- and 6-month forecasts can be used for budget reallocation, vendor negotiations, and portfolio reprioritisation. The finance teams, on the other hand, are not only able to anticipate what will happen if they combine anomaly flags with their work, but also can detect the concentration of risks, namely projects with high backlog and persistent schedule slips, enabling them to take the right actions [22, 25, 49].

The SHAP explanations and anomaly case summaries offer a clear account of the events to the stakeholders such as controllers, auditors, and risk committees, which they can rely on. The decision-makers are provided with forecasts that come along with the ranked drivers (e.g., open commitments and change orders) and the instances of unusual behaviour (e.g., vendor rate spikes) instead of the so-called "black-box scores," which can be used in the formal governance and escalation processes [17, 31,53]. In this way, the framework not only helps achieve higher predictive accuracy but also institutionalizes the learning process of how portfolio-level risks unfold over time [33, 40, 58].

Commercial solutions such as SAP Analytics Cloud, Oracle FCCS, Anaplan, and Microsoft Power BI, to varying extents, integrate forecasting with anomaly detection and are largely dependent on rule-based or proprietary black-box components. The framework presented here provides a comprehensive transparency and model logic, a unified forecasting–anomaly–explainability pipeline, and the capability to extend to custom portfolio behaviours that commercial tools cannot easily accommodate. Hence, the framework works with rather than against enterprise BI platforms, enabling greater analytical control and auditable reasoning. Table 7 presents a qualitative comparison of the proposed framework against leading enterprise analytics platforms across forecasting capability, anomaly detection, interpretability, customizability, and governance support.

5.4 Limitations

To begin with, all data are synthetic but retain a realistic structure; hence, while this setup enables controlled experiments, real ERP datasets can be noisier, less structured, and influenced by policies that may differ from the patterns revealed here. Therefore, the study's external validity must be verified in real-world enterprise environments [13,32]. Secondly, the presented framework is limited to batched historical data, with retraining performed periodically. It currently does not support real-time streaming data ingestion or online learning. However, the previous work has already established the necessity of real-time financial modelling and streaming architectures in dynamic risk scenarios [13,62]. Thirdly, the anomaly-detection evaluation being performed here is qualitative and not label-based. In the absence of ground-truth anomaly labels or expert validation logs, one cannot calculate precision, recall, or cost-weighted performance. This limits the ability to measure trade-offs between undetected anomalies and false alerts [11, 24].

Lastly, the model zoo is intentionally limited to tree ensembles and classical detectors. While this might be suitable for most ERP settings, it leaves out the latest developments in deep-sequence modelling, hybrid reinforcement learning, and knowledge-graph–augmented financial risk optimisation that can potentially enhance the performance in the highly complex scenarios [19,20, 54]. The framework does not explicitly represent drifting structural changes in the scenarios, for example, changes in procurement policies, macroeconomic shocks, or inflation-induced cost increases.

Table 7. Feature comparison of enterprise forecasting & risk systems

System	Forecasting	Anomaly Detection	XAI	Customizability	Governance / Audit
SAP Analytics Cloud (SAC)	Basic statistical forecasting; limited ML control	Rule-based alerts only	Minimal; no SHAP	Moderate (scripts only)	Strong logging; limited model transparency
Oracle FCCS	Regression-based predictive planning	Mostly thresholds/rules	None	Low; closed models	High audit trail; opaque ML
Anaplan	Proprietary time-series forecasting; strong scenarios	No unsupervised detection	None (rule explanations only)	High in formulas; low in ML	Good governance; formula transparency
Power BI + Azure Anomaly Detector	Exponential smoothing; optional Azure ML	Unsupervized detector with limited tuning	Very limited unless custom SHAP added	High only with custom pipelines	Good audit logging via Azure Monitor
This Unified ML Framework	Multi-horizon ML (XGB, LGBM, CatBoost); Optuna tuned	IForest, COPOD, LODA	Full SHAP global & local XAI	Very high; open-source, flexible	Strong auditability; reproducible, transparent

In fact, drift-monitoring tools should be in place to detect such changes promptly, thereby enabling recalibration or retraining cycles to maintain stability and governance. There was no formal statistical significance testing or confidence interval estimation for forecasting metrics, as the main focus was on comparative model behaviour across matched temporal folds. Subsequent research will add bootstrap confidence intervals and model-comparison tests to support inferential claims. This study version lacks a complete ablation table for all the features that have been engineered. As the dataset is synthetic and some interactions are structurally embedded, a systematic ablation analysis may not generalise significantly; however, future research will examine controlled ablations of real ERP datasets.

5.5 Future Directions

Future projects can broaden this framework in different ways. At the model layer, incorporating transformer-based tabular models, sequence models, or hybrid RL architectures may improve long-horizon forecasts and capture more complex temporal dependencies in vendor and project behaviour [19,20,54]. At the systems level, the next crucial step is to integrate the pipeline into a complete MLOps stack—continuous integration, automated drift detection, scheduled retraining, and shadow deployment in live ERP systems—to assess robustness under real production workloads [26,35,41]. Methodologically, subsequent studies must rely on expert-labelled anomalies and conduct prospective validations to measure operational value (e.g., avoided overruns or earlier detection of problematic projects). Scenario modelling could be incorporated into SHAP explanations to enable "what-if" analyses—e.g., illustrating how a 10% reduction in open commitments would affect forecast_remaining at the portfolio level [33,36].

Moreover, the ethical and governance aspects warrant further investigation. Among these are fairness analyses across departments or vendor groups, regulated human-in-the-loop interventions for overriding anomaly flags, and structured estimates of financial impact (e.g., savings from a 2–5% reduction in overruns in large portfolios) [31,42,53]. By resolving these issues, the adoption of integrated ML pipelines as central instruments of enterprise financial governance rather than mere experimental add-ons would be more convincing. As per the evidence in Table 8, the next steps of highest priority would be online-learning pipelines, real-world 12-month validation, and the integration of labelled anomalies for supervised risk modelling. from a practical point of view, very small gains in accuracy could lead to considerable financial advantages. Thus, a 2-5% decrease in forecast error or better use of the budget without overruns that can be achieved through earlier anomaly detection and improved spend visibility in a portfolio with an annual budget of \$500 million would mean that avoided losses would be around \$10-25 million per year.

A well-defined research plan to move the framework forward comprises the following points: (i) integration of online learning using libraries such as River, (ii) benchmarking of CatBoost and transformer-based models on multinational ERP datasets, (iii) creation of scenario-based "what-if'' explanations, (iv) prospective 12-month validation in a live ERP environment, (v) expert-labeled anomaly integration for supervised risk modelling, and (vi) fairness and governance audit formalization. These measures would gradually bring the framework closer to enterprise-grade operational deployment.

Table 8. Prioritized research agenda for advancing the unified ML framework

Priority	Research Direction	Estimated Timeline	Description / Expected Contribution
High	1. Integration of online learning pipelines (e.g., River)	6–12 months	Enables continuous adaptation to new financial behaviour, reduces drift, and supports near–real-time portfolio monitoring.
High	2. Prospective 12-month validation in live ERP environments	12–18 months	Tests robustness under real operational conditions, evaluates downstream impact on budgeting accuracy, and quantifies financial savings.
High	3. Incorporation of expert-labelled anomalies for supervised anomaly detection	6–12 months	Allows computation of precision, recall, cost-weighted scores, and reduces false positives in vendor/payment workflows.
Medium	4. Evaluation of transformer-based tabular models (FT-Transformer, TabNet)	6–12 months	Assesses whether deep architectures outperform gradient-boosted trees for long-horizon, nonlinear enterprise financial forecasting.
Medium	5. Development of what-if scenario engine for interactive SHAP-based counterfactuals	3–9 months	Provides controllers and auditors with actionable levers (e.g., "reduce commitments by 10% ⇒ forecast decreases by X").
Medium	6. Formal fairness and governance audits	6–12 months	Measures cross-department and cross-vendor bias, supports compliance with internal audit and regulatory standards.
Medium	7. Full MLOps deployment (MLflow/Kubeflow + CI/CD + drift monitoring)	9–18 months	Ensures scalable, reproducible, and traceable model deployment across enterprise environments.
Low	8. Expansion to multinational ERP datasets	12–24 months	Tests generalisability across regulatory regimes, currencies, vendor structures, and project governance models.
Low	9. Integration of hybrid RL + knowledge-graph methods	18–30 months	Supports adaptive portfolio optimisation and long-horizon strategic planning beyond static forecasting.
Low	10. Estimation of financial ROI from improved forecasting and anomaly detection	3–6 months	Quantifies monetary benefits (e.g., 2–5% reduction in overruns = \$10–25M annual savings in large portfolios).

6. Conclusion

This research designed and assessed a combined machine-learning architecture that integrates multi-horizon portfolio forecasting, unsupervised anomaly detection, and SHAP-based interpretability in a single enterprise-ready pipeline. Using a large, structurally realistic project-month dataset, gradient-boosted tree models (in particular XGBoost, LightGBM, and CatBoost) consistently outperformed linear baselines, achieving strong predictive accuracy at the 1-month horizon and retaining useful signal at 3 and 6 months. SHAP analysis indicated that open commitments, backlog, change orders, and schedule slippage are major drivers of future financial exposure, whereas milestone progress and healthy commitment conversion are stabilising factors. The anomaly detectors (Isolation Forest, COPOD, LODA) brought out small, high-risk subsets of project-months with plausible enterprise risk patterns, for example, vendor rate spikes, zero-commitment overspend, and stalled delivery. The implications for practice are of two kinds. Firstly, a single architecture lessens the disintegration that is usually between forecasting, risk monitoring, and reporting tools, thus allowing for consistent feature engineering, governance, and auditability. Secondly, the combination of high-performing models with clear explanations changes ML output from being just a set of difficult-to-understand scores into defensible financial narratives that can be easily understood by controllers, auditors, and portfolio leaders. Therefore, the framework not only works as a prediction engine but also as a means of institutional learning about the financial risk that is accumulated across portfolios. According to these results, practitioners should: (i) focus on gradient-boosted ensembles with systematic hyperparameter tuning for portfolio forecasting, (ii) combine forecasting with anomaly detection rather than working them as separate processes, and (iii) place SHAP-style explanations directly in the dashboards and review processes so that every high-risk signal has an interpretable explanation. Subsequent research should confirm the framework using genuine ERP data, develop it for streaming and online-learning scenarios, and include expert-labelled anomalies to measure precision, recall, and financial impact. Work is also required on fairness, human-in-the-loop governance, and “what-if” scenario analytics, which would make unified ML pipelines the main components of enterprise financial stewardship rather than isolated analytical experiments.

Ethical issue

The authors are aware of and comply with best practices in publication ethics, specifically regarding authorship (avoidance of guest authorship), dual submission, manipulation of figures, competing interests, and compliance with research ethics policies. The authors adhere to publication requirements that the submitted work is original and has not been published elsewhere.

Data availability statement

The manuscript contains all the data. However, more data will be available upon request from the authors.

Conflict of interest

The authors declare no potential conflict of interest.

References

[1] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic Minority Over-sampling Technique,” J. Artif. Intell. Res., vol. 16, pp. 321–357, 2002, doi: 10.1613/jair.953.

[2] T. Chen and C. Guestrin, “XGBoost: A Scalable Tree Boosting System,” in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, San Francisco, CA, USA, 2016, pp. 785–794, doi: 10.1145/2939672.2939785.

[3] F. T. Liu, K. M. Ting, and Z.-H. Zhou, “Isolation Forest,” 2008 Eighth IEEE International Conference on Data Mining, Pisa, Italy, 2008, pp. 413–422, doi: 10.1109/ICDM.2008.17.

[4] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, “The M4 Competition: Results, Findings, and Conclusions,” Int. J. Forecast., vol. 34, no. 4, pp. 802–808, 2018, doi: 10.1016/j.ijforecast.2018.06.001.

[5] S. M. Lundberg and S.-I. Lee, “A Unified Approach to Interpreting Model Predictions,” in Proc. 31st Conf. Neural Inf. Process. Syst. (NeurIPS), Long Beach, CA, USA, 2017. (no DOI — see arXiv:1705.07874)

[6] H. He and E. A. Garcia, “Learning from Imbalanced Data,” IEEE Trans. Knowl. Data Eng., vol. 21, no. 9, pp. 1263–1284, 2009, doi: 10.1109/TKDE.2008.239.

[7] M. Buda, A. Maki, and M. A. Mazurowski, “A Systematic Study of the Class Imbalance Problem in Convolutional Neural Networks,” Neural Netw., vol. 106, pp. 249–259, 2018, doi: 10.1016/j.neunet.2018.07.011.

[8] C. Cortes and V. Vapnik, “Support-Vector Networks,” Mach. Learn., vol. 20, no. 3, pp. 273–297, 1995, doi: 10.1007/BF00994018.

[9] H. Zou and T. Hastie, “Regularization and Variable Selection via the Elastic Net,” J. Roy. Stat. Soc. B, vol. 67, no. 2, pp. 301–320, 2005, doi: 10.1111/j.1467-9868.2005.00503.x.

[10] J. Platt, “Probabilistic Outputs for Support Vector Machines,” in Advances in Large Margin Classifiers, MIT Press, 1999, pp. 61–74.

[11] V. Chandola, A. Banerjee, and V. Kumar, “Anomaly Detection: A Survey,” ACM Comput. Surv., vol. 41, no. 3, pp. 1–58, 2009, doi: 10.1145/1541880.1541882.

[12] C. C. Aggarwal, Outlier Analysis, 2nd ed., Springer, 2017. https://doi.org/10.1007/978-3-319-47578-3

[13] A.Odunaike, “Integrating real-time financial data streams to enhance dynamic risk modeling and portfolio decision accuracy,” Int. J. Comput. Appl. Technol. Res., vol. 14, no. 8, pp. 1–16, 2025, doi: 10.7753/IJCATR1408.1001.

[14] V. Kalvala and A. Gupta, “Integrating Machine Learning and Statistical Models in Enterprise Risk Analysis,” 2025 4th International Conference on Sentiment Analysis and Deep Learning (ICSADL), Bhimdatta, Nepal, 2025, pp. 852–861, doi: 10.1109/ICSADL65848.2025.10933366.

[15] K. U. Apu, M. M. Rahman, A. B. Hoque, and M. Bhuiyan, “Forecasting future investment value with machine learning, neural networks, and ensemble learning,” Rev. Appl. Sci. Technol., vol. 1, no. 2, pp. 1–25, 2022, doi: 10.63125/edxgjg56.

[16] P. S. R. P. Muntala and S. K. Jangam, "Automated Risk Scoring in Oracle Fusion ERP Using Machine Learning," Int. J. Artif. Intell. Data Sci. Mach. Learn., vol. 5, no. 4, pp. 105–116, 2024, doi: 10.63282/3050-9262.IJAIDSML-V5I4P111.

[17] Q. Xin, "Construction of a machine-learning-based risk management evaluation model for enterprise financial reporting," Proc. Int. Conf. Commun., Inf., Digit. Technol. (CIDT2024), vol. 13185, pp. 19–26, 2024, doi: 10.1117/12.3032790.

[18] C. P. Vijay, "A Deep Learning Approach for Financial Risk Prediction in Enterprise Management Systems," Eksplorium, vol. 46, no. 1, pp. 1579–1597, 2025.

[19] X. Shi, Y. Zhang, M. Yu, and L. Zhang, "Deep learning for enhanced risk management: A novel approach to analyzing financial reports," PeerJ Comput. Sci., vol. 11, p. e2661, doi: 10.7717/peerj-cs.2661.

[20] Y. Cui and F. Yao, "Integrating deep learning and reinforcement learning for enhanced financial risk forecasting in supply chain management," J. Knowl. Econ., vol. 15, no. 4, pp. 20091–20110, 2024, doi: 10.1007/s13132-024-01946-5.

[21] C. Oko-Odion, A. Okunuga, and O. I. Okunbor, "Revolutionizing financial risk assessment through deep learning-driven business analytics," 2025, doi: 10.30574/wjarr.2025.25.1.0327.

[22] L. Hamzat, "Holistic enterprise risk and cost governance through real-time financial monitoring and predictive intelligence," 2025, doi: 10.5281/zenodo.15377535.

[23] O. O. Fagbore et al., "Predictive Analytics for Portfolio Risk Using Historical Fund Data," 2022, doi: 10.54660/.JFMR.2022.3.1.223-240.

[24] A.O. Ogedengbe et al., "A Predictive Compliance Analytics Framework Using AI," Shodhshauryam Int. Sci. Res. J., vol. 6, no. 4, pp. 171–195, 2023, doi: 10.32628/SHISRRJ.

[25] X. Wang, Z. Xue and Y. Chen, "Financial Risk Warning and Traceability for Manufacturing Enterprises: A Three-Stage Framework Based on Ensemble Learning Algorithm," IEEE Trans. Comput. Soc. Syst., vol. 12, no. 6, pp. 5061–5079, Dec. 2025, doi: 10.1109/TCSS.2025.3576991.

[26] S. Thambireddy et al., "AI-Enabled SAP Enterprise Systems: A Comprehensive Business Use Case Survey," Proc. ICSCSA, pp. 1045–1052, Aug. 2025, doi: 10.1109/ICSCSA66339.2025.11170704.

[27] N. Rane, S. Choudhary, and J. Rane, "Artificial intelligence and machine learning in business intelligence, finance, and e-commerce," 2024, doi: 10.2139/ssrn.4843988.

[28] K. Abiodun et al., "Risk-Sensitive Financial Dashboards with Embedded ML," Int. J. Sci. Res. Mod. Technol., vol. 3, no. 2, pp. 1–18, 2024, doi: 10.38124/ijsrmt.v3i2.678.

[29] N. L. Rane et al., "AI, ML, and DL for advanced business strategies," Partners Univ. Innov. J., vol. 2, no. 3, pp. 147–171, 2024, doi: 10.5281/zenodo.12208298.

[30] K. R. Ahmed et al., "Deep learning framework for interpretable supply chain forecasting," Sci. Rep., vol. 15, p. 26355, 2025, doi: 10.1038/s41598-025-11510-z.

[31] O. O. Elumilade et al., "The role of data analytics in strengthening financial risk assessment," Iconic Res. Eng. J., vol. 6, no. 10, pp. 324–338, 2023, doi: 10.64388/IREV6I10-1704359.

[32] J. G. George, "Enterprise architecture for post-merger financial systems integration," Aust. J. ML Res. App., vol. 3, no. 2, p. 429, 2023.

[33] G. Stephen, "Leveraging AI for Strategic Decision-Making in Biopharmaceutical Program Management," Int. J. Manag. Technol., vol. 12, no. 4, pp. 1–26, 2025, doi: 10.37745/ijmt.2013/vol12n4126.

[34] Oyeyipo et al., "A conceptual framework for transforming corporate finance through strategic growth," Int. J. Adv. Multidiscip. Res. Stud., vol. 3, no. 5, pp. 1527–1538, 2023.

[35] G. Tripathi, "AI in Finance Institutions: Multiplying Output using SageMaker," J. Comput. Sci. Technol. Stud., vol. 7, no. 5, pp. 94–101, 2025, doi: 10.32996/jcsts.2025.7.5.13.

[36] Nwoke, "Harnessing predictive analytics, machine learning, and scenario modeling," Int. J. Comput. Appl. Technol. Res., 2025, doi: 10.7753/IJCATR1404.1010.

[37] Rane, Jayesh and Amol Chaudhari, Reshma and Rane, Nitin, Artificial Intelligence and Machine Learning for Supply Chain Resilience: Risk Assessment and Decision Making in Manufacturing Industry 4.0 and 5.0 (July 24, 2025). Available at SSRN: https://ssrn.com/abstract=5366932 or http://dx.doi.org/10.2139/ssrn.5366932

[38] X. Han et al., "Symmetry-Aware Credit Risk Modeling," Symmetry, vol. 17, no. 3, 2025. DOI: 10.3390/sym17030341

[39] O. Irekponor, "Designing resilient AI architectures for predictive energy finance systems," Int. J. Res. Publ. Rev., vol. 6, no. 6, pp. 73–100, 2025.

[40] N. A. Siddiqui, "Optimizing Business Decision-Making through AI-Enhanced BI Systems," Strategic Data Manag. Innov., vol. 2, no. 1, pp. 202–223, 2025, doi: 10.71292/sdmi.v2i01.21.

[41] Rane, Nitin and Paramesha, Mallikarjuna and Choudhary, Saurabh and Rane, Jayesh, Business Intelligence and Business Analytics With Artificial Intelligence and Machine Learning: Trends, Techniques, and Opportunities (May 17, 2024). Available at SSRN: https://ssrn.com/abstract=4831920 or http://dx.doi.org/10.2139/ssrn.4831920.

[42] O. Famoti et al., "Data-Driven Risk Management in US Financial Institutions," 2023. https://www.multiresearchjournal.com/admin/uploads/archives/archive-1739352974.pdf

[43] A.Ajuwon et al., "AI-Powered Transformations in Financial Services," Eng. Technol. J., vol. 10, no. 6, pp. 5397–5415, 2025, doi: 10.47191/etj/v10i06.12.

[44] J. Rane, R. A. Chaudhari, and N. L. Rane, "Sustainable Supply Chain Resilience through AI," Deep Science Publishing, 2025.ISBN:9789371856218, 9371856211

[45] A.Ahmed et al., "AI-driven innovations in modern banking," J. Manag. Sci. Res. Rev., vol. 4, no. 3, pp. 1145–1183, 2025.
[46] Saiyed, "AI-Driven Innovations in Fintech," Int. J. Elect. Comput. Eng. Res., vol. 5, no. 1, pp. 8–15, 2025.
[47] Adekunle et al., "Integrating AI-driven risk assessment frameworks in financial operations," Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol., vol. 9, no. 6, pp. 445–464, 2023, doi: 10.32628/IJSRCSEIT.
[48] A.Khaddam and A. Alzghoul, "AI-driven business intelligence for strategic energy and ESG management," Int. J. Energy Econ. Policy, vol. 15, no. 4, p. 635, 2025, doi: 10.32479/ijeep.19820.
[49] A.Ojo, "A Data-Driven Framework for Project Risk Monitoring Using Predictive Analytics," J. Manag. Dev. Res., vol. 2, no. 2, pp. 125–136, 2025, doi: 10.69739/jmdr.v2i2.1171.
[50] S. A. Mohammed, "Integrated Financial Ecosystems," Indonesian J. Adv. Res. Technol., vol. 7, no. 7, 2025.
[51] O. Olubusola et al., "Machine learning in financial forecasting: A U.S. review," World J. Adv. Res. Rev., vol. 21, no. 2, pp. 1969–1984, 2024, doi: 10.30574/wjarr.2024.21.2.0444.
[52] R. Paul, M. A. Rahman, and M. Nuruzzaman, "AI-Enabled Decision Support Systems for Infrastructure Project Management," Rev. Appl. Sci. Technol., vol. 3, no. 4, pp. 29–47, 2024, doi: 10.63125/8d96m319.
[53] Oko-Odion and A. Angela, "Risk management frameworks for financial institutions," Int. J. Sci. Res. Arch., vol. 14, no. 1, pp. 1182–1204, 2025, doi: 10.30574/ijsra.2025.14.1.0155.
[54] S. Uddin et al., "Hybrid reinforcement learning and knowledge graph framework for financial risk optimization," Sci. Rep., vol. 15, no. 1, p. 29057, 2025, doi: 10.1038/s41598-025-14355-8.
[55] A.S. Adepoju, "Transforming administrative functions through AI," Int. J. Res. Publ. Rev., vol. 6, no. 1, pp. 4300–4316, 2025, doi: 10.55248/gengpi.6.0125.0622.
[56] Hasan and M. O. Faruq, "AI-Augmented Risk Detection in Cybersecurity Compliance," ASRC Procedia, vol. 1, no. 1, pp. 313–342, 2025, doi: 10.63125/49gs6175.
[57] Batz et al., "Integrating machine learning into business and management," Humanit. Soc. Sci. Commun., vol. 12, no. 1, pp. 1–20, 2025, doi: 10.1057/s41599-025-04361-6.
[58] Y. S. Ahirrao et al., "AI-Powered Financial Strategy," Emerg. Front. Libr. Amer. J. Eng. Technol., vol. 7, no. 9, pp. 126–151, 2025.
[59] U. Das, "Scaling Agile with AI," IJSAT, vol. 16, no. 2, 2025. https://doi.org/10.71097/IJSAT.v16.i2.5874
[60] G. Luo, M. A. Arshad, and G. Luo, "Decision Trees for Strategic Choice," Symmetry, vol. 17, no. 7, p. 976, 2025, doi: 10.3390/sym17070976.
[61] S. S. Parimi, "Automated Risk Assessment in SAP Financial Modules through Machine Learning," SSRN Electron. J., 2019, doi: 10.2139/ssrn.4934897.
[62] A.Odunaike, "Integrating Real-Time Financial Data Streams…," Int. J. Comput. Appl. Technol. Res., vol. 14, no. 8, pp. 1–16, 2025, doi: 10.7753/IJCATR1408.1001.
[63] Mendeley Data, "Dataset: w9mzf4vswh (Version 2)," Mendeley Data, Accessed: Feb. 10, 2025. [Online]. Available: https://data.mendeley.com/datasets/w9mzf4vswh/2
[64] U.S. General Services Administration, "Federal Procurement Data System (FPDS)," FPDS.gov, Accessed: Feb. 10, 2025. [Online]. Available: https://www.fpds.gov/
[65] California State Government, "Open Fiscal Data Portal," Open Fiscal California, Accessed: Feb. 10, 2025. [Online]. Available: https://open.fiscal.ca.gov/
[66] Stanford University, "MCC: The Modeling Civil Courts Project," Stanford Law School, Accessed: Feb. 10, 2025. [Online]. Available: https://mcc.law.stanford.edu/
[67] Government Transparency Project, "Databases Overview," GovTransparency.eu, Accessed: Feb. 10, 2025. [Online]. Available: https://www.govtransparency.eu/category/databases/