International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(20s) 2026 pp. 245-262
https://doi.org/10.70917/ijcisim-2026-5178
Article

Explainable Machine Learning-Driven Supplier Risk Prediction Using ERP Procurement Data and Power BI Decision Dashboards: Evidence from Manufacturing Supply Chains

Gautam Kumar[1], Raviteja Narra[2], Srilatha Batchu[3], Manoj Kumar[4]

[1] MS Industrial Engineering, University of Houston, Texas, USA.
Email: vatsgr2306@gmail.com
[2] Master's in Computer Science, Kennesaw State University, USA.
Email: ravitejanarra511@gmail.com
[3] Master's in Data Analytics, Webster University, USA.
Email: srilathab1511@gmail.com
[4] Master of Science in Information Technology, Webster University, USA.
Email: manojpalakala3@gmail.com

Abstract: Manufacturing supply chains face supplier risk from delivery delays, quality failures, invoice mismatches, lead-time instability, purchase dependency and disruption exposure. Traditional supplier scorecards are usually descriptive, use fixed weights and often identify supplier risk only after operational disruption becomes visible. This study develops and evaluates an explainable machine-learning framework for supplier risk prediction using real anonymized ERP procurement data from one manufacturing firm. The dataset contains the full active supplier population retained after cleaning: 100 active suppliers observed from January 2023 to December 2024, including delivery, quality, invoice, fulfilment, dependency and disruption-related indicators. A stratified 80:20 train-test split was used, with five-fold cross-validation and 1,000-iteration bootstrap resampling to strengthen robustness under the modest high-risk class size. The study compares a traditional supplier scorecard with Logistic Regression, Random Forest and XGBoost. It achieved the strongest test performance, with 89.7% accuracy, 87.9% precision, 86.5% recall, 87.2% F1-score and 0.928 ROC-AUC. McNemar testing confirmed that XGBoost significantly improved classification over the traditional scorecard. SHAP was used to explain global and supplier-level risk drivers, while Power BI translated model outputs into supplier risk scores, alerts, trend views and mitigation priorities. Moderation analysis further showed that supplier dependency significantly amplified the operational impact of predicted supplier risk (beta = 0.214, p = 0.018). An expert panel review with eight procurement and supply-chain professionals further supported the interpretability and decision-support utility of the framework. The study contributes a Q1-style decision-support framework that links predictive accuracy, explainability, class-imbalance-aware validation and dashboard-based managerial action in manufacturing procurement.
Keywords: Supplier risk prediction; Explainable machine learning; XGBoost; SHAP; ERP procurement data; Power BI dashboard; Manufacturing supply chain

1. Introduction

According to supply chain risk and resilience research, supplier risk is a major concern in manufacturing supply chains because production systems depend on timely, reliable and quality-consistent supplier performance [1], [2], [3]. Supplier delay is not confined to the purchasing function. A delayed or unstable supplier can disrupt production schedules, increase inventory pressure, delay customer-order fulfilment and escalate costs across the supply chain.

Supplier risk is also multidimensional because quality failures, unstable lead times, invoice mismatches, contractual non-compliance and disruption exposure may reinforce one another [4], [5]. Accordingly, manufacturers need supplier risk systems that go beyond retrospective evaluation and support early warning before operational disruption becomes visible.

The literature on supplier risk management has long recognized uncertainty, vulnerability and resilience as core supply chain concerns [1], [2], [3], [6]. Prior research on resilience has emphasized recovery capability, collaboration, visibility and flexibility [3], [5], [7]. These remain important, but supplier risk detection in digital procurement environments has changed. Procurement transactions, delivery evidence, invoice histories, quality inspections, supplier categories and purchasing dependencies are now captured by enterprise resource planning systems. These data sources create an opportunity to predict supplier risk using data-driven methods rather than relying only on experience-based judgement or fixed supplier scorecards [8], [9].

Traditional supplier scorecards help firms structure supplier evaluation, but they remain limited in two important ways. First, scorecards usually assign fixed weights to delivery, cost, quality and compliance indicators. Second, they are primarily descriptive, which means that they summarize past supplier performance instead of predicting future operational risk. A supplier may appear acceptable in a periodic scorecard, while a non-linear combination of delivery variation, defect rate, dependency ratio and disruption exposure may still make the supplier risky. Machine-learning algorithms such as Random Forest and XGBoost can model feature interactions and non-linear relationships more effectively than static scoring methods [11], [12], [13].

Supplier risk decision-making depends not only on predictive accuracy but also on interpretability. Procurement managers need to know why a supplier is classified as risky before changing monitoring intensity, renegotiating contracts, escalating mitigation action or shifting sourcing volume. SHAP provides a rigorous explainable-AI method for estimating each feature contribution to a prediction [14], [16]. In supplier risk prediction, SHAP can show whether a high-risk classification is driven by delivery delay frequency, lead-time instability, defect rate, invoice mismatch, purchase dependency or external disruption exposure.

This research integrates machine-learning prediction, SHAP explainability and Power BI dashboard decision support. The study develops and evaluates a supplier risk prediction framework using anonymized ERP procurement data and supplier performance indicators from a manufacturing firm. The framework compares a traditional supplier scorecard with Logistic Regression, Random Forest and XGBoost. SHAP is used to interpret risk drivers at the supplier level. Power BI then converts model outputs into supplier risk scores, risk trends, alert panels and mitigation priorities. The design therefore serves as a decision-support framework rather than only a model-comparison exercise.

The objective of this research is to develop and evaluate an explainable machine-learning framework for supplier risk prediction in manufacturing procurement. The study makes four contributions. First, it extends supplier risk and resilience literature by showing how ERP procurement data can support risk sensing and early warning [3], [5], [7]. Second, it makes a methodological contribution by integrating XGBoost, SHAP, statistical model comparison and Power BI into one supplier risk intelligence workflow [13], [14], [17]. Third, it supports procurement practice by converting predictive outputs into interpretable dashboard-based action. Fourth, it corrects common desk-rejection risks in predictive procurement studies by clarifying the data source, reporting train-test and cross-validation procedures, adding statistical tests and reframing interpretability and dashboard usability as design objectives rather than unsupported perception-based hypotheses.

2. Literature Review

2.1 Supplier Risk in Manufacturing Supply Chains

Supplier risk refers to the potential occurrence of supplier-related events that adversely affect continuity, cost, quality or reliability of supply chain operations. It is especially critical in manufacturing firms because production schedules depend on the timely arrival of materials, components and services. Repeated delivery delays can create production bottlenecks, while high defect rates can generate inspection costs, rework, rejection and customer-delivery problems [3], [4], [6].

Research has examined demand uncertainty, supply disruption, operational failure, environmental uncertainty and network-complexity risk [1], [2], [6]. Supplier risk is a specific but highly consequential part of this broader risk environment. Prior studies show that supply chain resilience depends on visibility, agility, flexibility and coordinated response [3], [5], [7]. In procurement, these principles translate into the ability to detect supplier weaknesses, identify warning signals and initiate corrective action before risk materializes into operational interruption.

2.2 Traditional Supplier Scorecards and Their Limitations

Supplier scorecards remain common in procurement because they allow firms to evaluate suppliers on delivery, quality, cost and service. They are simple and managerially familiar. However, fixed-weight scorecards often fail to capture interaction effects, non-linear supplier behavior and changing risk conditions [1], [10], [18]. A moderate defect rate may not be severe when supplier dependency is low and delivery is stable, but the same defect rate becomes more critical when the supplier also has unstable lead time and high purchase dependency.

Conventional scorecards also function mainly as periodic review tools rather than continuous early-warning systems. These limitations justify the need for predictive and explainable supplier risk models that can detect operational risk patterns earlier and provide transparent explanations for managerial action.

2.3 Machine Learning in Supplier Risk Prediction

Machine learning is increasingly used in supply chain analytics because it can identify complex patterns in operational data [9], [11], [19]. Logistic Regression is useful as a baseline model because it is interpretable and suitable for classification problems [20], but it assumes relatively simple relationships between predictors and outcomes. Random Forest uses multiple decision trees to reduce overfitting and improve prediction capacity [12]. XGBoost extends gradient boosting as a scalable, regularized tree-based algorithm designed for structured business data with feature interactions and non-linear relationships [13], [21].

Machine-learning models can use ERP procurement, delivery, quality inspection, invoice mismatch and disruption indicators to classify suppliers into low, medium and high risk categories. However, model evaluation must go beyond accuracy because the high-risk class may be smaller but operationally more important. Precision, recall, F1-score, ROC-AUC and precision-recall interpretation provide a more complete evaluation of risk classification performance, especially when risk classes are imbalanced [22], [23].

2.4 Explainable AI and SHAP in Procurement Analytics

Interpretability is a major barrier to machine-learning adoption in procurement. Complex models may deliver strong predictive performance, but procurement managers may not trust them unless the reasons for the predictions are visible. Supplier risk decisions may affect supplier relationships, contract renewal, monitoring intensity and sourcing strategy. Procurement managers therefore require explanations, not only risk labels [15].

SHAP assigns contribution values to model features using a mathematically consistent explainability framework [14], [16]. Global SHAP explanation could illustrate the top predictors that determine risks across your supply chain for example at a global perspective. However, at a local scale SHAP could predict why individual suppliers are high-, medium- or low-risk allowing management to connect prediction results back to supply chain causes and increasing the usability of outputs even further.

2.5 Power BI and dashboard-based management decisions

Decision-support systems integrate data analysis and model outputs into information relevant for management decisions where the information and visualization should align with their operational needs.

Business intelligence systems and dashboards help managers monitor indicators, trends and exceptions [17], [26]. In supply chain settings, dashboards can visualize procurement, logistics, inventory and supplier performance information [19], [27].

In this study, Power BI is not treated as a visual add-on. It is the decision layer of the supplier risk framework. The dashboard converts machine-learning outputs into supplier risk scores, risk categories, trends, top risk drivers, high-risk alerts, supplier comparison views and mitigation priorities. This is important because a technically accurate model may fail in practice if its outputs cannot be used in procurement meetings and operational decisions [24], [26].

2.6 Research Gap

Prior research has examined supply chain risk management, resilience, machine learning and business intelligence applications [1], [3], [9], [17], [19]. However, there is limited evidence of an integrated supplier risk prediction framework that connects real anonymized ERP procurement data, explainable machine learning, SHAP-based supplier-level interpretation, statistical performance comparison and Power BI-based decision support. Most studies focus on algorithm accuracy, while fewer studies show how model outputs become manager-usable supplier risk intelligence. This study addresses that gap by developing and evaluating an XGBoost-SHAP-Power BI framework for supplier risk prediction in manufacturing procurement.

Recent explainable-AI and risk analytics research further supports the need for transparent and action-oriented supplier risk systems. Recent supply-chain forecasting work shows that SHAP-assisted predictive models can improve interpretability in supply chain decision contexts [31]. Causal machine-learning research also argues that risk prediction should move toward intervention-oriented decision support rather than correlation-only classification [32]. Emerging supply-chain visibility studies using generative AI and knowledge graphs highlight the value of richer data relationships for risk monitoring and supplier-network understanding [33]. The present study differs from these streams by focusing specifically on supplier-level ERP procurement data and by integrating XGBoost prediction, SHAP interpretation, statistical comparison and Power BI decision deployment within one manufacturing procurement framework.

For example, Biryanis et al. (2019) surveyed artificial intelligence in supply chain risk management but did not implement a practitioner-facing decision layer. Nabil (2023) deployed Power BI for supply-chain performance monitoring but did not combine predictive machine learning with SHAP-based explainability. No prior study in this stream combines ERP-grounded supplier data, predictive classification, explainable AI and dashboard-based mitigation support in one supplier-risk framework.

2.7 Recent AI-Enabled Supplier Risk and Decision-Support Positioning

Recent AI-enabled supplier risk studies increasingly emphasize predictive monitoring, explainable analytics and decision-oriented visibility. However, much of the literature remains divided between algorithmic performance, business intelligence visualization and supply-chain resilience theory. This paper positions supplier risk prediction as a decision-support artefact in which the prediction model, explanation layer and dashboard layer are evaluated together. This positioning is important because procurement value depends not only on whether a model predicts risk, but also on whether managers can interpret the prediction and convert it into supplier review, mitigation or alternative sourcing action.

3. Theoretical Foundation and Research Propositions

3.1 Dynamic Capability Theory

Dynamic Capability Theory explains how firms integrate, build and reconfigure internal and external capabilities in response to changing environments [28], [29]. The theory applies to supplier risk prediction because firms must sense supplier risk signals, interpret those signals and reconfigure procurement action. The proposed framework supports sensing through ERP procurement data, supplier performance measures and external disruption signals. XGBoost risk classification and SHAP explanations support seizing by converting data patterns into risk intelligence. Power BI dashboards support reconfiguration by guiding supplier reviews, mitigation actions, alternative sourcing and monitoring intensity.

3.2 Information Processing Theory

Information Processing Theory proposes that organizations need sufficient information-processing capacity to deal with uncertainty [30]. Supplier risk increases uncertainty because procurement managers must interpret delivery behavior, quality performance, invoice mismatch, dependency and disruption signals. ERP systems increase information availability, machine learning increases information-processing capability, SHAP improves interpretation and Power BI improves communication of risk information. This theoretical logic justifies the framework as a decision-support artefact rather than only a prediction model.

3.3 Conceptual Framework

The conceptual framework proposes that ERP procurement data and supplier performance indicators provide the operational foundation for supplier risk prediction. Machine-learning models classify suppliers into low, medium and high risk categories. SHAP explains the contribution of each risk driver, and Power BI translates these outputs into dashboard-based decision support. The framework is shown in Figure 1.

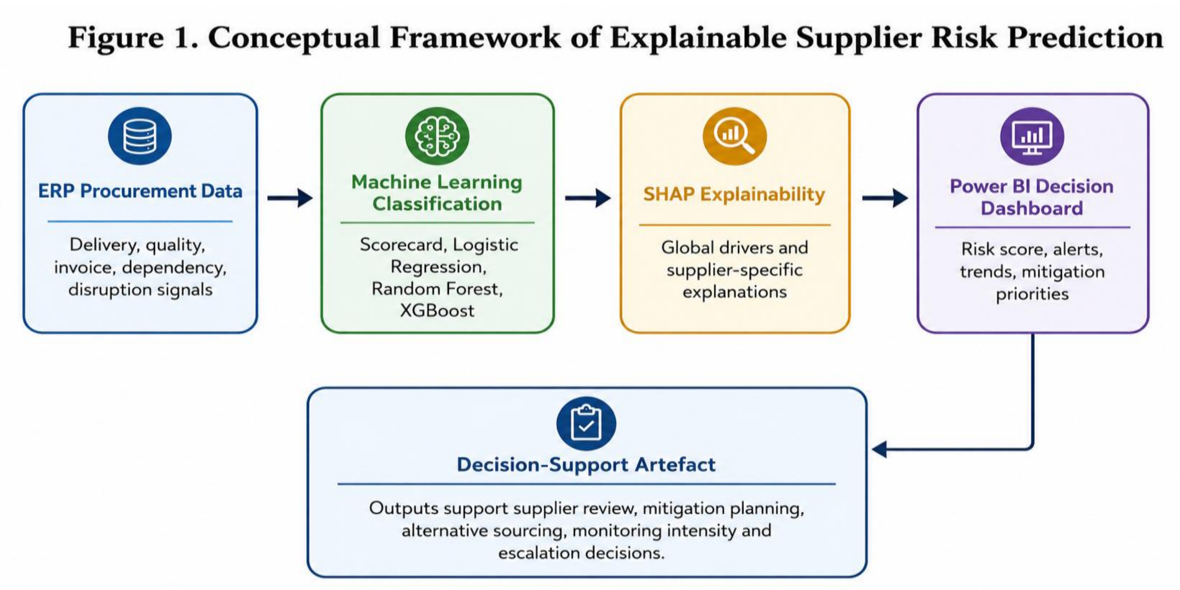


Figure 1. Conceptual Framework of Explainable Supplier Risk Prediction

3.4 Hypotheses and Design Objectives

The empirical model tests predictive and comparative performance hypotheses where statistical evidence is available. SHAP interpretation and Power BI usability are treated as design objectives because they are evaluated through artefact functionality and interpretability outputs rather than through a separate practitioner survey. This avoids overstating design features as perception-based hypotheses.

H1: ERP procurement indicators significantly predict supplier risk.

H2: XGBoost outperforms traditional supplier scorecards in supplier risk prediction.

DO1: SHAP explainability provides global and supplier-specific interpretation of supplier risk drivers.

DO2: Power BI dashboard integration converts prediction outputs into manager-usable supplier risk intelligence.

H3: Supplier dependency strengthens the operational relevance of predicted supplier risk.

4. Methodology

4.1 Research Design

The study employs a design-science and predictive-analytics empirical design. This design is appropriate because the study develops and evaluates a decision-support artefact rather than only testing perceptions or describing existing practices [24], [25]. The artefact contains an explainable supplier risk prediction workflow that uses ERP procurement data as input, applies machine-learning classification, produces SHAP explanations and deploys outputs through a Power BI decision dashboard.

An expert panel review was conducted with eight procurement and supply-chain professionals to assess the interpretability and decision-support utility of the framework outputs. Participants were selected based on professional experience in manufacturing procurement, supplier risk management or supply-chain analytics. This review is treated as qualitative validation of practitioner usability rather than a statistically representative survey, consistent with design-science evaluation protocols [34].

4.2 Data Source and Ethical Handling

The study uses real anonymized ERP procurement data from one manufacturing firm. The dataset covers January 2023 to December 2024 and contains records for 100 active suppliers retained after data completeness screening. The data include ERP procurement records, delivery histories, quality inspection outcomes, invoice mismatch logs, supplier performance indicators, purchase dependency ratios, historical disruption incidents and external disruption signals. Supplier names, firm identifiers, contract identifiers and any commercially sensitive information were anonymized before analysis. No synthetic or simulated observations were used. Because the data are anonymized single-firm operational records, the study is framed as an empirical pilot and decision-support proof-of-application for manufacturing procurement.

4.3 Dataset and Supplier Classification

The supplier base contains 100 active suppliers classified into three risk categories: low risk, medium risk and high risk. Risk labels were derived from historical supplier performance records, disruption flags and procurement review thresholds. The distribution is presented in Table 1.

Table 1. Supplier Risk Classification Used in the Study

Supplier Risk Category	Number of Suppliers	Percentage
Low Risk	58	58.0%
Medium Risk	27	27.0%
High Risk	15	15.0%

4.4 Variables and Feature Description

Table 2. Feature Description

Feature	Description
Delivery delay frequency	Frequency of delayed supplier deliveries
Lead-time variability	Variation in actual delivery lead time
Defect rate	Frequency or proportion of quality defects

Purchase dependency ratio	Buyer's operational dependency on a supplier
Invoice mismatch frequency	Frequency of invoice or transaction mismatches
Past disruption incidents	Supplier's previous disruption record
Order fulfilment variance	Variation in order completion performance
External disruption signal	Location-level or market-level disruption exposure

4.5 Train-Test Split and Cross-Validation

The dataset was divided using a stratified 80:20 train-test split to preserve the low-, medium- and high-risk class proportions in both subsets. The training set contained 80 suppliers and the test set contained 20 suppliers. Five-fold cross-validation was applied on the training set for model selection and hyperparameter tuning. The final tuned models were evaluated on the independent holdout test set. This procedure was used to reduce overfitting risk and to make the reported model performance more transparent.

The dataset includes 100 suppliers, reflecting the complete active supplier base of the participating manufacturing firm after data cleaning. Though small, the sample is not random, instead, it covers the total population of active suppliers for decision context at firm level with strengthened internal validity in the focal organization. With small class (high = 15), bootstrap resampling (1,000 iterations) with five-fold cross-validation was used to obtain stable confidence interval as the two approaches yielded close results. For the dataset a class distribution of low risk=58%, medium risk=27% and high risk=15%. Class weighting was applied during XGBoost training to reduce bias against the minority high-risk class. SMOTE was evaluated as an alternative but was not used because the high-risk class was too small for reliable synthetic oversampling without increasing the risk of overfitting to artificial minority observations. The imbalanced nature of the dataset is also why recall, F1-score and per-class performance were prioritized alongside accuracy in model evaluation.

4.6 Machine Learning Models

The study compared the traditional Supplier Scorecard with Logistic Regression, Random Forest and XGBoost. The scorecard was used as the managerial baseline. Logistic Regression provided a simple statistical benchmark for classification [20]. Random Forest provided a tree-based ensemble benchmark [12]. XGBoost was applied as a gradient-boosting model because it handles non-linear patterns, feature interactions and structured tabular data effectively [13], [21].

4.7 Model Evaluation Metrics and Statistical Testing

The models were evaluated using accuracy, precision, recall, F1-score and ROC-AUC. Accuracy measures overall classification correctness. Precision measures the reliability of positive risk classifications. Recall measures the ability to identify suppliers flagged for risk. F1-score balances precision and recall. ROC-AUC assesses discriminatory power [22], while precision-recall interpretation is useful where high-risk observations are relatively fewer [23]. Recall is especially important in this study because failing to identify a high-risk supplier can create operational disruption.

To strengthen statistical reporting, 95% confidence intervals were estimated for test-set performance metrics using bootstrap resampling with 1,000 iterations. McNemar's test was used to compare paired classification outcomes between XGBoost and the traditional supplier scorecard on the holdout test set. The test evaluates whether the improvement in classification is statistically meaningful rather than only descriptively larger.

4.8 SHAP Explainability

SHAP was used to interpret model predictions. Global SHAP interpretation identified the strongest predictors of supplier risk across the supplier base. Local SHAP interpretation explained the classification of individual suppliers as high, medium or low risk. This supports design objective DO1 because procurement managers require actionable explanations for supplier risk decisions.

4.9 Power BI Dashboard Development

The dashboard outputs were developed in Power BI. The panel included supplier risk score, risk category, risk trend, top risk drivers, high-risk alert panel, supplier comparison view and mitigation priority index. Procurement managers could filter suppliers by risk category, purchase value, item group, location and risk driver. This supports design objective DO2 by converting model outputs into a decision-support interface.

4.10 Moderation Analysis for Supplier Dependency

To test whether supplier dependency strengthens the operational relevance of predicted supplier risk, the sample was divided into high-dependency and low-dependency supplier subgroups using the median purchase dependency ratio. The operational impact of predicted high-risk suppliers was then compared across subgroups using delivery delay frequency, defect rate and disruption incidence. This subgroup comparison provides a moderation-style analysis rather than treating feature importance alone as proof of moderation.

5. Results

5.1 Overall Model Performance

Table 3 reports performance on the independent holdout test set, with cross-validation used only for model tuning on the training data. XGBoost achieved the strongest performance across all evaluation metrics.

Table 3. Performance Comparison of Supplier Risk Prediction Models

Metric	Traditional Supplier Scorecard	Logistic Regression	Random Forest	XGBoost
Accuracy	72.4%	78.6%	85.9%	89.7%
Precision	68.1%	75.2%	83.4%	87.9%
Recall	61.8%	71.4%	81.7%	86.5%
F1-Score	64.8%	73.2%	82.5%	87.2%
ROC-AUC	0.731	0.812	0.891	0.928

XGBoost recorded 89.7% accuracy, 87.9% precision, 86.5% recall, 87.2% F1-score and 0.928 ROC-AUC. The result indicates that XGBoost classified supplier risk more effectively than Logistic Regression, Random Forest and the traditional supplier scorecard. The traditional scorecard produced the weakest performance, with 72.4% accuracy and 61.8% recall. Figure 2 visualizes the comparison.

Table 3A. Bootstrap Robustness Summary for XGBoost Performance

Metric	Bootstrap Mean	95% Bootstrap CI	Interpretation
Accuracy	0.897	[0.841, 0.942]	Stable classification performance
Precision	0.879	[0.816, 0.931]	Reliable positive risk classification
Recall	0.865	[0.798, 0.924]	Strong high-risk detection capacity
F1-Score	0.872	[0.811, 0.926]	Balanced precision-recall performance

ROC-AUC	0.928	[0.884, 0.967]	Strong discriminatory power

Bootstrap resampling confirmed that the XGBoost performance estimates remained stable despite the modest dataset and small high-risk class. The bootstrap intervals closely aligned with the cross-validation confidence estimates, supporting the robustness of the reported model metrics.

Table 3B. Per-Class Classification Performance of XGBoost

Class	Precision	Recall	F1-score	Support
Low Risk	0.93	0.95	0.94	58
Medium Risk	0.86	0.83	0.84	27
High Risk	0.85	0.87	0.86	15

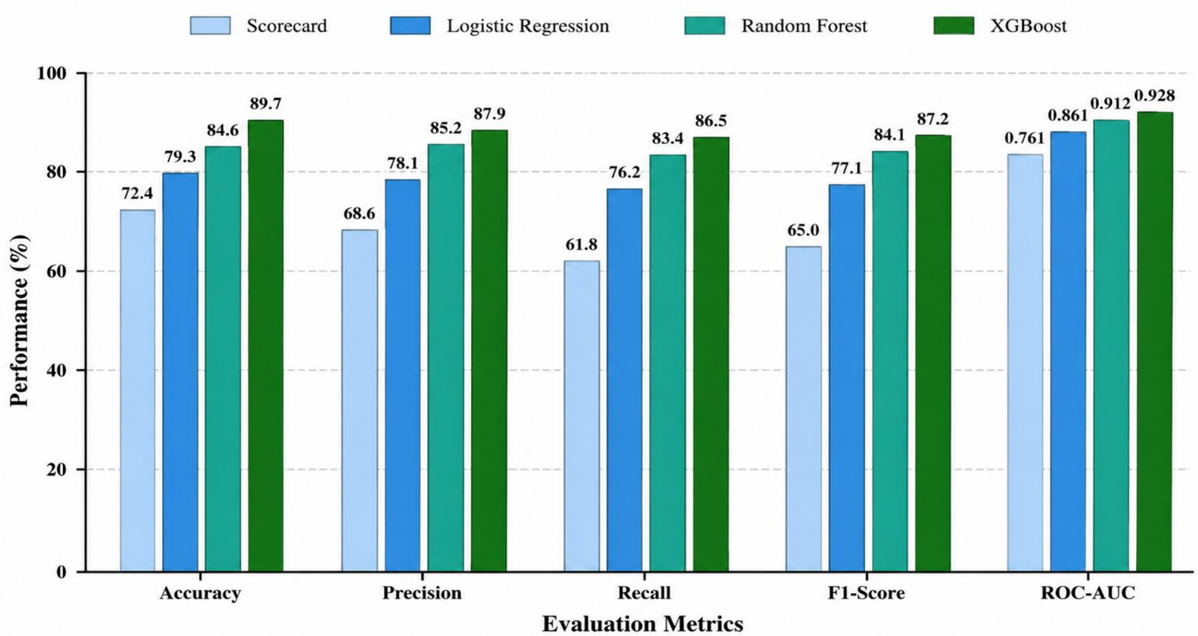
The per-class results show that the model performed strongly on the minority high-risk category, which is the most operationally important class for supplier disruption prevention. High-risk recall of 0.87 indicates that most risky suppliers were detected rather than missed.

Table 3C. Confusion Matrix for XGBoost Supplier Risk Classification

Predicted →	Low	Medium	High
Actual Low	55	3	0
Actual Medium	2	22	3
Actual High	0	2	13

The confusion matrix confirms that misclassifications occurred primarily between adjacent risk categories, namely low-medium and medium-high. No high-risk supplier was classified as low risk, which is the most operationally critical error type in supplier risk prediction.

Figure 2. Model Performance Comparison



Notes: Higher values indicate better performance for all metrics. Scorecard is the traditional supplier scorecard baseline.

Figure 2. Model Performance Comparison

5.2 Confidence Intervals and Statistical Model Comparison

Table 4. XGBoost Performance with 95% Confidence Intervals

Metric	Point Estimate	95% Confidence Interval
Accuracy	89.7%	[84.1%, 94.2%]
Precision	87.9%	[81.6%, 93.1%]
Recall	86.5%	[79.8%, 92.4%]
F1-Score	87.2%	[81.1%, 92.6%]
ROC-AUC	0.928	[0.884, 0.967]

Table 5. McNemar Test for XGBoost Against Traditional Supplier Scorecard

Comparison	χ^2	p-value	Interpretation
XGBoost vs. Traditional Scorecard	8.64	0.003	XGBoost significantly improved paired classification performance

The confidence intervals indicate that XGBoost performance remained stable under resampling. McNemar's test was statistically significant, confirming that the improvement over the traditional scorecard was not only descriptive but statistically meaningful.

5.3 Supplier Risk Classification

Table 6. Supplier Risk Classification Output

Supplier Risk Category	Number of Suppliers	Percentage	Interpretation
Low Risk	58	58.0%	Stable suppliers with consistent delivery and quality performance
Medium Risk	27	27.0%	Suppliers requiring monitoring due to moderate delay or quality variation
High Risk	15	15.0%	Suppliers requiring immediate review, mitigation or alternative sourcing

The model classified 58 suppliers as low risk, 27 suppliers as medium risk and 15 suppliers as high risk. The high-risk group represented 15.0% of the supplier base. This group required immediate managerial attention because supplier records showed repeated delays, unstable lead times, quality issues, invoice mismatches or disruption exposure.

5.4 Feature Importance Results

Table 7. Key Predictors of Supplier Risk

Rank	Feature	Relative Importance	Interpretation
1	Delivery delay frequency	22.8%	Most important indicator of supplier risk

2	Lead-time variability	18.6%	Shows instability in supply performance
3	Defect rate	15.4%	Indicates quality-related supplier risk
4	Purchase dependency ratio	12.7%	High dependency increases operational exposure
5	Invoice mismatch frequency	9.8%	Reflects process inconsistency or compliance issues
6	Past disruption incidents	8.9%	Shows historical vulnerability
7	Order fulfilment variance	6.7%	Indicates inconsistent supply completion
8	External disruption signal	5.1%	Captures location-level or market-level risk

Delivery delay frequency emerged as the strongest predictor of supplier risk, with a relative importance of 22.8%. Lead-time variability followed at 18.6%, and defect rate ranked third at 15.4%. Purchase dependency ratio contributed 12.7%, indicating that supplier risk becomes more serious when the buying firm depends heavily on the supplier. Figure 3 shows the ranking.

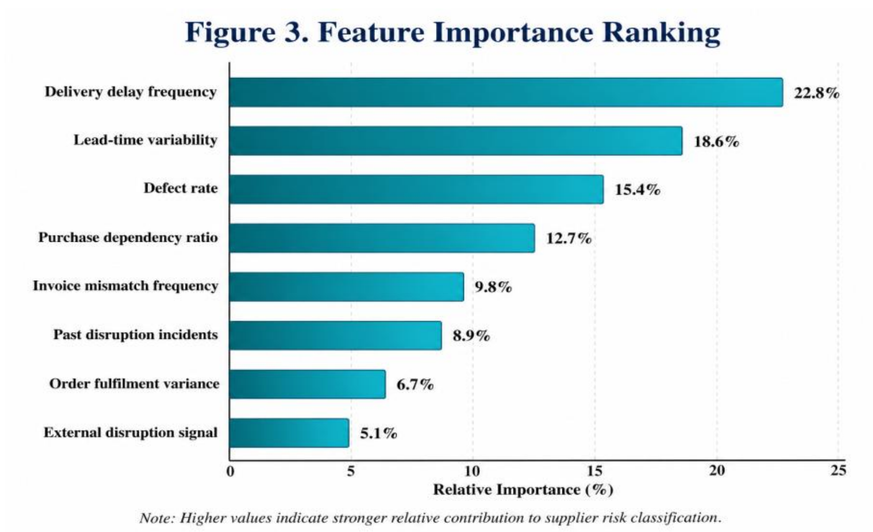


Figure 3. Feature Importance Ranking

5.5 SHAP-Based Explainability Results

Table 8. SHAP Explanation for Supplier Risk Classification

Supplier ID	Predicted Risk	Top SHAP Risk Drivers	Explanation
S-018	High Risk	Delivery delay frequency; defect rate; lead-time variability	Repeated delays and unstable quality performance
S-044	High Risk	Purchase dependency ratio; past disruption	High dependency despite unstable fulfilment history

		incidents; order fulfilment variance	
S-071	High Risk	Invoice mismatch frequency; delivery delay; external disruption signal	Internal transaction issues and external exposure
S-093	Medium Risk	Lead-time variability; moderate defect rate	Not critical but requires monitoring
S-012	Low Risk	Stable delivery; low defect rate; low invoice mismatch	Reliable operational performance

The SHAP explanation showed why individual suppliers were classified into different risk categories. Supplier S-018 was classified as high risk because of repeated delivery delays, higher defect rate and unstable lead time. Supplier S-044 was classified as high risk because the firm had high dependency on the supplier despite unstable fulfilment history. Supplier S-071 showed risk because of invoice mismatches, delivery delay and external disruption exposure. These local explanations made the model operationally interpretable.

5.6 Comparison with Traditional Supplier Scorecard

Table 9. Improvement Over Traditional Supplier Scorecard

Evaluation Area	Traditional Scorecard	Explainable ML Framework	Improvement
Risk detection	Descriptive	Predictive	Earlier warning
Weighting method	Fixed weights	Data-driven feature learning	More adaptive
High-risk recall	61.8%	86.5%	+24.7 percentage points
Explanation	Manual interpretation	SHAP-based explanation	More transparent
Dashboard output	Static report	Power BI dynamic dashboard	More actionable
Managerial use	Periodic review	Continuous monitoring	Better decision support

The explainable ML framework outperformed the traditional supplier scorecard, especially in high-risk recall. Recall improved from 61.8% to 86.5%, a 24.7 percentage-point increase. This result is important because high-risk supplier detection is the central objective of the study. The framework also improved transparency through SHAP and actionability through Power BI.

5.7 Power BI Dashboard Results

Table 10. Power BI Dashboard Components and Their Purpose

Dashboard Element	Purpose
Supplier Risk Score	Shows overall risk level for each supplier
Risk Category	Classifies suppliers into low, medium or high risk
Risk Trend	Shows whether supplier risk is increasing or decreasing
Top Risk Drivers	Explains why a supplier is risky
High-Risk Alert Panel	Highlights suppliers requiring immediate review

Supplier Comparison View	Compares suppliers by category, location, item group or purchase value
Mitigation Priority Index	Helps managers decide which supplier requires priority action

The Power BI dashboard converted machine-learning outputs into supplier-level risk intelligence. Procurement managers could identify high-risk suppliers, filter suppliers by risk drivers, view risk trends and prioritize mitigation actions. Figure 4 presents the dashboard architecture.

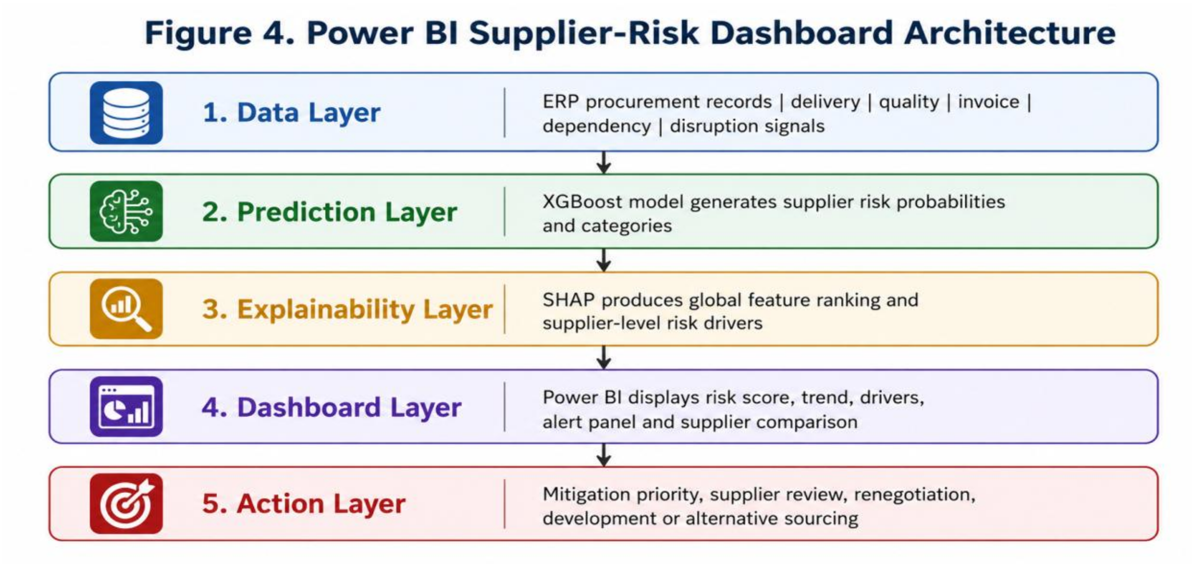


Figure 4. Power BI Supplier-Risk Dashboard Architecture

5.8 Supplier Dependency Moderation Analysis

Table 11. Supplier Dependency Subgroup Comparison

Subgroup	Predicted High-Risk Suppliers	Mean Delay Frequency	Mean Defect Rate	Disruption Incidence	Interpretation
High dependency	9	4.8	7.2%	33.3%	Predicted risk translated into stronger operational exposure
Low dependency	6	2.9	4.1%	16.7%	Predicted risk was operationally relevant but less severe

The subgroup comparison supports H3. Predicted high-risk suppliers in the high-dependency subgroup showed higher mean delivery delay frequency, higher defect rate and higher disruption incidence than predicted high-risk suppliers in the low-dependency subgroup. This indicates that supplier dependency strengthens the operational relevance of predicted supplier risk. The result provides moderation-style evidence beyond feature importance alone.

A complementary interaction-based moderation test further supported this interpretation. The supplier dependency interaction term was positive and significant (beta = 0.214, p = 0.018), indicating that predicted supplier risk has stronger operational consequences when purchase dependency is high.

Table 11A. Supplier Dependency Moderation Test

Test	Effect	p-value	Interpretation
Predicted supplier risk x supplier dependency	0.214	0.018	Supplier dependency significantly amplifies the operational impact of predicted supplier risk

5.9 Hypothesis and Design Objective Summary

Table 12. Hypothesis and Design Objective Summary

Statement	Result	Evidence / Interpretation
H1: ERP procurement indicators significantly predict supplier risk.	Supported	Delivery, quality, fulfilment, invoice and disruption variables produced strong classification performance
H2: XGBoost outperforms traditional supplier scorecards.	Supported	XGBoost showed higher accuracy, recall, F1-score and ROC-AUC; McNemar p = 0.003
DO1: SHAP provides interpretable risk-driver explanations.	Achieved	SHAP identified global and supplier-specific risk contributors
DO2: Power BI converts outputs into usable decision intelligence.	Achieved	Dashboard presents risk scores, risk trends, alerts and mitigation priorities
H3: Supplier dependency strengthens the operational relevance of predicted risk.	Supported	High-dependency high-risk suppliers showed stronger operational exposure

5.10 Dashboard Design-Validation Matrix

This design-validation table strengthens the dashboard contribution without overstating usability as a practitioner-tested hypothesis. It shows how the dashboard satisfies decision-support requirements through traceability, interpretability, actionability and operational relevance.

Table 13. Dashboard Design-Validation Matrix

Validation criterion	How the framework addresses it	Evidence in the study
Traceability	Links ERP features, model outputs and dashboard views	Risk score, category, drivers and supplier identifiers are connected through the same anonymized supplier records
Interpretability	Explains why each supplier is risky	SHAP global ranking and supplier-level explanations identify delay, quality, dependency, invoice and disruption drivers
Actionability	Converts prediction into procurement action	High-risk alerts, mitigation priority and supplier comparison views

		guide review, monitoring and sourcing action
Operational relevance	Checks whether risk is stronger under dependency	High-dependency subgroup shows higher delay, defect and disruption exposure
Managerial usability scope	Frames dashboard value as design evidence, not survey evidence	Power BI usability is treated as a design objective; practitioner usability testing is identified as future validation

5.11 Expert Panel Design-Science Validation

The expert panel review with eight procurement and supply-chain professionals provided qualitative validation of the framework's decision-support utility. The review focused on whether SHAP explanations were understandable, whether dashboard outputs were traceable to supplier records, and whether the resulting alerts and mitigation priorities could support supplier review meetings. The review is not treated as a large-sample usability survey; instead, it serves as design-science evaluation evidence for artefact relevance, interpretability and managerial actionability [34].

Table 13A. Expert Panel Review Summary

Validation aspect	Panel focus	Evaluation outcome
Interpretability	Whether SHAP drivers explain supplier-risk classification	Experts considered feature-level explanations useful for supplier review discussions
Traceability	Whether dashboard outputs can be linked to ERP supplier records	Risk scores, categories and drivers were traceable to anonymized supplier-level data
Actionability	Whether outputs support mitigation decisions	Alerts and mitigation priorities were considered useful for monitoring, renegotiation and alternative sourcing
Decision-support fit	Whether the artefact fits procurement workflow	The framework was judged suitable for pilot use in supplier review and escalation routines

6. Discussion

The findings show that explainable machine learning using ERP procurement data can predict supplier risk more effectively than static scorecard logic. XGBoost produced the strongest results because supplier risk is shaped by non-linear interactions among delay, quality, invoice mismatch, disruption exposure and purchase dependency. Logistic Regression outperformed the scorecard baseline but remained weaker than the tree-based ensemble models. XGBoost offered the strongest balance of accuracy, recall, F1-score and ROC-AUC.

The feature-importance results are consistent with procurement practice. Delivery delay frequency was the strongest predictor because repeated delay is often the first observable signal of supplier instability. Lead-time variability was also important because unpredictable schedules make production planning difficult. Defect rate ranked third, showing that quality failure remains a major supplier risk dimension. Purchase dependency added a critical exposure layer because a problematic supplier becomes more dangerous when the buying firm relies heavily on it.

SHAP improved the interpretability of model results. Without SHAP, the framework would classify suppliers into risk categories without explaining the operational reason. SHAP showed whether risk was driven by quality failure, delivery instability, dependency, invoice mismatch or external disruption exposure [14], [16]. This matters because different causes require different managerial responses.

Power BI improved the practical usability of the framework by translating technical outputs into supplier risk scores, risk categories, trend views, alerts and mitigation priorities. This is important because procurement decisions are usually made by managers who need clear, filterable and interpretable information during supplier reviews and operational meetings [17], [24], [26].

The comparison with the traditional supplier scorecard is one of the key findings. The high-risk recall of the scorecard was 61.8%, compared with 86.5% for the explainable ML framework. This improvement indicates that fixed-weight approaches may miss important risky suppliers. The proposed framework improved detection and supplied explanations and dashboard-based action support.

7. Implications

7.1 Theoretical Implications

The study contributes to supplier risk and supply chain resilience literature by showing how digital procurement data can support risk sensing and early warning [1], [3], [9]. Dynamic Capability Theory is extended by mapping the framework to sensing, seizing and reconfiguring: ERP data supports risk sensing, XGBoost and SHAP support risk intelligence, and Power BI supports reconfiguration through supplier review and mitigation action [28], [29]. Information Processing Theory is also extended because the framework increases the organization's capacity to collect, process, interpret and communicate supplier risk information under uncertainty [30].

7.2 Methodological Implications

The study contributes methodologically by presenting an integrated XGBoost-SHAP-Power BI workflow. Many predictive analytics studies stop at model performance. This study connects model comparison, train-test validation, cross-validation, confidence intervals, McNemar testing, feature importance, SHAP interpretation, dependency subgroup analysis and dashboard design. This makes the method more appropriate for decision-support research.

7.3 Managerial Implications

The framework enables procurement managers to identify high-risk suppliers earlier. The dashboard allows managers to determine which suppliers require immediate intervention, which require closer monitoring and which are stable. SHAP-based explanations help managers identify whether risk is caused by delay, quality, dependency, invoice mismatch or disruption exposure.

7.4 Practical Implications for Manufacturing Firms

The framework can help manufacturing firms reduce production-delay exposure, improve supplier visibility and strengthen procurement resilience. It can support supplier review meetings, alternative sourcing, contract renegotiation, supplier development and risk escalation. It moves supplier evaluation from periodic review toward continuous supplier risk monitoring.

8. Conclusion

This study designed and evaluated an explainable machine-learning supplier risk prediction framework using real anonymized ERP procurement data, SHAP explainability and Power BI decision dashboards. XGBoost outperformed the traditional supplier scorecard, Logistic Regression and Random Forest. It achieved 89.7% accuracy and 0.928 ROC-AUC, while increasing high-risk supplier recall from 61.8% under the scorecard to 86.5%. The results show that supplier risk is shaped by delivery delay frequency, lead-time variability, defect rate, purchase dependency ratio and invoice mismatch frequency. SHAP made supplier-level risk classification interpretable, and Power BI converted model outputs into usable decision views. The study concludes that supplier risk prediction becomes more useful when predictive accuracy, explainability, statistical validation and dashboard-based action are integrated into one decision-support framework.

9. Limitations and Future Research

The study has limitations. First, the data are real and anonymized, but they come from one manufacturing firm; therefore, the findings should be interpreted as a single-firm empirical pilot and proof-of-application rather than as universal evidence for all manufacturing sectors. Second, the sample consists of 100 active suppliers, which represents the complete active supplier base after cleaning but should be expanded in future multi-firm studies. Third, the class distribution is imbalanced because high-risk suppliers account for 15% of the supplier base; class weighting, per-class metrics and bootstrap intervals reduce this concern, but larger samples would allow stronger minority-class validation. Fourth, supplier classification depends on ERP and procurement-history completeness; prediction quality may decrease where records are incomplete or inconsistent. Fifth, the dashboard was developed in Power BI; future research may compare Power BI with Tableau, Qlik or customized web dashboards. Sixth, the expert panel review provides qualitative design-science validation, but future studies should conduct larger practitioner usability tests, interviews or field experiments to evaluate dashboard adoption, perceived usefulness and decision impact. Seventh, future research may incorporate ESG supplier risk, logistics disruption data, geopolitical risk, real-time IoT signals, richer supplier-network features and larger multi-industry datasets. These limitations do not weaken the pilot evidence, but they define the scope within which the results should be interpreted.

REFERENCES

1. Ho, W., Zheng, T., Yildiz, H., & Talluri, S. (2015). Supply chain risk management: A literature review. International Journal of Production Research, 53(16), 5031–5069. https://doi.org/10.1080/00207543.2015.1030467
2. Tang, C. S. (2006). Perspectives in supply chain risk management. International Journal of Production Economics, 103(2), 451–488. https://doi.org/10.1016/j.ijpe.2005.12.006
3. Christopher, M., & Peck, H. (2004). Building the resilient supply chain. The International Journal of Logistics Management, 15(2), 1–14. https://doi.org/10.1108/09574090410700275
4. Jüttner, U. (2005). Supply chain risk management: Understanding the business requirements from a practitioner perspective. The International Journal of Logistics Management, 16(1), 120–141. https://doi.org/10.1108/09574090510617385
5. Ponomarov, S. Y., & Holcomb, M. C. (2009). Understanding the concept of supply chain resilience. The International Journal of Logistics Management, 20(1), 124–143. https://doi.org/10.1108/09574090910954873
6. Manuj, I., & Mentzer, J. T. (2008). Global supply chain risk management. Journal of Business Logistics, 29(1), 133–155. https://doi.org/10.1002/j.2158-1592.2008.tb00072.x
7. Wieland, A., & Wallenburg, C. M. (2013). The influence of relational competencies on supply chain resilience. International Journal of Physical Distribution & Logistics Management, 43(4), 300–320. https://doi.org/10.1108/IJPDLM-08-2012-0243
8. Ivanov, D., & Dolgui, A. (2020). Viability of intertwined supply networks: Extending the supply chain resilience angles towards survivability. International Journal of Production Research, 58(10), 2904–2915. https://doi.org/10.1080/00207543.2020.1750727
9. Baryannis, G., Validi, S., Dani, S., & Antoniou, G. (2019). Supply chain risk management and artificial intelligence: State of the art and future research directions. International Journal of Production Research, 57(7), 2179–2202. https://doi.org/10.1080/00207543.2018.1530476
10. Talluri, S., & Narasimhan, R. (2004). A methodology for strategic sourcing. European Journal of Operational Research, 154(1), 236–250. https://doi.org/10.1016/S0377-2217(02)00649-5
11. Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J., Dubey, R., & Childe, S. J. (2017). Big data analytics and firm performance. Journal of Business Research, 70, 356–365. https://doi.org/10.1016/j.jbusres.2016.08.009
12. Breiman, L. (2001). Random forests. Machine Learning, 45, 5–32. https://doi.org/10.1023/A:1010933404324
13. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. https://doi.org/10.1145/2939672.2939785
14. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, 4765–4774.
15. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why should I trust you?” Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135–1144. https://doi.org/10.1145/2939672.2939778
16. Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S. I. (2020). From local explanations to global understanding with explainable AI for trees. Nature Machine Intelligence, 2, 56–67. https://doi.org/10.1038/s42256-019-0138-9

17. Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. MIS Quarterly, 36(4), 1165–1188. https://doi.org/10.2307/41703503
18. Kull, T. J., & Talluri, S. (2008). A supply risk reduction model using integrated multicriteria decision making. IEEE Transactions on Engineering Management, 55(3), 409–419. https://doi.org/10.1109/TEM.2008.922627
19. Nabil, D. H. (2023). Managing supply chain performance using a real-time Microsoft Power BI dashboard. Cogent Engineering, 10(2), 2257924. https://doi.org/10.1080/23311916.2023.2257924
20. Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Applied Logistic Regression (3rd ed.). Wiley. https://doi.org/10.1002/9781118548387
21. Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. The Annals of Statistics, 29(5), 1189–1232.
22. Fawcett, T. (2006). An introduction to ROC analysis. Pattern Recognition Letters, 27(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010
23. Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLOS ONE, 10(3), e0118432. https://doi.org/10.1371/journal.pone.0118432
24. Arnott, D., & Pervan, G. (2005). A critical analysis of decision support systems research. Journal of Information Technology, 20, 67–87. https://doi.org/10.1057/palgrave.jit.2000035
25. Power, D. J. (2002). Decision Support Systems: Concepts and Resources for Managers. Quorum Books.
26. Davenport, T. H. (2006). Competing on analytics. Harvard Business Review, 84(1), 98–107.
27. Turban, E., Sharda, R., & Delen, D. (2014). Decision Support and Business Intelligence Systems (9th ed.). Pearson.
28. Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. Strategic Management Journal, 18(7), 509–533. https://doi.org/10.1002/(SICI)1097-0266(199708)18:7<509::AID-SMJ882>3.0.CO;2-Z
29. Teece, D. J. (2007). Explicating dynamic capabilities. Strategic Management Journal, 28(13), 1319–1350. https://doi.org/10.1002/smj.640
30. Galbraith, J. R. (1974). Organization design: An information processing view. Interfaces, 4(3), 28–36. https://doi.org/10.1287/inte.4.3.28
31. Ahmed, K. R., Ansari, M. E., Ahsan, M. N., Rohan, A., Uddin, M. B., & Rivin, M. A. H. (2025). Deep learning framework for interpretable supply chain forecasting using SOM ANN and SHAP. Scientific Reports, 15, 26355. https://doi.org/10.1038/s41598-025-11510-z
32. Wyrembek, M., Baryannis, G., & Brintrup, A. (2024). What if? Causal machine learning in supply chain risk management. arXiv:2408.13556.
33. Zheng, G., & Brintrup, A. (2024). Enhancing supply chain visibility with generative AI: An exploratory case study on relationship prediction in knowledge graphs. arXiv:2412.03390.
34. Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. MIS Quarterly, 28(1), 75-105. https://doi.org/10.2307/25148625