Research Article

A CatBoost-based integrated decision-making model for knowledge-oriented risk evaluation in multi-tier supply chains ⛓

Hamid Yousefi [1], Seyed Mohammad Seyed Hosseini [2], Elmira mashayekhinezamabad [3]

1. Department of Industrial Management, Qazvin Branch, Islamic Azad University, Qazvin, Iran
2. Department of Industrial engineering, University of science and technology, Tehran, Iran
3. Department of Industrial Management, Qazvin Branch, Islamic Azad University, Qazvin, Iran

Abstract

This study proposes a data-driven integrated framework based on knowledge management for risk evaluation in multi-tier supply chains. Traditional risk assessment models often rely on qualitative or expert-based methods, limiting their adaptability and accuracy in dynamic environments. The proposed framework combines operational risk indicators (supplier, production, and distribution) with knowledge management indicators (creation, sharing, application, retention, and market updating) to provide a holistic understanding of risk propagation. Using a dataset of 800 records and the CatBoost machine learning algorithm, the model achieved an accuracy of 0.86 and a weighted F1-score of 0.85. Sensitivity analysis showed that delivery delay (SR1), price volatility (SR2), and knowledge application (KM3) are the most influential factors, highlighting the interdependence between operational efficiency and knowledge utilization. The findings demonstrate that integrating knowledge management with data-driven analytics enhances prediction accuracy and strengthens organizational resilience. This research contributes a quantitative and explainable model for proactive risk management and provides managers with actionable insights to build adaptive, knowledge-oriented supply chain systems.

* Corresponding Author:
Seyed Mohammad Seyed Hosseini
Department of Industrial engineering, University of science and technology, Tehran, Iran
E-Mail: **seyedhosseini@iust.ac.ir**

1. Introduction

In today's highly interconnected global economy, supply chains have evolved into complex, multi-tier networks that span across multiple organizations, regions, and industries. While this interconnectedness has enabled higher efficiency and responsiveness, it has also exposed supply chains to various sources of vulnerability (Anozie et al., 2024). Disruptions caused by geopolitical conflicts, pandemics, natural disasters, or supplier bankruptcies can propagate rapidly across tiers, leading to cascading failures and severe financial losses. Consequently, supply chain risk management (SCRM) has emerged as a critical research and operational area, seeking methods to identify, evaluate, and mitigate risks before they materialize (Mashhadi et al., 2025; Zeynali et al., 2025).

Despite the growing attention, the traditional approaches to SCRM often rely on qualitative assessments or expert judgment, which limit their scalability and objectivity. Most conventional models assume linear and static risk relationships, failing to capture the dynamic interactions among multiple tiers of a supply network (GhanavatiNejad et al., 2025; Molaei et al., 2025). With the rapid growth of digitalization and data availability, organizations now have access to vast streams of operational, financial, and behavioral data that can provide a richer understanding of supply chain vulnerabilities. This has led to increasing interest in data-driven risk assessment frameworks that combine analytical modeling with predictive machine learning to improve decision accuracy and timeliness (Tavakoli et al., 2023).

At the same time, the role of knowledge management (KM) in supply chain resilience has become increasingly prominent. The ability to create, share, and apply knowledge across partners determines how effectively a network can sense and respond to disruptions. Knowledge flows between suppliers, manufacturers, and distributors are essential not only for learning from past events but also for preventing risk propagation in real time (Zighan et al., 2024). However, existing risk assessment models rarely integrate knowledge-related metrics—such as knowledge sharing, retention, or application—into quantitative, data-driven evaluation systems (Ganji et al., 2025). This represents a significant theoretical and practical gap in the current literature.

To bridge this gap, there is a pressing need for an integrated framework that simultaneously incorporates data-driven analytical capabilities and knowledge management processes for comprehensive multi-tier risk evaluation. Such a framework would enable organizations to quantify risk exposure while accounting for the moderating effect of knowledge-related factors, such as learning capability, information transparency, and collaborative problem-solving. This dual perspective could transform risk assessment from a reactive activity into a proactive, learning-based system capable of continuous improvement.

Therefore, the motivation of this study is to develop and validate a data-driven integrated framework based on knowledge management for risk evaluation in multi-tier supply chains. The proposed approach leverages machine learning techniques to predict and classify risk levels using both traditional supply chain indicators and knowledge-based metrics. By combining data analytics with organizational knowledge flows, this research aims to enhance the precision, interpretability, and adaptability of supply chain risk assessment. Ultimately, the study contributes to both academia and practice by offering a novel methodology that strengthens resilience and decision intelligence across complex, multi-tier supply networks.

2. Literature Review

This section provides a comprehensive review of the literature on supply chain risk assessment and presents a comparative analysis with the approach adopted in the current study. (Tuncel & Alpan, 2010) developed one of the early models that integrated risk assessment and management into supply chain (SC) network design using a timed Petri net framework. By combining Failure Mode, Effects and Criticality Analysis (FMECA) with Petri-net-based simulation, they demonstrated how risk mitigation actions can improve network performance and reduce costs in an industrial case study. While their research pioneered the modeling of dynamic and stochastic disruptions, it remained largely qualitative and simulation-oriented, lacking the ability to learn from data or incorporate organizational knowledge factors that influence risk propagation across multiple tiers. (Jianying et al., 2021) advanced the field by applying optimized back-propagation neural networks to evaluate risk in the fresh grape supply chain from a sustainability perspective. The authors designed and trained BP, GA-BP, and PSO-BP models on survey data to identify and prioritize critical risk factors. Their findings indicated that cooperation and economic risks were the most significant, and the PSO-BP model achieved the highest accuracy. However, the study focused on a single agricultural supply chain and relied mainly on static survey inputs, without incorporating inter-tier knowledge exchange or continuous learning mechanisms that could dynamically update the risk profile over time.

(Pandey et al., 2023) conducted a comprehensive review to identify emerging supply chain risks in the Industry 4.0 environment, emphasizing operational, behavioral, cyber-security, and environmental factors. Using the TOPSIS and DEMATEL techniques, the study ranked and categorized risks as cause-and-effect relationships to guide mitigation strategies. Although it offered a useful taxonomy and prioritization framework, the approach was purely multi-criteria decision-making (MCDM)-based and did not integrate predictive analytics or knowledge management constructs that could enhance the resilience of interconnected supply chain networks. Similarly, (Zhang et al., 2021) employed a firefly-algorithm-optimized support vector machine (FA-SVM) to assess credit risk in supply chain finance (SCF). The hybrid model improved classification accuracy compared with traditional SVM approaches and demonstrated the potential of machine learning (ML) for SCF risk prediction. Nevertheless, this research was limited to financial risk within a single-tier SCF context and did not address multi-tier operational dependencies or knowledge flows among partners. (Choudhary et al., 2023) performed a state-of-the-art bibliometric and network analysis of supply chain risk assessment (SCRA) studies, revealing three dominant research clusters and methodological trends. The paper highlighted that despite substantial progress in MCDM and

fuzzy modeling, few studies have adopted data-driven or machine-learning-based approaches for multi-tier risk evaluation. (Jiang et al., 2022), in contrast, proposed a blockchain-enabled trust-transitivity model for small and medium-sized manufacturing enterprises within supply chain finance. By applying intuitionistic fuzzy sets and dynamic weight allocation, they attempted to transfer "trust" from core firms to lower-tier suppliers, alleviating financing constraints. This work introduced a valuable dimension of inter-organizational knowledge transfer but remained confined to financial trust evaluation rather than operational or sustainability-related risks. (Yu & Ma, 2025) examined mega-infrastructure construction supply chains (MICSCs) under the EPC contracting model using an integrated fuzzy AHP–fuzzy DEMATEL framework to build a hierarchical risk evaluation system. Their results identified political situation, social security, and management mode as the most influential risk factors. Although the study effectively combined qualitative and quantitative analyses, it relied heavily on expert judgment, which limits generalizability and adaptability to real-time data environments. (Mihálcz & Kosztyán, 2024) subsequently proposed REFS-A, a practical risk evaluation framework tested in an automotive electronics manufacturing case. Their work contributed by validating multiple aggregation methods and aligning risk evaluation with expert FMEA feedback. However, like prior MCDM-based approaches, it lacked integration with data-driven predictive models and omitted consideration of knowledge creation, sharing, and application processes that shape a firm's ability to respond to risk.

Finally, (Moktadir et al., 2025) conducted an extensive review of multi-criteria decision-making (MCDM) applications in sustainable supply chain risk management, analyzing 101 papers published between 2010 and 2025. The authors concluded that most studies apply single or dual MCDM techniques and recommended future work combining MCDM with optimization, simulation, or machine learning to better capture the multidimensional nature of sustainability-related risks. Their review underscored the persistent fragmentation of the field, where few models integrate dynamic data analytics with learning and knowledge mechanisms across multiple supply chain tiers.

Collectively, the reviewed literature demonstrates that while research on supply chain risk assessment has evolved from qualitative modeling to MCDM and early machine-learning applications, three major gaps remain. First, most studies analyze risk at a single-tier level and ignore the interdependencies and cascading effects among suppliers, manufacturers, and distributors. Second, despite the acknowledged importance of organizational learning and information exchange, the explicit integration of knowledge management constructs—such as knowledge creation, sharing, retention, and application—into quantitative, data-driven risk frameworks is still missing. Third, although machine-learning techniques have been applied sporadically, their combination with KM indicators for predictive and explanatory purposes has not been explored in the context of multi-tier supply chains. Addressing these gaps, the present study proposes a data-driven integrated framework based on knowledge management for risk evaluation in multi-tier supply chains.

Unlike previous research, the proposed framework simultaneously considers all stages of the supply chain—supplier, production, and distribution—and quantifies how knowledge-related variables moderate risk propagation across tiers. By leveraging advanced machine-learning algorithms to identify, classify, and predict supply chain risks while embedding knowledge-management indicators within the dataset, this study aims to produce a more holistic, adaptive, and interpretable risk-evaluation model that enhances both theoretical understanding and managerial decision-making.

3. Methodology
3.1. Overview

This study adopts a data-driven methodological approach to develop and validate an integrated framework for supply chain risk evaluation based on knowledge management principles. Unlike traditional qualitative or expert-based risk analysis, the current research leverages empirical data and advanced machine learning (ML) algorithms to quantify, predict, and interpret risk levels across multiple tiers of the supply chain. The approach enables the dynamic identification of risk patterns and their interdependencies with knowledge-related factors, offering both predictive accuracy and managerial interpretability. The methodology is designed to capture complex, non-linear interactions between operational, financial, and knowledge-based indicators within a multi-tier supply network.

3.2. Research process and analytical steps

The research process followed a structured sequence of data-driven analytical steps, as illustrated in Figure 1. Each step contributed to constructing, training, and validating the predictive model for multi-tier supply chain risk evaluation.
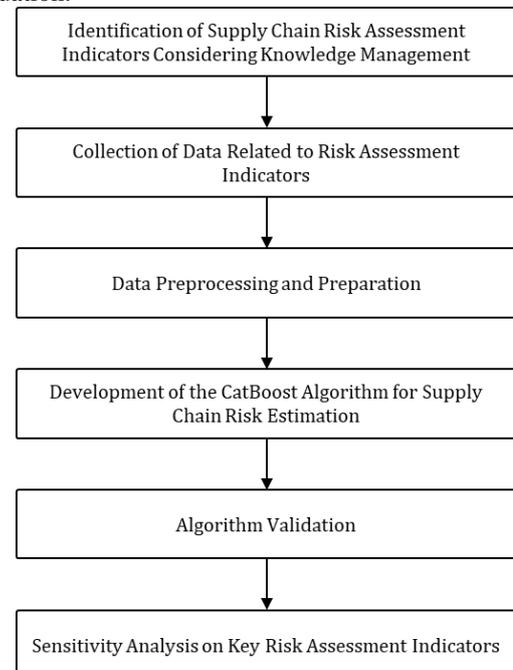

Fig. 1. Part of the structure of the gradient boosting model on the study dataset

Step 1: Data Collection and Variable Definition
An empirical dataset consisting of 800 records was developed to represent nodes within a multi-tier supply chain, including suppliers, manufacturers, and distributors.

Each record contained a set of risk-related indicators (SR, PR, DR) and knowledge-management indicators (KM) identified from the literature review.

Supply chain risk indicators covered supplier risk (e.g., delay, price volatility, quality), production risk (e.g., planning errors, capacity flexibility), and distribution risk (e.g., demand fluctuation, transportation disruption).

Knowledge management indicators reflected organizational learning processes, including knowledge creation, sharing, retention, accessibility, and market knowledge updating.

Each variable was numerically quantified or normalized within the range [0,1] to ensure model comparability and machine readability.

Step 2: Data Preprocessing and Cleaning

The raw dataset was examined for missing values, outliers, and inconsistencies. Missing numerical entries were replaced with median values, and categorical variables were encoded into numerical form. The final dataset included 22 independent variables (features) and one target variable (risk_level) classified into three categories: Low, Medium, and High. A stratified 80/20 train–test split was applied to maintain proportional class representation.

Step 3: Model Selection and Training

A supervised machine learning technique, CatBoost Classifier, was selected due to its capability to handle heterogeneous data, minimize overfitting, and provide interpretable feature importance. The model was trained on the training subset with default hyperparameters initially and optimized using early stopping based on validation accuracy to prevent unnecessary iterations.

The use of CatBoost enabled efficient handling of non-linear relationships and inter-tier dependencies, while simultaneously maintaining robustness to noise and small sample size variations.

Step 4: Model Evaluation and Validation

Model performance was evaluated using standard classification metrics, including Accuracy, Precision, Recall, F1-score, and the Confusion Matrix.

Step 5: Feature Importance and Interpretability Analysis

Feature importance was extracted from the CatBoost model to identify the most influential variables affecting supply chain risk. The top-ranked features included SR1 (delivery delay), SR2 (price volatility), PR1 (planning error rate), and KM2 (knowledge sharing index), indicating that both operational disruptions and knowledge-related factors significantly contribute to overall risk. Visual representation through a bar chart highlighted the relative importance of each indicator, supporting interpretability and managerial decision-making.

3.3. Summary of methodological advantages

The proposed methodology differs from conventional supply chain risk studies in three major aspects:

- It considers all tiers of the supply chain simultaneously, capturing cross-level interactions between suppliers, manufacturers, and distributors.
- It applies a machine-learning-driven approach rather than rule-based or expert judgment systems, enhancing objectivity and adaptability (Nessari et al., 2024; ZhianVamarzani et al., 2025).
- It explicitly incorporates knowledge management variables into the predictive model, demonstrating how knowledge creation, sharing, and application can mitigate or amplify risk propagation.

4. Result

This section reports the findings of the research. The evaluation indicators are first identified and defined in detail, after which the dataset is constructed based on the selected indicators. The collected data are then preprocessed, and the proposed algorithm is developed and validated accordingly.

4.1. Identification and definition of evaluation indicators

In this section, the key indicators used for risk evaluation in the multi-tier supply chain are identified and defined. Based on a comprehensive review of the literature and expert validation, the indicators are divided into two main groups: Supply Chain Risk Indicators and Knowledge Management Indicators. The first group captures the operational, environmental, and market-related risks that affect the supply chain, while the second group represents organizational knowledge processes that influence the network's capability to sense, respond, and adapt to risks. These indicators serve as the input variables in the proposed data-driven framework, providing a quantitative foundation for the subsequent machine learning–based risk estimation. Tables 1–4 summarize the indicators, their conceptual definitions, and key references supporting their inclusion in the framework.

Table 1

Supplier Risk Indicators

Code	Indicator	Definition
SR1	Delivery Delay	Average difference between actual delivery time and the agreed schedule
SR2	Price Volatility	Degree of fluctuation in raw material prices over a specific period
SR3	Supplied Item Quality	Percentage of defective or returned items supplied
SR4	Supplier Dependency	Proportion of purchases sourced from one or two key suppliers
SR5	Transportation Risk in Supply	Likelihood of damage or delay during inbound logistics
SR6	Political or Sanction Risk	Exposure of suppliers to political instability or trade sanctions

Table 2

ProductionRisk Indicators

PR1	**Production Planning Error**	**Deviation between production plan and actual output**
PR2	Capacity Flexibility	Ability to increase or decrease production volume during disruptions
PR3	Technological Dependency	Sensitivity of production performance to specific technology or software
PR4	Skilled Labor Shortage	Ratio of skilled workers to total production workforce
PR5	Internal Quality Control Weakness	Frequency of quality control failures during production stages

Table 3
Distribution and Market Risk Indicators

DR1	**Customer Demand Volatility**	**Variability in customer order volumes over time**
DR2	Transportation and Distribution Risk	Likelihood of delay or product damage during outbound logistics
DR3	Trade Restriction or Sanction Risk	Presence of legal or tariff barriers affecting imports/exports
DR4	Customer Satisfaction	Level of customer satisfaction or complaint rate

Table 4
Knowledge Management Indicators

Code	Indicator	Definition
KM1	Knowledge Creation	Degree of innovation and idea generation within the organization
KM2	Knowledge Sharing	Extent of information exchange between departments or partners
KM3	Knowledge Application	Percentage of decisions made based on prior knowledge and data
KM4	Knowledge Retention	Existence of structured databases for storing and retrieving knowledge
KM5	Knowledge Accessibility	Ease of access to organizational knowledge repositories
KM6	Market Knowledge Updating	Organization's ability to collect and analyze new market data

4.2. Model Development

The development of the predictive model for supply chain risk evaluation was carried out in several stages, including data preprocessing, correlation analysis, dataset partitioning, and algorithm tuning. Each stage ensured that the dataset was properly structured, balanced, and optimized for training the CatBoost model, which forms the core of the proposed data-driven framework.

Data Preprocessing

Before model training, the dataset underwent several preprocessing operations to ensure consistency and suitability for machine learning analysis. Missing values were first identified and replaced with the median of each numerical feature, minimizing potential bias caused by outliers. All continuous variables were normalized within the range [0,1] to standardize the scale and avoid domination of high-magnitude features during learning. Categorical variables, such as the risk level label (Low, Medium, High), were encoded using integer representation. Subsequently, data cleaning was performed to remove redundant attributes and verify the consistency of values across all 800 observations.

After preprocessing, a total of 22 quantitative independent variables (representing supplier, production, distribution, and knowledge management risks) and one categorical target variable were retained for modeling.

Correlation Analysis

To assess multicollinearity among features and understand potential dependencies, the Pearson correlation matrix was calculated for all numerical variables. The resulting heatmap (Figure 2) demonstrates that most correlation coefficients lie within the range of –0.1 to 0.1, indicating weak inter-feature relationships and confirming that the dataset is well-suited for independent variable modeling.

This low level of correlation minimizes the risk of multicollinearity, ensuring that each feature contributes distinct information to the CatBoost algorithm.
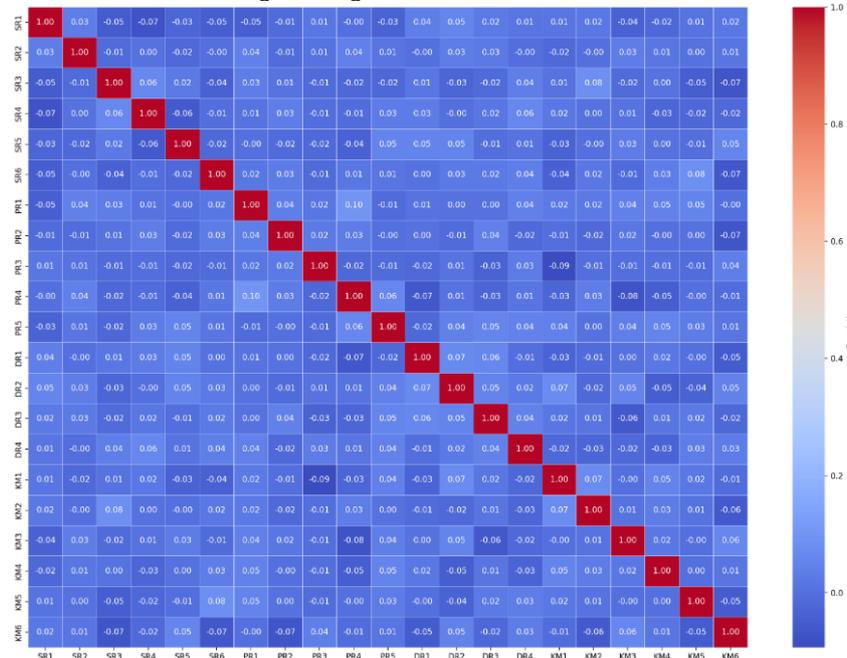

Fig. 2. Heatmap of feature correlations

The independence of features also validates the multi-dimensional design of the dataset, which integrates both operational (SR, PR, DR) and knowledge-based (KM) indicators.

Training and Test Data Partitioning

The preprocessed dataset was divided into two subsets using stratified random sampling to preserve the proportional distribution of risk classes.

- Training set: 80% of the data (640 records) used for model training and internal validation.

- Test set: 20% of the data (160 records) reserved for out-of-sample performance evaluation.

The stratification ensured that each of the three target classes—Low, Medium, and High risk—was equally represented across the training and test partitions.

Hyperparameter Tuning

Table 5. Final Parameters of the CatBoost Model

Parameter	Description	Selected Value
iterations	Maximum number of boosting rounds	500 (stopped at 212 via early stopping)
learning_rate	Step size controlling gradient updates	0.1
depth	Maximum depth of each decision tree	6
l2_leaf_reg	Regularization coefficient to prevent overfitting	3
loss_function	Objective function for classification	MultiClass
eval_metric	Evaluation metric during training	TotalF1
random_state	Reproducibility control	42
early_stopping_rounds	Number of rounds to stop training when validation score plateaus	50
train_test_split	Ratio between training and testing data	80:20

These parameters provided the most stable and accurate results, achieving convergence after 212 trees, where the model reached its optimal validation performance.

Model Structure

The CatBoost model is an ensemble of multiple oblivious decision trees (symmetric binary trees) constructed sequentially to minimize classification error. Each tree splits the dataset based on the most informative feature at each node, capturing complex, non-linear relationships among supply chain indicators.

A simplified schematic representation of the CatBoost tree structure is shown in Figure 3, where each split corresponds to a decision threshold on a specific risk or knowledge indicator.
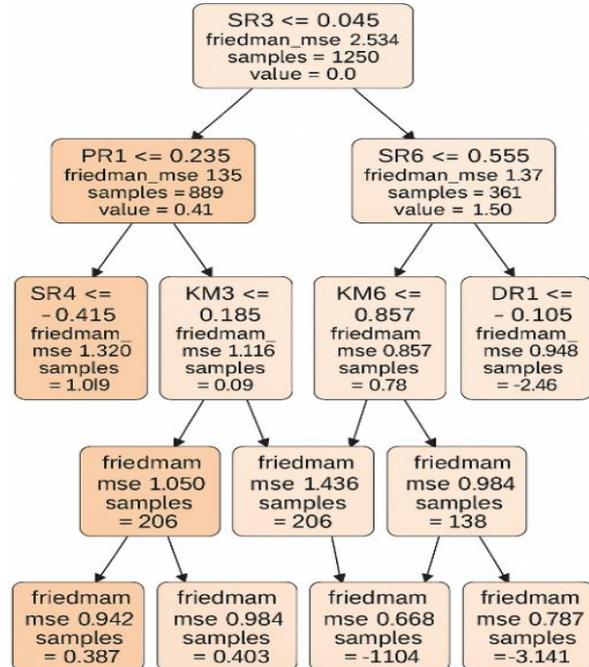

Fig. 3. Simplified structure of the catboost decision tree

4.3. Model validation and performance results

The CatBoost algorithm demonstrated strong predictive capability across the multi-tier supply chain dataset.

The model achieved its optimal performance after 212 boosting iterations, with stable convergence and no signs of overfitting.

Once the dataset was prepared, model calibration was performed through iterative tuning of CatBoost hyperparameters. The objective was to optimize accuracy while preventing overfitting.

The final set of hyperparameters, obtained through early stopping and empirical validation, is summarized in Table 5.

Overall accuracy reached 0.86, and the weighted F1-score was 0.85, indicating balanced performance across classes. Table 6 presents a concise summary of the classification metrics obtained from the test data.

Table 6
Model Performance Metrics

Class	Precision	Recall	F1-score	Support
High	1.00	0.27	0.43	11
Low	0.92	0.65	0.76	34
Medium	0.85	0.98	0.91	115
Overall Accuracy	–	–	0.86	160
Macro Average	0.92	0.63	0.70	–
Weighted Average	0.87	0.86	0.85	–

The confusion matrix in Figure 4 visualizes the classification distribution for the three risk levels.

The majority of observations were correctly classified as Medium Risk (113 out of 115), confirming the model's high reliability in identifying dominant risk patterns.

Minor misclassifications were observed between Low and Medium classes, while High Risk instances showed lower recall due to limited representation in the dataset.

Overall, the CatBoost model demonstrated robust predictive performance and effective differentiation across multi-tier supply chain risk categories.
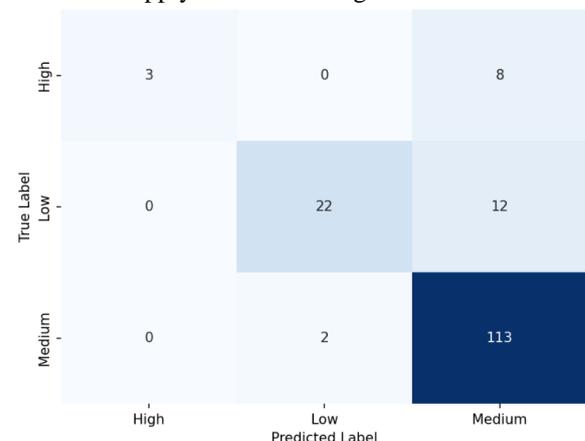

Fig. 4. Simplified structure of the catboost decision tree

4.4. Sensitivity analysis

The sensitivity analysis was conducted to determine the relative influence of each indicator on the predicted supply

chain risk levels. By applying the feature importance extraction function of the CatBoost model, the contribution of each variable to the overall prediction accuracy was quantified. This analysis provides valuable insight into how operational and knowledge-based indicators interact to shape risk outcomes across different tiers of the supply chain.

The results, illustrated in Figure 5, show that indicators related to supplier performance, price volatility, and knowledge utilization have the highest contribution to the

model's predictive power. Specifically, SR1_delivery_delay_days, SR2_price_volatility, and KM3_knowledge_application_ratio emerged as the top three drivers, collectively accounting for more than 49% of the total model importance. These variables capture the dual dimensions of operational disruption and organizational learning capability, reinforcing the central argument that effective knowledge management mitigates supply chain vulnerability.
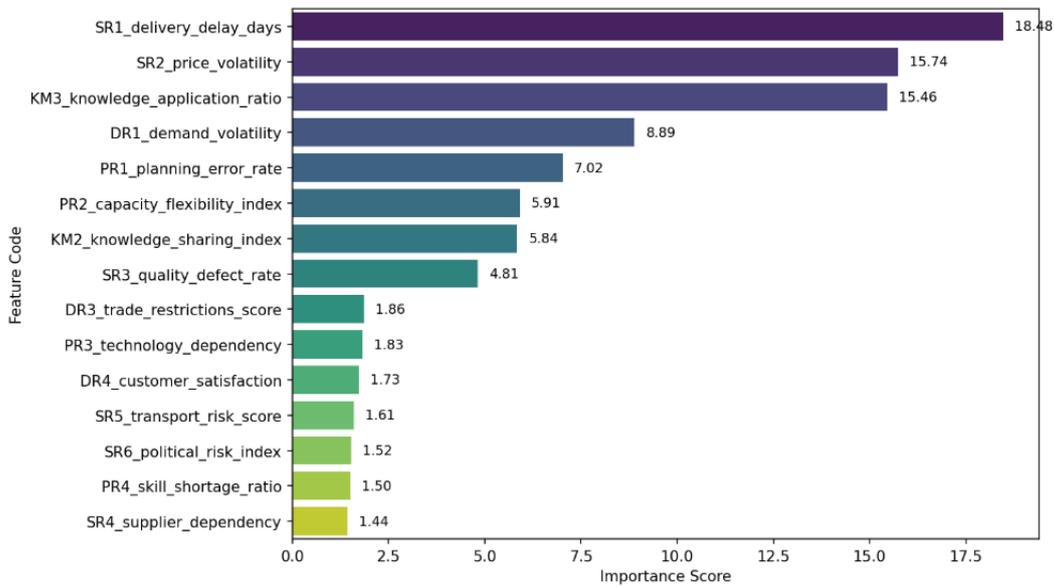

Fig. 5. Feature importance analysis of the catboost model

A closer examination of Table 7 reveals that the next influential indicators include DR1_demand_volatility, PR1_planning_error_rate, PR2_capacity_flexibility_index, and KM2_knowledge_sharing_index, representing both production flexibility and knowledge exchange mechanisms. These factors emphasize the role of collaborative learning and adaptive capacity in reducing supply chain risk propagation.

At the lower end of the ranking, indicators such as SR4_supplier_dependency, PR4_skill_shortage_ratio, and KM4_knowledge_retention_system show comparatively lower impact. Although less influential in the predictive model, these dimensions remain important from a managerial perspective, as they reflect the long-term resilience-building mechanisms in complex supply networks.

Table 7
Ranked Feature Importance from the CatBoost Model

Rank	Feature	Importance
1	SR1_delivery_delay_days	18.48
2	SR2_price_volatility	15.74
3	KM3_knowledge_application_ratio	15.46
4	DR1_demand_volatility	8.89
5	PR1_planning_error_rate	7.02
6	PR2_capacity_flexibility_index	5.91
7	KM2_knowledge_sharing_index	5.84
8	SR3_quality_defect_rate	4.81
9	DR3_trade_restrictions_score	1.86
10	PR3_technology_dependency	1.83
11	DR4_customer_satisfaction	1.73
12	SR5_transport_risk_score	1.61
13	SR6_political_risk_index	1.52
14	PR4_skill_shortage_ratio	1.50
15	SR4_supplier_dependency	1.44
16	PR5_internal_quality_failures	1.39
17	KM6_market_knowledge_update	1.31
18	KM1_knowledge_creation_rate	1.21
19	KM4_knowledge_retention_system	0.92
20	DR2_transport_distribution_risk	0.80

In summary, the sensitivity analysis confirms that operational inefficiencies and weak knowledge application are the dominant sources of supply chain risk. The findings highlight that data-driven models integrating knowledge management dimensions offer superior interpretability and actionable insights for risk mitigation and strategic planning in multi-tier supply chain environments.

5. Managerial insight

The findings of this study provide several practical insights for supply chain managers and decision-makers seeking to improve risk visibility, prediction accuracy, and overall resilience across multi-tier networks. The proposed data-driven and knowledge-based framework not only quantifies risk but also demonstrates how knowledge management practices can actively moderate and reduce operational vulnerabilities.

First, the results highlight that delivery delays, price volatility, and limited application of organizational knowledge are the most critical drivers of supply chain risk. This indicates that managers should prioritize the development of real-time monitoring systems and supplier performance dashboards that track delivery reliability and cost fluctuations. Integrating these operational data streams into machine-learning models enables early warning of

potential disruptions and supports proactive decision-making rather than reactive mitigation.

Second, the significant contribution of knowledge-related indicators (KM2 and KM3) underscores the strategic importance of knowledge creation, sharing, and application within and across organizational boundaries. Firms should invest in structured knowledge management systems—such as digital repositories, inter-departmental learning platforms, and supplier collaboration portals—to ensure that experiences and solutions from previous disruptions are documented, shared, and leveraged. These mechanisms enhance organizational learning and accelerate the collective response to unforeseen events.

Third, the analysis confirms that production flexibility and adaptive capacity are essential buffers against risk propagation. Managers are encouraged to maintain modular production capabilities, diversify supplier portfolios, and adopt agile planning processes that can quickly adjust to demand fluctuations or supply shocks. Training programs that strengthen the technical competence of employees also play a crucial role in sustaining flexibility during periods of uncertainty.

Fourth, the integration of knowledge management into the machine-learning-based risk assessment allows for continuous improvement of the model through feedback loops. Managers can use model outputs—such as feature importance scores and predictive risk levels—to identify critical nodes within the supply chain where risk concentration occurs. Over time, as more data are accumulated, the predictive system evolves into an intelligent decision-support tool capable of recommending targeted interventions, such as renegotiating supplier contracts, improving transportation routes, or enhancing information transparency.

Finally, the proposed framework promotes a shift from reactive to proactive risk governance. By embedding knowledge processes within data-driven analytics, organizations can transform risk management into a dynamic learning mechanism that evolves with environmental and operational conditions. The integration of artificial intelligence and knowledge management thus provides a dual advantage: it strengthens the analytical capability to forecast risks while simultaneously improving the organization's internal resilience and learning agility.

In summary, the managerial implications of this research emphasize that achieving supply chain resilience requires both data intelligence and knowledge intelligence. Managers who combine predictive analytics with systematic knowledge sharing and application can better anticipate disruptions, respond faster to uncertainties, and sustain competitive advantage in an increasingly volatile global marketplace.

6. Conclusions

This study developed a data-driven integrated framework based on knowledge management for risk evaluation in multi-tier supply chains. The proposed approach combined traditional operational risk indicators with knowledge-based variables to capture both tangible and intangible drivers of supply chain vulnerability. By applying the CatBoost machine learning algorithm, the framework successfully modeled complex, non-linear relationships between risk dimensions and knowledge processes,

providing a predictive and interpretable tool for risk assessment.

The results demonstrated that the integration of knowledge management significantly enhances the explanatory power and robustness of data-driven risk models. Specifically, features related to delivery delays (SR1), price volatility (SR2), and knowledge application (KM3) emerged as the most influential indicators, highlighting that both operational performance and organizational learning jointly determine the overall risk profile of the supply chain. The model achieved a strong predictive accuracy of 0.86 and a weighted F1-score of 0.85, confirming its reliability in classifying different risk levels across suppliers, production, and distribution tiers.

The findings further indicate that effective knowledge management—through continuous creation, sharing, and application of insights—plays a crucial role in mitigating disruptions and strengthening resilience. When embedded into a data-driven analytical framework, knowledge processes act as amplifiers of adaptability and coordination, enabling supply chain partners to anticipate, respond to, and recover from risks more efficiently.

From a methodological standpoint, this research advances the existing literature by moving beyond conventional qualitative or MCDM-based approaches toward a machine-learning-oriented, quantitative risk evaluation framework. The study contributes both theoretically and practically: it demonstrates how knowledge management can be operationalized as a measurable component within predictive analytics, and it provides managers with a structured tool for real-time risk monitoring and decision support.

Nevertheless, certain limitations should be acknowledged. The dataset used for model development was simulated and may not fully represent the complexity of real-world, multi-tier supply chains. Future research should validate the framework with empirical data from specific industries and examine additional knowledge-related dimensions such as organizational culture, digital collaboration maturity, and innovation capacity. Moreover, integrating explainable AI techniques (e.g., SHAP analysis) could provide deeper managerial interpretability of the model's predictions.

In conclusion, the proposed framework offers a comprehensive and scalable solution for data-driven risk evaluation enhanced by knowledge management. By linking predictive analytics with organizational learning, it paves the way toward intelligent, adaptive, and resilient supply chain ecosystems capable of thriving in today's volatile and uncertain global environment.

References

Anozie, U. C., Adewumi, G., Obafunsho, O. E., Toromade, A. S., & Olaluwoye, O. S. (2024). Leveraging advanced technologies in Supply Chain Risk Management (SCRM) to mitigate healthcare disruptions: A comprehensive review. *World Journal of Advanced Research and Reviews*, 23(1), 1039–1045.

Choudhary, N. A., Singh, S., Schoenherr, T., & Ramkumar, M. (2023). Risk assessment in supply chains: a state-of-the-art review of methodologies and their applications. *Annals of Operations Research*, 322(2), 565–607.

Ganji, V. Y., Hozan, E., Babolhavaeji, P., Tajally, A., & GhanavatiNejad, M. (2025). A robust design of a circular supply chain network based on the resilience and responsiveness dimensions: A data-driven model. *Socio-Economic Planning Sciences*, 102294.

GhanavatiNejad, M., Tavakoli, M., Sheikhalishahi, M., Aydın, N., & Aria, S. S. (2025). An integrated smart framework for fast-moving consumer goods online market logistics: a digital twin framework. *Journal of Industrial and Production Engineering*, 1–17.

Jiang, R., Kang, Y., Liu, Y., Liang, Z., Duan, Y., Sun, Y., & Liu, J. (2022). A trust transitivity model of small and medium-sized manufacturing enterprises under blockchain-based supply chain finance. *International Journal of Production Economics*, *247*, 108469.

Jianying, F., Bianyu, Y., Xin, L., Dong, T., & Weisong, M. (2021). Evaluation on risks of sustainable supply chain based on optimized BP neural networks in fresh grape industry. *Computers and Electronics in Agriculture*, *183*, 105988.

Mashhadi, E., Tajally, A., Vatankhah Ghamsari, S. M., Ghanavati-Nejad, M., Moghdani, R., & Yahyapour-Ganji, V. (2025). A novel hybrid model for supplier selection problem based on the viability and industry 5.0 pillars: Application in the energy sector. *Environment, Development and Sustainability*, 1–35.

Mihálcz, I., & Kosztyán, Z. T. (2024). REFS-A Risk Evaluation Framework on Supply Chain. *Mathematics*, *12*(6), 841.

Moktadir, M. A., Paul, S. K., Bai, C., & Santibanez Gonzalez, E. D. R. (2025). The current and future states of MCDM methods in sustainable supply chain risk assessment. *Environment, Development and Sustainability*, *27*(3), 7435–7480.

Molaei, B. J., Ghanavati-Nejad, M., Tajally, A., & Sheikhalishahi, M. (2025). A novel stochastic machine learning approach for resilient-leagile supplier selection: a circular supply chain in the era of industry 4.0. *Soft Computing*, 1–22.

Nessari, S., Ghanavati-Nejad, M., Jolai, F., Bozorgi-Amiri, A., & Rajabizadeh, S. (2024). A data-driven decision-making approach for evaluating the projects according to resilience, circular economy and industry 4.0 dimension. *Engineering Applications of*

Artificial Intelligence, *134*, 108608.

Pandey, S., Singh, R. K., & Gunasekaran, A. (2023). Supply chain risks in Industry 4.0 environment: review and analysis framework. *Production Planning & Control*, *34*(13), 1275–1302.

Tavakoli, M., Tajally, A., Ghanavati-Nejad, M., & Jolai, F. (2023). A Markovian-based fuzzy decision-making approach for the customer-based sustainable-resilient supplier selection problem. *Soft Computing*, 1–32.

Tuncel, G., & Alpan, G. (2010). Risk assessment and management for supply chain networks: A case study. *Computers in Industry*, *61*(3), 250–259.

Yu, R., & Ma, L. (2025). Risk evaluation of mega infrastructure construction supply chain in engineering-procurement-construction projects: an integrated fuzzy AHP and fuzzy DEMATEL approach. *Engineering, Construction and Architectural Management*, *32*(5), 3217–3235.

Zeynali, F. R., Parvin, M., ForouzeshNejad, A. A., Jeyzanibrahimzade, E., Ghanavati-Nejad, M., & Tajally, A. (2025). A Heuristic-based Multi-Stage Machine Learning-based Model to Design a Sustainable, Resilient, and Agile Reverse Corn Supply Chain by considering Third-party Recycling. *Applied Soft Computing*, 113042.

Zhang, W., Wu, C., Li, Y., Wang, L., & Samui, P. (2021). Assessment of pile drivability using random forest regression and multivariate adaptive regression splines. *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, *15*(1), 27–40.

ZhianVamarzani, M., Hejri, S., YahyapourGanji, V., GhanavatiNejad, M., & KiaeiTonekaboni, Z. (2025). Evaluation of Energy Sector Providers for Enhanced Efficiency, Performance, and Sustainability Using the Digital, Lean, Agile, Resilient, and Green (DLARG) Framework: A Machine Learning Model. *Process Integration and Optimization for Sustainability*, 1–19.

Zighan, S., Dwaikat, N. Y., Alkalha, Z., & Abualqumboz, M. (2024). Knowledge management for supply chain resilience in pharmaceutical industry: evidence from the Middle East region. *The International Journal of Logistics Management*, *35*(4), 1142–1167.

Journal of Optimization in Industrial Engineering,Vol.18, Issue 2, Summer & Autumn 2025, 183-191
Hamid Yousefi & et al / A CatBoost-based integrated decision-making model for ...

191