

Research on Supply Chain Disruption Risk Prediction Model Based on Machine Learning

Ye Chen

*College of International Economics & Trade, Ningbo University of Finance and Economics,
Ningbo, China
3275502782@qq.com*

Abstract. Supply chain disruptions feature sudden occurrence, transmissibility and severe losses. Traditional risk management approaches relying on manual scoring and post-incident response fail to identify complex nonlinear signals in a timely manner. Targeting the demand for predictive decision-making in operations and supply chain management, this paper constructs a machine learning prediction framework consisting of data governance, feature engineering, model training, risk interpretation and response linkage, and compares Logistic regression, support vector machine, random forest and gradient boosting tree. The study demonstrates the model evaluation process using reproducible exemplary simulation data, and measures prediction quality comprehensively through accuracy, recall, F1-score and area under the receiver operating characteristic curve. The study holds that the value of machine learning does not lie in the mere pursuit of algorithmic accuracy, but in converting risk probabilities into operational actions such as inventory adjustment, multi-source procurement, transportation rerouting and supplier collaboration. The implementation of the model should be accompanied by the simultaneous establishment of data standards, threshold governance, interpretability mechanisms and rolling update systems. This framework can effectively integrate professional capabilities including supply chain management, risk statistical modeling, predictive analytics and AI-enhanced business analytics, providing actionable research insights for supply chain resilience building.

Keywords: supply chain disruption, machine learning, risk prediction, predictive analytics, supply chain resilience

1. Introduction

Global procurement networks, lean inventory and cross-border logistics have improved the efficiency of resource allocation, but also made local delays more prone to spread along multi-echelon supply networks. Research on digital and intelligent operations points out that supply chains are shifting from single-chain control to cross-entity data collaboration and ecological decision-making [1]. Against this backdrop, disruption risk management should not only answer "where problems have occurred", but also predict "which order, supplier or route is most likely to fail in the coming period, and what measures should be taken in advance". Machine learning can identify nonlinear correlations from heterogeneous data such as order fulfillment, delivery cycle, quality,

inventory, transportation and external shocks. However, if algorithms are detached from business objectives, they may produce accurate but unexecutable black boxes. Therefore, starting from management issues, this paper proposes a binary classification risk prediction model for daily operations, uses reproducible simulation samples to illustrate the training and evaluation process, and focuses on how model outputs can be embedded in procurement, planning, inventory and logistics decisions [1].

2. Research foundation and data system of supply chain disruption risk

2.1. Connotation and management objectives of disruption risk

Supply chain disruption refers to a state where supply, production, transportation or information flow deviates from the plan due to internal failures and external shocks, leading to stockouts, production line stoppages, delayed delivery or abnormal cost increases. It differs from general fluctuations: general fluctuations can be absorbed by safety stock and regular scheduling, while disruptions often break through established buffers and trigger cascading effects across the network. The management objectives of the prediction model should be divided into three levels: first, identify high-risk orders and nodes before events occur to extend the response lead time; second, estimate risk probability and potential impact to help managers allocate limited resources according to severity; third, connect early warning results to alternative procurement, capacity rearrangement and logistics rerouting. Xu Shuang et al. applied Bayesian decision trees to assess agricultural supply chain risks, showing that probabilistic inference and tree models can conduct risk grading in combination with business indicators [2]. Accordingly, this paper sets "whether an event causing over three days of delay in promised delivery occurs within the next seven days" as the prediction label, endowing the model with a clear time window, object and consequence boundary [2].

2.2. Definition of prediction tasks and indicators

The model treats each pending fulfillment order as an observation unit, aggregates the status of its supplier, material, inventory and in-transit transportation at the prediction time point, and outputs the disruption probability p . A label y of 1 indicates a major delay within the observation window, and 0 indicates completion as scheduled or within tolerance. This definition can be converted into a binary classification task of supervised learning, and can also be directly aligned with enterprise service level indicators. This study does not take overall accuracy as the sole objective, since disruption samples usually account for a small proportion. A model that classifies all orders as normal may achieve high accuracy but delivers no management value. This paper focuses more on recall, i.e., the proportion of actual disruptions identified in advance, while taking precision into account to control false alarm costs. The threshold is jointly determined by production line stoppage losses, expedited procurement costs and early warning processing capacity. The basic expression of risk probability is as follows, where $f(x_i)$ represents the comprehensive score generated by the model based on the order feature vector. The higher the probability, the higher the priority of the order to enter enhanced monitoring and emergency decision-making [3].

$$p_i = P(y_i = 1|x_i) = \frac{1}{1+e^{-f(x_i)}} \quad (1)$$

2.3. Data sources and feature system

High-quality features should cover three categories of information: "node capability – operational status – external environment". Node capability includes suppliers' historical on-time rate, quality defect rate, capacity load, financial stability and number of alternative sources; operational status includes order fulfillment rate, coefficient of variation of delivery cycle, in-transit dwell time, inventory available days, demand deviation and frequency of plan changes; external environment covers extreme weather, port congestion, geopolitical conflicts, and abnormal exchange rates and raw material prices. Research on machine learning for supply chain risk shows that multi-source data can improve identification speed, but data availability, class imbalance and feature interpretation remain major obstacles [3]. For enterprise implementation, ERP, supplier relationship management, warehouse management and transportation management systems should serve as main data sources, supplemented by public event and weather data. All fields require unified timestamps, supplier codes and order granularity, and data owners should be designated to prevent inconsistent calibers of the same indicator across procurement, logistics and finance departments [4].

2.4. Model research framework

First, extract data on orders, suppliers, inventory, transportation and external events, and perform missing value processing, anomaly identification and time alignment. Second, convert raw data into interpretable features such as delay rate, cycle fluctuation, inventory coverage and shock index. Third, divide the dataset into training set and test set, and compare the prediction performance of linear, kernel and tree models. Fourth, calibrate the output of the optimal model into risk probabilities, and set early warning thresholds according to the enterprise's loss function. Finally, push high-risk orders to procurement, planning and logistics positions, record disposal actions and actual results, and generate new samples to support rolling training. This process corresponds to the interdisciplinary capabilities of predictive analytics, risk statistical modeling, process modeling and supply chain management, emphasizing that the model is only one module in the decision-making system. Without follow-up responsible persons, response time limits and feedback fields, even a high AUC can hardly truly reduce disruption losses [5].

3. Construction and evaluation of machine learning prediction model

3.1. Data preprocessing and sample division

Before modeling, features are first frozen at the prediction time point to avoid data leakage caused by using fields generated after the event occurs. For example, the actual arrival date cannot be used to predict whether the arrival will be delayed, and post-event expediting costs cannot be used as ex-ante risk features. Continuous variables can be imputed with the median and standardized, and categorical variables can be one-hot encoded. For long-tailed outliers, quantile winsorization is adopted, while extreme events that truly represent risks should be retained. Samples are divided into training set and test set by chronological order or order batches. If random division is adopted, it should be ensured that the same order and its derived records do not appear in both sets simultaneously. In the example of this paper, 2000 operational records are generated and split into training set and test set at a 7:3 ratio, with the random seed fixed to ensure result reproducibility. For the problem of insufficient disruption samples, class weights or stratified sampling can be applied

during the training phase, while the test set maintains the original proportion, so as to more realistically evaluate false alarms and missed alarms after the model goes live.

3.2. Candidate algorithms and ensemble logic

Candidate models balance benchmarking, nonlinear processing capability and interpretability. Logistic regression has a clear structure and its coefficient directions are intuitive for managers to understand, serving as the benchmark for judging whether complex models deliver genuine gains. Support vector machines are suitable for processing small-to-medium-scale, high-dimensional and nonlinear boundaries, but probability interpretation requires additional calibration. Random forest reduces single-tree fluctuation through voting by multiple decision trees and can output feature importance. Gradient boosting tree fits the residuals of the previous round iteratively, and usually captures well the interaction between supply delays, external shocks and inventory buffers. Yi Jingtao and Yan Huan combined ARIMA and GRU for foreign trade risk prediction, demonstrating that hybrid models can handle non-stationary and highly volatile sequences [4]. However, this paper prioritizes four types of models with relatively simple structures, because enterprise sample size, maintenance cost and interpretation requirements are often more important than pursuing minor accuracy improvements from deep networks. The training objective adopts binary cross-entropy loss to penalize the deviation between predicted probabilities and true labels.

$$L = -(1/n) \sum_{i=1}^n [y_i \ln p_i + (1 - y_i) \ln(1 - p_i)] \quad (2)$$

3.3. Model training and threshold optimization

During the training phase, cross-validation should be conducted within the training set to search for hyperparameters such as number of trees, maximum depth, learning rate and regularization intensity within a limited range, and the test set should be reserved for final evaluation to avoid over-optimistic results from repeated testing. After the probability model generates outputs, the calibration curve should also be checked: if only 30% of orders with a predicted probability of 0.70 actually experience disruption, managers will overestimate risks and excessively occupy buffer resources. The threshold does not need to be fixed at 0.50, but should be determined based on business losses. Let the expected loss of one missed disruption be C_{FN} , and the disposal cost of one false alarm be C_{FP} . When production line stoppage losses are much higher than manual verification costs, enterprises can moderately lower the threshold to improve recall; when the early warning team has limited capacity, the threshold should be raised and only high-impact orders should be addressed. Camur et al. compared multiple regression algorithms based on real inbound logistics data and found that tree models have good generalization ability in predicting product availability dates [5], indicating that model selection should be determined by out-of-sample performance rather than algorithm popularity.

3.4. Evaluation indicators and interpretability mechanism

Model evaluation considers discrimination ability, classification effect and management usability simultaneously. AUC measures the overall ability of the model to distinguish high-risk and low-risk orders under various thresholds. Precision P represents the proportion of actual disruptions among warned orders, reflecting whether verification resources are effectively utilized. Recall R represents the proportion of actual disruptions detected in advance, which is directly associated with missed

alarm losses. F1-score is the harmonic mean of P and R , suitable for class imbalance scenarios. In addition, confusion matrix, probability calibration error and stability across different supplier tiers should also be reported. For interpretability, linear models can use coefficient directions, and tree models can use permutation importance or SHAP values to explain why a given order is high-risk. For example, "rising supplier delay rate in the past 30 days" and "inventory coverage below transportation cycle" jointly drive up the risk. Interpretation results should be translated into business language and audit trails should be kept, which not only facilitates review by procurement managers, but also helps identify erroneous data and unreasonable rules, preventing black-box scores from directly replacing professional judgment.

$$F_1 = \frac{2PR}{P+R} \quad (3)$$

4. Model application, management value and implementation suggestions

4.1. Exemplary results and model comparison

To illustrate the evaluation method, this paper generates 2000 exemplary order records with a fixed random seed. Features include supplier delay rate, order fulfillment rate, delivery cycle fluctuation, inventory days, transportation on-time rate, demand fluctuation, external shock and quality problem rate; disruption labels are generated through a transparent nonlinear probability process. Table 1 shows that different algorithms present trade-offs among accuracy, precision, recall, F1-score and AUC. Linear models are easy to interpret but have limited ability to capture variable interactions; kernel methods and tree models usually improve discrimination ability. Model selection should not mechanically pick the one with the highest single indicator, but should integrate early warning capacity and loss structure: if the cost of production line stoppage for key materials is high, recall should be prioritized; if disposal costs are high, precision needs to be improved. The exemplary results serve to verify the process and table specifications, rather than claiming to represent real enterprise performance. Formal research should import enterprise data for more than 12 consecutive months and conduct cross-time, cross-supplier and cross-category external validation.

Table 1. Comparison of model performance under exemplary simulation data

Model	Accuracy	Precision	Recall	F1-score	AUC
Logistic Regression	0.817	0.287	0.837	0.427	0.887
Support Vector Machine	0.925	0.600	0.245	0.348	0.876
Random Forest	0.937	0.677	0.429	0.525	0.899
Gradient Boosting Tree	0.923	0.552	0.327	0.410	0.892

Note: The data are 2000 exemplary operational records generated with a fixed random seed, with disruption samples accounting for 8.2%. The results are only used to illustrate the modeling process and do not replace verification with real enterprise data.

4.2. Early warning grading and decision linkage

Risk probabilities only generate value when converted into action rules. Enterprises can classify orders into three levels: low, medium and high. Low-risk orders run according to conventional plans and are automatically documented. Medium-risk orders are verified by procurement or logistics personnel within 24 hours regarding supplier capacity, in-transit nodes and inventory coverage, with

delivery commitments adjusted if necessary. High-risk orders immediately enter cross-departmental disposal, where planning, procurement, production and logistics jointly evaluate alternative supply, emergency transportation, order splitting and capacity rearrangement. Grading thresholds should be set according to material criticality; the same probability of 0.65 carries different management implications for general auxiliary materials and exclusive core components. In addition to displaying probabilities, the system interface should also list the top three risk drivers, estimated impact amount and recommended actions, enabling managers to make quick judgments. After disposal, record whether suggestions are adopted, response time, costs and final outcomes, so as to evaluate "whether the model reduces losses" rather than just "whether the model predicts accurately", thus forming a closed loop of prediction, response, review and retraining [6].

4.3. Supply chain resilience strategies and organizational guarantee

Prediction models should be designed in conjunction with resilience strategies. For suppliers repeatedly identified as high-risk, enterprises can promote dual-source or multi-source procurement, sign flexible capacity agreements, increase strategic inventory of key raw materials, and improve quality and delivery through supplier development programs. For port congestion and transportation delay risks, alternative ports, routes and carrier combinations can be preset. For demand shocks, rolling forecasts and sales and operations planning can be adopted to coordinate inventory and capacity. Research on technology strategic supply chains emphasizes that resilience design, risk management and control, and collaborative optimization need to be considered holistically [6]. At the organizational level, it is recommended to establish a data and model governance team led by the supply chain director, clarifying that the procurement department provides supplier data, the planning department maintains demand and inventory calibers, the logistics department feeds back transportation events, the information department ensures interfaces and access permissions, and risk management staff are responsible for thresholds and audits. Monitor model drift monthly, review thresholds and features quarterly, and assess improvements in stockout rate, expediting costs and on-time delivery brought by the model semi-annually, to avoid the system being left unmaintained after launch.

5. Conclusions

Supply chain disruption risk prediction is not an isolated algorithm competition, but a data-driven management process targeting operational continuity. Starting from risk definition, data system, feature engineering, candidate algorithms, threshold optimization, interpretability mechanism and response process, this paper constructs a machine learning framework that can be implemented step by step by enterprises, and illustrates the multi-model comparison method with exemplary simulation data. The study shows that tree models are suitable for identifying complex interactions, while linear models can serve as transparent benchmarks. Practical selection should strike a balance among recall, false alarm cost, interpretability and maintainability. More importantly, risk probabilities must be linked with inventory, procurement, capacity and transportation decisions, and receive continuous feedback through disposal results. This study embeds machine learning methods into the context of supply chain management and risk governance, reflects the integrated application of predictive analytics, statistical modeling and operational decision-making, and also lays a solid foundation for subsequent empirical research using real enterprise data.

References

- [1] Chen, J., & Liu, Y. H. (2021). Digital intelligence enables operational management transformation: From supply chain to supply chain ecosystem. *Management World*, 37(11), 227–240.
- [2] Xu, S., Cai, H. M., Zhao, L. C., et al. (2024). A Bayesian decision tree algorithm model for risk prediction and assessment in the agricultural product supply chain. *Journal of Southwest University (Natural Science Edition)*, 46(3), 189–200.
- [3] Yang, M., Lim, M. K., Qu, Y., et al. (2023). Supply chain risk management with machine learning technology: A literature review and future research directions. *Computers & Industrial Engineering*, 175, 108859. DOI:10.1016/j.cie.2022.108859.
- [4] Yi, J. T., & Yan, H. (2023). Research on foreign trade risk prediction and early warning based on wavelet decomposition and ARIMA-GRU hybrid model. *Chinese Management Science*, 31(6), 100–110. DOI:10.16381/j.cnki.issn1003-207x.2021.1174.
- [5] Camur, M. C., Ravi, S. K., & Saleh, S. (2024). Enhancing supply chain resilience: A machine learning approach for predicting product availability dates under disruption. *Expert Systems with Applications*, 247, 123226. DOI:10.1016/j.eswa.2024.123226.
- [6] Yang, S. L., Li, X. J., Mo, H. J., et al. (2025). Basic characteristics and key scientific issues of technology strategic supply chains. *China Management Science*, 33(1), 1–13. DOI:10.16381/j.cnki.issn1003-207x.2025.01.001.