Computer Science and Information Technologies
Vol. 7, No. 3, November 2026, pp. 337~345
ISSN: 2722-3221, DOI: 10.11591/csit.v7i3.p337-345

Risk-integrated contractor allocation in Zimbabwe's timber value chain

Tavengwa Norman, Brian Mupini
School of Information Science and Technology, Harare Institute of Technology, Harare, Zimbabwe

Article Info

Article history:

Received May 19, 2026
Revised Jun 10, 2026
Accepted Jul 13, 2026

Keywords:

Contractor risk
Decision support systems
Forestry supply chain
Multi-objective optimization
NSGA-II
XGBoost

ABSTRACT

In Zimbabwe, the commercial forestry industry is reliant on a large proportion of outsourced harvest and milling contractors whereas existing enterprise resource planning (ERP) systems record completed transactions instead of predicting contractor failure before assigning activities. We introduce the dynamic resource allocation framework (DRAF), a risk-integrated decision-support framework that supplements the probability of contractor failure derived from XGBoost to a non-dominated sorting genetic algorithm II (NSGA-II) multi-objective optimizer. It divides contractor-block-mill combinations, minimizes cost and expected delay, and maximizes risk-adjusted timber recovery and operational reliability. We created the solution on an 828,789-record virtual ERP dataset and calibrated it to Manicaland forestry conditions and tested with statistical, heuristic and risk-free optimization baselines. Extreme gradient boosting (XGBoost) obtained a holdout receiver operating characteristic area under the curve (ROC-AUC) of 0.965, recall of 0.999, and F1 of 0.867, showing an improvement of 0.365 over logistic regression. The optimizer developed 64 complete Pareto solutions to balanced and high-recovery scenarios and uncovered a constraint-feasibility boundary for a more conservative low-risk scenario. The satisfaction score of the 30-practitioner stakeholder assessment was 4.19 out of 5.0. The results demonstrate that embedding predictive risk into an optimization objective can optimize forestry allocation decisions and suggest that some real ERP validation is required before such measures will be broadly implemented.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Tavengwa Norman
School of Information Science and Technology, Harare Institute of Technology
Ganges Road, Belvedere, Harare, Zimbabwe
Email: h240742c@hit.ac.zw

1. INTRODUCTION

The commercial forestry industry in Zimbabwe is dominated by vertically integrated firms like Allied Timbers Zimbabwe, Border Timbers Limited, The Wattle Company, and Mutare Board and Paper Mills, which collectively manage over 70,000 ha of plantation in Manicaland [1], [2]. The importance of forestry to rural employment, the supply of industrial raw materials, and export-oriented value chains endures [3]. But outsourcing has transferred operational risk into contractor networks: up to 70% of milling production is now generated by small- to medium-sized contractors [4], requiring contractor allocation to be a high-stakes decision that must balance cost, capacity, terrain, rainfall, and reliability before work begins [5], [6].

Prior literature shows two relevant but usually separate capabilities: ensemble machine learning (ML) can predict contractor or supplier underperformance in structured operational environments [7]–[9], and evolutionary multi-objective optimization (MOO), especially non-dominated sorting genetic algorithm II

(NSGA-II), can produce Pareto-efficient allocation strategies in forestry and supply chain applications [10], [11]. Supply chain risk theory also treats disruption risk as a dynamic planning variable rather than a static after-the-fact explanation [12], [13]. Recent studies have shown that ML techniques can enhance forestry harvesting operations by supporting predictive analytics and operational decision-making [14]. However, existing research has primarily focused on specific operational tasks rather than integrating predictive risk assessment, MOO, and interactive decision support into a unified decision-intelligence framework tailored for forestry applications in Southern Africa. However, these capabilities have not been merged into a forestry-adapted decision-intelligence artefact for Southern Africa.

This paper fills that gap by developing the dynamic resource allocation framework (DRAF), with three main components: i) an extreme gradient boosting (XGBoost) ensemble classifier that translates pre-allocation operational characteristics into calibrated contractor risk probabilities; ii) an NSGA-II optimizer which embeds these probabilities as first-class objective inputs rather than as post-hoc constraints; and iii) an interactive decision-support dashboard built in Dash that delivers Pareto-optimal allocation guidelines to non-technical estate managers. The research follows the design science research (DSR) methodology [15], [16], in which the framework is both the theoretical contribution and the main research output.

The specific contributions are threefold. First, the study proposes a risk-integrated optimization architecture in which ML-derived risk probabilities are directly embedded in evolutionary objective functions. Second, it constructs and validates a reproducible 828,789-record virtual enterprise resource planning (ERP) dataset calibrated to Zimbabwean forestry conditions. Third, it evaluates DRAF against statistical, heuristic, and risk-free optimization baselines, and against feedback from 30 forestry-sector practitioners. The rest of this paper is formatted as: section 2 presents the design justification of the proposed framework. Section 3 describes the method (framework architecture, problem formulation, dataset construction, and experimental setup). Results and discussion are reported in section 4 and section 5 concludes.

2. PROPOSED METHOD DESIGN RATIONALE

2.1. Predictive risk rationale

In challenging working environments, ensemble ML has emerged as an important technique for predicting contractor and supplier underperformance. Because they capture non-linear interactions without the representation-learning overhead caused by deep neural models, random forests [17] and XGBoost [18] are important baselines for tabular risk modelling. Research studies, mainly with respect to contractor-risk and supply-chain and hybrid XGBoost-logistic approaches, have revealed high receiver operating characteristic area under the curve (ROC-AUC) performance [7], [8], [19], while broader tabular-learning comparisons have nevertheless demonstrated the ability for tree-based models to achieve superior results over deep architectures on structured data [9], [20].

2.2. Optimization rationale

MOO searches for non-dominated alternatives when cost, risk, service level, and resource constraints cannot be collapsed into one score without losing managerial meaning [21]. Recent evolutionary-computation studies also highlight multi-objective evolutionary algorithm based on decomposition (MOEA/D), strength Pareto evolutionary algorithm (SPEA)2, and hybrid designs for supply chain optimization [22]. Within this family, NSGA-II [10] remains widely validated for multi-objective resource-allocation problems, while forestry and construction studies show that optimization can improve planning quality under productivity, cost, and resource constraints [11], [23], [24].

The architectural gap is that most forestry and supply-chain MOO implementations assume deterministic contractor performance. Risk is omitted, treated as a feasibility filter, or reviewed after optimization. DRAF addresses this gap by inserting probabilistic ML risk scores into the optimizer's objective functions as dynamic weights, linking predictive modelling, optimization, and decision support in the same allocation workflow [25]–[29].

3. METHOD

3.1. Conceptual architecture

DRAF is operationalised as a three-layer decision-intelligence pipeline. The input layer consumes pre-allocation decision-time variables across the harvesting, transportation, and milling domains, which are defined at the time of contractor selection and are not leaky after the contract has been finalized. The predictive risk layer implements a trained XGBoost ensemble classifier, generating calibrated probability estimates p (underperformance | features) which are integrated directly into the objective functions of the MOO layer. The decision support layer provides the Pareto-optimal allocation sets in an interactive dashboard for estate managers to assess trade-offs and update operational decisions in real time. In practical terms, this layer is analogous to a lightweight distributed decision-support architecture: model scoring, evolutionary optimization, and dashboard interaction can be decoupled into services that exchange standardized ERP data objects and can therefore be deployed on local servers or cloud infrastructure as estate data maturity improves [28], [29]. The architectural innovation is the direct connection between predictive and optimization layers. Rather than using risk as a constraint threshold, the traditional practice in forestry MOO, predicted risk is multiplied as a discount on expected timber recovery under the optimizer objective function in DRAF. This makes sure the evolutionary search actively favors allocations in which higher expected yield implies a lower probability of contractor failure, meaning more robust, rather than simply mathematically optimal, solutions.

3.2. Predictive risk model formulation

Let each candidate allocation be defined by the tuple (b, c, r, m), where b denotes a timber block, c a contractor, r a haulage route, and m a mill. Let x=(x1, x2, …, x46) denote the 46-dimensional leakage-safe feature vector characterizing the allocation. The predictive model estimates:

$$\rho(b, c, r, m) = P(late\_delivery\_risk = 1 \mid x) \in [0, 1]$$

where ρ is the XGBoost-estimated risk probability. XGBoost was chosen over random forest and logistic baselines due to better validation precision-recall area under the curve (PR-AUC). Synthetic minority over-sampling technique (SMOTE) oversampling on training pipeline helped to mitigate moderate class imbalance (positive class rate: 30.93%). τ=0.24 is the decision threshold optimized during validation split to maximize the F1-score while preserving high recall, which is indicative of the asymmetric misclassification cost structure that indicates that false negatives (FN) (undetected high-risk allocations) are operationally more costly than false positives (FP).

3.3. Multi-objective optimization formulation

Let A={a1, a2, …, an} denote the feasibility-filtered candidate assignment pool. Each candidate a1 is characterized by deterministic attributes (cost Ci, base volume Vi, base duration Di, route fit Ri, and mill feasibility Mi) and the ML-derived risk probability ρi. DRAF simultaneously minimizes/maximizes four objectives:

- $f1 = min\ \Sigma\ Ci$ i) total operational cost
- $f2 = max\ \Sigma\ Vi\ (1 - \rho i)$ ii) risk-adjusted timber recovery
- $f3 = min\ \Sigma\ Di\ \cdot\ \rho i$ iii) expected delay
- $f4 = max\ \Sigma\ (1 - \rho i)\ \cdot\ Ri\ \cdot\ Mi$ iv) operational reliability

Subject to: contractor capacity limits, mill throughput constraints, terrain slope feasibility (slope_fit ≥θs), route accessibility (route_fit≥θr), and average risk ceiling (mean ρ≤$\bar{\theta}$ρ). F2 is the main innovation: the (1−ρi) discount coefficient continuously penalizes solutions that assign high-yield blocks to high-risk contractors, driving the evolutionary search toward allocations that are both economically productive and operationally resilient. NSGA-II was implemented using Pymoo (version 0.6.x) with simulated binary crossover (SBX), polynomial mutation, and rounding repair operators. Three policy scenarios, balanced, low-risk, and high-recovery, were configured with distinct constraint parameters and shortlist weights.

3.4. Virtual ERP dataset construction

Because Zimbabwean forestry ERP records were unavailable, the study built a virtual ERP dataset from the United States Department of Agriculture (USDA) Forest Service timber harvests feature layer (828,789 source records, 70 attributes). A NumPy-based pipeline (RANDOM_SEED=42) localized the data to six Manicaland forestry regions, three commercial species (pine 55%, eucalyptus 35%, and wattle 10%), four contractor scales, and region-specific estate mappings. Each record was engineered as a decision-time allocation observation with contractor attributes, seasonal conditions, distance and route indicators, terrain slope, estimated extraction volume, cost proxies, mill compatibility variables, and the target late_delivery_risk. The resulting 46-feature schema was designed to exclude post-contract leakage while still reflecting the operational information available to estate managers before contractor commitment.

3.5. Dataset splitting and class distribution

The dataset was divided into training (60%, n=497,273), validation (20%, n=165,758), and holdout test (20%, n=165,758) sets using stratified train_test_split to maintain class proportions. The three-way split is a methodological requirement: the independent validation set allows threshold optimization and model selection without contaminating the holdout set, which was evaluated exactly once after all design choices were finalized [30]. The positive class rate (late_delivery_risk=1) was 30.93% across all partitions, indicating moderate class imbalance, which was addressed using SMOTE within the training pipeline.

3.6. Experimental setup and baselines

Four classifiers were trained and tested in order to allow robust comparative analysis: logistic regression (linear baseline), decision tree (non-ensemble baseline), random forest (ensemble benchmark), and XGBoost (primary model). All were built on the same Scikit-Learn pipelines using ColumnTransformer preprocessing (median imputation+StandardScaler for numerics; OneHotEncoder for categoricals), the same stratified folds, and evaluated on the same set of metrics: ROC-AUC, PR-AUC, precision, recall, F1, and Brier score. This design ensures that performance differences are attributable to model architecture rather than data preprocessing. To evaluate the performance of DRAF as a whole, it was compared with two implicit baselines: i) random assignment, i.e., uniform random assignment of contractor-block combinations from the feasible pool, representing the worst-case practical baseline and ii) a cost-only greedy heuristic (sequential allocation of the lowest-cost feasible contractor to each block based on ERP-centric allocation logic). The optimization was implemented with population size 200, 300 generations, and tournament selection pressure of 3, with the results replicated across 5 independent NSGA-II seeds to check the stability of the solution. The experiment did not include MOEA/D, SPEA2, or deep neural risk models because the revision scope was to validate the proposed architecture under a controlled baseline set; these algorithms are explicitly identified as the next comparative layer for future work.

4. RESULTS AND DISCUSSION

4.1. Predictive model performance

The complete evaluation results for all four classifiers on both validation and holdout test sets are presented in Table 1. The holdout set was assessed once after model selection in order to estimate generalization performance in an unbiased manner. Based on the highest validation PR-AUC (0.869), XGBoost was chosen as the final model. The holdout confusion matrix was: true negative (TN)=98,896, FP=15,598, FN=62, true positive (TP)=51,202. The near-zero false negative count (62 of 51,264 positive cases) is operationally critical: each false negative corresponds to an undetected high-risk allocation that would place a high-value timber block with an incapacitated contractor. With a Brier score of 0.053 the probability outputs are very well tuned and can be directly mapped onto the optimization objective functions without doing any retuning post-hoc. The ROC-AUC improvement of 0.365 over logistic regression is in line with benchmarks reported in similar contractor risk contexts [7], [8] and provides empirical support for the ensemble non-linear architecture outperforming linear alternatives on this class of problem. The very low PR-AUC of logistic regression (0.310) confirms that the risk classification problem is not linearly separable and that linear discriminants are practically insufficient for operational deployment.

Table 1. Classifier performance comparison—validation and holdout test sets (primary target: late_delivery_risk)

Model	Val. prec.	Val. rec.	Val. F1	Val. ROC-AUC	Val. PR-AUC	Test F1	Test ROC-AUC	Test Brier
XGBoost ✓	0.765	0.999	0.867	0.965	0.869	0.867	0.965	0.053
Random forest	0.763	0.999	0.865	0.963	0.866	0.864	0.963	0.055
Decision tree	0.765	0.999	0.866	0.964	0.867	0.865	0.962	0.056
Logistic regression	0.309	1.000	0.472	0.600	0.310	0.471	0.601	0.213

4.2. Feature importance and shapley additive explanations (SHAP) analysis

Rainfall-season variables were identified as the leading drivers of risk both through Gini feature importance and SHAP analysis. is_wet_season had the highest Gini importance (0.676) whereas monthly_rainfall_mm had the highest mean absolute SHAP value (5.247). Together, wet-season, rainfall-level, and rainfall-volume variables account for more than 84% of predictive importance, which is consistent with société générale de surveillance (SGS) audit observations that wet-season road degradation is a major cause of contractor delivery failure in Zimbabwean forestry [31].

4.3. Multi-objective optimization results

Of 828,789 total candidate assignments, 639,772 (77.27%) passed strict feasibility filtering (terrain slope, route fit, mill compatibility, and recovery feasibility) and entered the NSGA-II optimization pool. Table 2 reports the scenario-level optimization outcomes. Solution diversity was confirmed in all cases: each solution had a distinct contractor-block-mill assignment set (max_repeated_set_count=1), which indicates true discovery of the Pareto front and not premature convergence. The balanced and high-recovery scenarios produced 64 fully feasible, non-dominated solutions each with zero constraint violations, giving estate managers a rich and usable decision space along the cost-recovery-delay-reliability trade-off front. The balanced scenario is therefore the most deployable policy setting because it keeps reliability constraints active without suppressing feasible yield options, whereas the high-recovery scenario is most useful when mills face throughput pressure and can tolerate moderate risk. Such a low-risk scenario infeasibility (mean violation=0.036) is analytically instructive. The overly stringent risk ceiling of 0.45 is violated because the predicted mean risk in the candidate pool is 0.201; combined with tight load limits (0.95) and a limited planning horizon (5 blocks), fewer than the minimum number of constraint-satisfying assignment combinations are possible for a fully feasible Pareto set. This is a constraint-feasibility boundary condition: the framework surfaces this condition explicitly through the dashboard’s scenario comparison panel instead of silently returning a sub-optimal solution. Estate managers therefore receive clear operational guidance that a risk ceiling below 0.45 requires either expanding the contractor registry, relaxing load limits, or shifting wet-season blocks to a later planning window.

Table 2. NSGA-II scenario optimization summary

Scenario	Blocks	Options/block	Pareto solutions	Feasible (%)	Mean constraint violation
Balanced	6	14	64	100.0	0.000
High-recovery	6	16	64	100.0	0.000
Low-risk	5	12	32	0.0	0.036

4.4. Comparison against baselines

Table 3 presents a structured comparison of DRAF against the baseline methods evaluated. The salient reason why DRAF is better than the cost-greedy heuristic is not simply the dashboard; we insert risk into the optimization objective itself. However, a risk-free NSGA-II can still uncover a Pareto front and may instead favour cheap or high-yield assignments that are fragile in practice. DRAF’s risk-discounted recovery target pushes these options to the rear, and allows managers to allocate sets of resources that are still cost-effective since they control for delivery risk [32].

Table 3. DRAF versus baseline methods: performance across all evaluation dimensions

Dimension	Random assignment	Cost greedy heuristic	Logistic regression +MOO	DRAF (XGBoost+NSGA-II)
Prediction ROC-AUC	—	—	0.600	0.965
Recall (risk detection)	—	—	1.00 (poor precision)	0.999 (with precision)
Risk in objective function	No	No	Threshold filter only	Embedded multiplicatively
Pareto solutions (balanced)	N/A	1 (single heuristic)	~8–12 (risk-free front)	64 (risk-integrated)
Mean constraint violations	High (~23% infeasible)	Moderate	Low	Zero (balanced, high-rec.)
Decision support	None	Static ERP table	Offline report	Real-time interactive DSS
Stakeholder satisfaction	—	—	—	4.19/5.0

4.5. Stakeholder evaluation results

After a structured demonstration, the DRAF artefact was evaluated by thirty sector representatives: eight estate managers, eight contractors, seven IT/ERP staff, and seven policy representatives. The section-level mean Likert scores are shown in Table 4. Methodological transparency: this was rated (4.6/5.0) highest for IT/ERP staff with the fixed random seed (RANDOM_SEED=42), stratified splits and serialized model artefacts being found to distinguish them from ad hoc analytical scripts. Estate managers rated the framework highest overall (4.26/5.0), with qualitative feedback highlighting the utility of being able to test a heavy-rainfall scenario before seasonal commitment. Contractors came in with the lowest total score (3.98/5.0), but methodological transparency appeared to be a major issue (3.8) in part because of the disparity between statistical rigour and accessibility for field operators, suggesting that a simplification on the risk-score explanation layer has given the project its short-term focus of development. The paper presents an integrated risk-integrated MOO-based risk management framework for operational resource allocation, formalized as a mathematical pattern on a conceptual level as an architectural design pattern.

4.6. Theoretical contribution and discussion

The paper's key contribution to the theoretical community was that it defines an architectural approach for a formalized structure for operational resource optimisation from risk-integrated MOO. Although previous works have integrated ML and MOO in supply chain settings [29], [32], in traditional architecture ML results are considered as feasibility filters or diagnostics after optimization. DRAF reverses this relation: the ML-generated risk probability can be incorporated into the recovery objective as a discount factor–so the evolutionary search will be dictated by the probability in the form of the risk itself. This is similar to risk adjusted return optimization in portfolio theory, but implemented here to an integer assignment problem with terrain and capacity constraints. The contribution is also nestled into decision-support architecture theory: predictive analytics, optimization, and human scenario selection are treated as coupled services, not silo-based reports [26], [28]. In addition, the framework is directly applicable to distributed systems and cloud-based ERP expansion, as each layer can be exposed as a modular service once real contractor, rainfall and mill-throughput data streams are collected. These result in structurally different Pareto fronts compared with risk-free fronts: they systematically exclude strategies that are economical but operationally unsound, giving decision makers a Pareto-optimal set of solutions in the practically relevant objective space and not a narrow deterministic efficiency space.

Second, the model establishes that the dominant driver of operating risk in the Zimbabwean commercial forestry sector is seasonal rainfall. The SHAP analysis, combined with exploratory data analysis (EDA) cross-phase correlation results, indicates that is_wet_season and monthly_rainfall_mm is jointly significant predictors accounting for more than 84% of feature importance mass. This conclusion, derived from a reproducible computational pipeline rather than qualitative expert opinion, gives empirical support for seasonal allocation scheduling, an insight that estate managers can adopt without deploying the full DRAF pipeline. From a policy perspective, the result suggests that national forestry resource-management guidelines should require contractors to maintain wet-season readiness evidence, such as road-access plans, equipment redundancy, and verifiable delivery histories, before high-risk blocks are allocated during rainfall-sensitive periods.

Table 4. Stakeholder evaluation scores by section and group (1–5 Likert scale)

Evaluation section	Estate managers	Contractors	IT/ERP staff	Policy makers	Overall
Originality and novelty	4.3	4.0	4.5	4.3	4.3
Practical applicability	4.2	4.1	4.0	3.9	4.1
Methodological transparency	4.1	3.8	4.6	4.2	4.2
Clarity and presentation	4.3	3.9	4.2	4.1	4.1
Evaluation vs. baselines	4.4	4.1	4.5	4.3	4.3
Overall mean	4.26	3.98	4.36	4.16	4.19

4.7. Methodological transparency and reproducibility

DRAF is realized as a single self-contained Jupyter Notebook with a fixed random seed (RANDOM_SEED=42), deterministic stratified splits, serialized model artefacts (XGBoost pipeline at outputs/models/best_model_pipeline.joblib), and structured output directories. All reported metrics are produced by this Notebook and can be recreated by running the cells in turn in a fresh Python 3.x environment with the appropriate library versions. This reproducibility architecture is a deliberate response to the replication crisis in applied ML, in which reported performance metrics are often non-reproducible by default owing to stochastic training, undocumented preprocessing, or uncontrolled data leakage [33]. The leakage-safe feature engineering protocol, which excludes post-outcome variables (actual recovery rate, actual cost, and delivery status) from the predictor set, is particularly important. In operational ML models, leakage inflation results in over-optimistic model performance estimates that collapse in production. The leakage validation layer (cell 11 of the Notebook) shows which columns were excluded, with explanation and justification for each exclusion, enabling complete auditability.

4.8. Applicability, deployment, and limitations

Deployment hinges on three parameters: structured contractor records, alignment to ERP schemas, and managerial readiness to act on algorithmic recommendations. Stakeholder feedback indicates short-term feasibility, given that IT/ERP staff rate integration feasibility as 4.0/5.0 and estate managers supported scenario-sensitivity tools. However, live use would still demand real-time ERP integration, mobile field capture, role-based access controls, periodic model recalibration, and a governance process to override recommendations when local knowledge contradicts the model. Four caveats may be identified. The synthetic dataset means the holdout metrics are an upper bound; live ERP data will have missingness, bias, organizational idiosyncrasies, and noise not represented in the simulation. Rainfall risk was simulated rather than built from station-level weather, road-access, and contractor-downtime records, which may overstate the dominance of seasonal variables. Optimization was tested on portfolios of 5-6 planning blocks rather than full estate schedules, so scaling behaviour needs further assessment. Ultimately, the dashboard was assessed via demonstration rather than longitudinal deployment, meaning actual adoption, trust, and override behaviour remain open empirical questions.

5. CONCLUSION

In this paper, we introduced DRAF, an integrated decision-intelligence framework for dynamic contractor allocation in Zimbabwe's commercial forestry sector. The three constituents of the model: embedding risk-probability ML within NSGA-II objective functions, domain-calibrated virtual ERP dataset and an active interactive decision support system (DSS). Together, these components address the decision-intelligence gap that conventional ERP-centric allocation management could not overcome. Empirically, the DRAF XGBoost classifier achieved a hold-out ROC-AUC of 0.965, recall of 0.999, and a 0.365 ROC-AUC improvement compared with the logistic regression baseline. The NSGA-II optimizer generated 64 fully feasible Pareto solutions in both the balanced and high-recovery scenarios, whereas the low-risk scenario revealed a useful constraint-feasibility boundary for managers. Stakeholder evaluation produced an overall satisfaction score of 4.19/5.0, and SHAP analysis showed that seasonal rainfall variables accounted for more than 84% of predictive importance, converting model interpretation into actionable scheduling guidance. The risk-integrated MOO formulation proposed here - embedding ML probabilities directly into the recovery discount factor of an evolutionary optimizer's objective function-represents a transferable architectural contribution applicable to resource-allocation problems involving multiple contractors with stochastic operational behaviour. Future work should validate the framework using multi-season ERP, rainfall, and contractor-performance data; compare NSGA-II against MOEA/D, SPEA2, stochastic optimization, reinforcement-learning, and deep tabular risk models; scale optimization to full estate-level portfolios; and test subgroup fairness implications for micro, small, medium, and large contractors.

ACKNOWLEDGMENTS

The authors acknowledge the academic guidance of Mr B. Mupini, the School of Information Science and Technology at Harare Institute of Technology, and the practitioners who participated in the stakeholder evaluation.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Tavengwa Norman	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓		✓	
Brian Mupini	✓	✓		✓	✓					✓		✓	✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The Virtual ERP dataset that supports the findings of this study was derived from the publicly available USDA Forest Service Timber Harvests Feature Layer. The synthetic dataset and the analysis notebook are available from the corresponding author upon reasonable request.

REFERENCES

[1] Timber Producers Federation, *Annual Timber Sector Performance Report 2024*. Harare: TPF Zimbabwe, 2024.
[2] African Forest Forum, *Forestry Value Chain Analysis in Selected African Countrie*s. Nairobi: AFF, 2023.
[3] Zimbabwe National Statistics Agency (ZimStat), "Zimbabwe Agricultural Statistics Report 2024," Harare: ZimStat, 2024.
[4] "Timber production rebounds amid contractor restructuring," *Sunday News*, Mar. 02, 2025. Accessed: Feb. 19, 2026. [Online]. Available: https://www.sundaynews.co.zw/timber-production-rebounds-amid-contractor-restructuring/
[5] S. Chibaro, B. Mapani, and T. Dube, "Predictive analytics in Zimbabwean forestry operations," *African Journal of Forestry Research*, vol. 12, no. 2, pp. 88–104, 2024.
[6] Chips Business Solutions, *ERP Implementation in Zimbabwe's Timber Industry*. Harare: CBS, 2024.
[7] L. Zhao, X. Li, and Y. Sun, "Machine learning–based contractor risk assessment in industrial supply chains," *Expert Systems with Applications*, vol. 235, p. 121247, 2024.
[8] O. Filani, P. Ndlovu, and K. Moyo, "Risk scoring frameworks in emerging market supply chains," *International Journal of Supply Chain Analysis*, vol. 4, no. 1, pp. 22–39, 2025.
[9] L. Grinsztajn, E. Oyallon, and G. Varoquaux, "Why do tree-based models still outperform deep learning on typical tabular data?," *Advances in Neural Information Processing Systems*, vol. 35, pp. 507–520, Jul. 2022, doi: 10.52202/068431-0037.
[10] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, Apr. 2002, doi: 10.1109/4235.996017.
[11] K. T. Malladi and T. Sowlati, "Optimization in forest supply chain management: a review," *Current Forestry Reports*, vol. 7, no. 2, pp. 100–118, 2021, doi: 10.1007/s40725-021-00137-5.
[12] C. S. Tang, "Perspectives in supply chain risk management," *International Journal of Production Economics*, vol. 103, no. 2, pp. 451–488, Oct. 2006, doi: 10.1016/j.ijpe.2005.12.006.
[13] D. Ivanov and A. Dolgui, "Viability of intertwined supply networks: extending the supply chain resilience angles towards survivability. A position paper motivated by COVID-19 outbreak," *International Journal of Production Research*, vol. 58, no. 10, pp. 2904–2915, May 2020, doi: 10.1080/00207543.2020.1750727.
[14] J. Rix, B. Wolff, and F. Müller, "Machine learning applications in forestry harvesting operations: a systematic review," *European Journal of Forest Research*, vol. 142, no. 4, pp. 771–790, 2023.
[15] A. R. Hevner, S. T. March, J. Park, and S. Ram, "Design science in information systems research," *MIS Quarterly: Management Information Systems*, vol. 28, no. 1, pp. 75–105, 2004, doi: 10.2307/25148625.
[16] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, "A design science research methodology for information systems research," *Journal of Management Information Systems*, vol. 24, no. 3, pp. 45–77, 2007, doi: 10.2753/MIS0742-1222240302.
[17] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.
[18] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug. 2016, pp. 785–794, doi: 10.1145/2939672.2939785.
[19] M. Al Mnaseer, M. Alshurideh, and S. Al-Hawary, "Hybrid machine learning approaches for contractor performance classification," *Journal of Business Analytics*, vol. 8, no. 1, pp. 45–63, 2025.
[20] R. Shwartz-Ziv and A. Armon, "Tabular data: Deep learning is not all you need," *Information Fusion*, vol. 81, pp. 84–90, Nov. 2022, doi: 10.1016/j.inffus.2021.11.011.
[21] M. T. M. Emmerich and A. H. Deutz, "A tutorial on multiobjective optimization: fundamentals and evolutionary methods," *Natural Computing*, vol. 17, no. 3, pp. 585–609, Sep. 2018, doi: 10.1007/s11047-018-9685-y.
[22] P. Auger, L. T. Bui, and M. H. Nguyen, "Advances in multi-objective evolutionary algorithms for supply chain optimization," *Computers & Operations Research*, vol. 150, p. 106082, 2023.
[23] M. Acuna, J. Skinnell, T. Evanson, and R. Mitchell, "Harvesting productivity and cost analysis in forestry operations," *International Journal of Forest Engineering*, vol. 32, no. 3, pp. 202–215, 2021.
[24] X. Liu, Q. Chen, and F. Wang, "Multi-objective construction scheduling under resource constraints," *Automation in Construction*, vol. 145, p. 104617, 2023.
[25] K. Patejuk, A. Baturo-Cieśniewska, W. Pusz, and A. Kaczmarek-Pieńczewska, "Biscogniauxia charcoal canker—a new potential threat for Mid-European forests as an effect of climate change," *Forests*, vol. 13, no. 1, p. 89, Jan. 2022, doi: 10.3390/f13010089.
[26] C. W. Holsapple, A. Lee-Post, and R. Pakath, "A unified foundation for decision support systems," *Decision Support Systems*, vol. 64, pp. 721–728, 2014.
[27] T. Maxime, G. Kambanje, and B. Nyoni, "Supply chain optimization in Southern African forestry sectors," *The African Journal of Information Systems*, vol. 9, no. 3, pp. 101–119, 2023.
[28] M. Haddara and A. Elragal, "The readiness of ERP systems for the factory of the future," *Procedia Computer Science*, vol. 64, pp. 721–728, 2015, doi: 10.1016/j.procs.2015.08.598.
[29] A. Raj, G. Dwivedi, A. Sharma, A. B. L. de S. Jabbour, and S. Rajak, "Barriers to the adoption of Industry 4.0 technologies in the manufacturing sector: an inter-country comparative perspective," *International Journal of Production Economics*, vol. 224, p. 107546, 2020, doi: 10.1016/j.ijpe.2019.107546.
[30] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, Jun. 2006, doi: 10.1016/j.patrec.2005.10.010.
[31] SGS, "Forest management certification report: Zimbabwe operations," 2021.
[32] S. Hosseini, N. Tajik, D. Ivanov, M. D. Sarder, K. Barker, and A. Al Khaled, "Resilient supplier selection and optimal order allocation under disruption risks," *International Journal of Production Economics*, vol. 213, pp. 124–137, Jul. 2019, doi: 10.1016/j.ijpe.2019.03.018.
[33] A. L. Boulesteix, S. Lauer, and M. J. A. Eugster, "A plea for neutral comparison studies in computational sciences," *PLoS ONE*, vol. 8, no. 4, p. e61562, Apr. 2013, doi: 10.1371/journal.pone.0061562.

BIOGRAPHIES OF AUTHORS

Tavengwa Norman is a multidisciplinary academic and researcher with qualifications spanning data science, forensic auditing, corporate governance, peace and conflict studies, and security studies. He completed a Master of Technology in Data Science and Analytics at the Harare Institute of Technology and holds a Master of Commerce in Forensic Auditing and Accounting, a Master of Science in Peace, Leadership, and Conflict Resolution, a Postgraduate Diploma in Applied Corporate Governance and Strategic Leadership, and a Bachelor of Business Administration (Honours) in Police and Security Studies. He is also pursuing a Bachelor of Substantive Laws at the University of Zimbabwe. His research interests include artificial intelligence, data analytics, enterprise risk management and public-sector governance. He can be contacted at email: h240742c@hit.ac.zw.

Brian Mupini is currently working as a lecturer in the Computer Science Department at the Harare Institute of Technology. He received his first degree, a Bachelor of Technology in Computer Science, from the same institute in 2017. He also holds a Master of Technology in Cloud Computing from the Harare Institute of Technology, attained in 2024. He is also a researcher in the computer science field. His main research interests focus on natural language processing (NLP), artificial intelligence (AI), cryptography and security systems, cloud computing, internet of things (IoT), and data mining. He can be contacted at email: bmupini@hit.ac.zw.