GAS Journal of Economics and Business Management (GASJEBM)
Volume 3 | Issue 6, 2026 Homepage: https://gaspublishers.com/gasjebm-home/ ISSN: 3048-782X

Research on the Evaluation of Third-Party Logistics Based on Random Forest Methods

XiaJie Chen

College of Economics and Management, Southwest Petroleum University, Chengdu 610500, China

Received: *05.05.2026* | ***Accepted:*** *06.06.2026* | ***Published:*** *06.06.2026*

***Corresponding Author:** XiaJie Chen

DOI: *10.5281/zenodo.20569546*

Abstract **Original Research Article**

With the acceleration of global economic integration and increasing supply chain complexity, comprehensive capability assessment of third-party logistics service providers has become a core component in optimizing supply chain management. Addressing limitations of traditional evaluation methods—including high subjectivity, unreasonable indicator weighting, and difficulty handling nonlinear relationships—we propose a third-party logistics evaluation model based on the random forest algorithm. By integrating three key dimensions—functional indicators, operational metrics, and stability indicators—we establish a comprehensive evaluation framework and detail the specific application process and optimization strategies for the random forest model in logistics assessment.

Keywords: Random Forests; Third-party Logistics; Evaluation Metrics; Machine Learning.

1. Introduction

The core challenge in evaluating third-party logistics lies in the multidimensional, complex, and nonlinear nature of its evaluation indicators. On one hand, logistics services encompass multiple stages—including transportation, warehousing, distribution, and information management—with each stage comprising numerous sub-indicators; on the other hand, these indicators exhibit intricate interrelationships that traditional linear models struggle to accurately capture. With the advancement of digitalization in the logistics sector, enterprises have accumulated vast amounts of operational data. Extracting valuable insights from this data and establishing a scientific, objective evaluation system has become a critical issue for the development of the third-party logistics industry.

The advancement of machine learning algorithms, particularly random forest algorithms, has provided novel approaches to addressing the aforementioned challenges. Random forest is an ensemble learning algorithm that enhances model accuracy and stability by constructing multiple decision trees and combining their predictions. This study aims to develop a third-party logistics evaluation model based on the random forest algorithm, establishing a well-designed indicator system to enable objective and precise assessment of service quality. This will assist logistics demand-side entities in making informed partner selection decisions and guide service providers in effectively improving service quality.

2. Theoretical Foundation

2.1 The Essence of Third-party Logistics

Third-party logistics refers to a business model in which specialized logistics providers, independent of both suppliers and demanders, deliver logistics services. As logistics practices have evolved, the concept of third-party logistics has continuously expanded, progressing from basic transportation and warehousing services to comprehensive modern logistics solutions that encompass entire supply chain management.

2.2 Principles of the Random Forest Algorithm

The Random Forest is an ensemble learning algorithm introduced by Leo Breiman in 2001, which constructs multiple decision trees and combines their predictions to produce a more accurate and robust prediction model. Its core methodology integrates Bootstrap Aggregation with random feature selection.

The mathematical expression for the Random Forest algorithm is given as: $\hat{y} = \frac{1}{K}\sum_{k=1}^{K} f\kappa(x)$, where $\hat{y}$ denotes the predicted outcome, K represents the number of decision trees, and $f\kappa(x)$ indicates the prediction output of the k-th decision tree.

3. Demand for Evaluation within the Third-party Logistics Industry

3.1 The Limitations of Traditional Evaluation Methods

In the face of complex logistics operational environments, traditional evaluation methods exhibit significant limitations. Firstly, indicator weighting is subjectively determined and overly reliant on expert experience, making them ill-suited for dynamic changes. Secondly, they struggle to handle nonlinear relationships and fail to fully capture the intricate interactions between indicators. Finally, their inadequate capacity to process massive datasets hinders the extraction of deep insights from enterprises' operational data. Therefore, there is an urgent need to adopt more objective, data-driven intelligent evaluation methods such as Random Forests.

3.2 The Advantages of Random Forests in Logistics Evaluation

Compared to traditional evaluation methods, random forests demonstrate significant advantages in third-party logistics assessment. High prediction accuracy: By integrating multiple decision trees, random forests effectively reduce model variance and enhance prediction precision. Handling nonlinear relationships: They automatically capture complex nonlinear relationships and interactions between indicators without requiring predefined functional forms. Feature importance assessment: They calculate importance scores for each evaluation metric to identify key factors influencing third-party logistics service quality. Anti-overfitting capability: Random sampling and random feature selection effectively mitigate overfitting risks. Insensitivity to missing data: They can handle datasets containing missing values with strong robustness.

4. Construction of an Evaluation Model Based on Random Forests

4.1 Data Preprocessing and Feature Engineering

Data preprocessing is a fundamental step in model construction and directly impacts model performance. The preprocessing of third-party logistics evaluation data primarily involves the following steps: Data cleaning: Handling missing values, outliers, and duplicate data to ensure data quality. For missing values, methods such as deletion, filling, or interpolation can be employed depending on their severity. Data transformation: Encoding categorical variables and discretizing continuous variables to make the data suitable for model processing. Data standardization: Converting metrics with different units into a unified format; common methods include Z-score normalization and minimum-maximum scaling.

4.2 Construction and Training of Random Forest Models

The construction of a random forest model involves the following key steps:

1. Determine the model type—select either a random forest classifier or regressor depending on whether the evaluation objective is classification (e.g., excellent, good, qualified, unqualified) or regression (e.g., composite score).
2. Set model parameters: Key parameters of a random forest include the number of decision trees (n_estimators), maximum tree depth (max_depth), minimum sample size required for splitting nodes (min_samples_split), minimum sample size for leaf nodes (min_samples_leaf), and number of features to select (max_features).
3. Training: Train the random forest model using training data by constructing multiple decision trees via bootstrap sampling to form the random forest.
4. Model Validation: Evaluate model performance using validation set data and adjust parameters to avoid overfitting or underfitting.

4.3 Model Optimization and Interpretation of Evaluation Results

Model optimization is a critical step in enhancing prediction performance, primarily involving parameter tuning and model integration: Parameter tuning involves identifying the optimal parameter combination using methods such as grid search, stochastic search, or Bayesian optimization.

Model Integration: Combining random forests with other algorithms (such as neural networks and support vector machines) to create more powerful integrated models.

A key advantage of the Random Forest model is its ability to assess feature importance, i.e., the contribution of each evaluation metric to the final outcome. Feature importance is typically calculated using Gini impurity or information gain; a higher score indicates greater contribution of the feature to the model. Through feature importance analysis, critical factors influencing third-party logistics service quality can be identified, providing clear guidance for improvement.

5. Conclusion and Outlook

5.1 Research Conclusion

This study systematically investigates evaluation indicators for third-party logistics based on random forests and develops a corresponding evaluation model. The key findings are as follows: First, the random forest algorithm is particularly well-suited for addressing complex challenges in third-party logistics evaluation, effectively capturing nonlinear relationships among indicators and enhancing assessment accuracy and reliability. Through Bootstrap sampling and random feature selection, random forests demonstrate strong resistance to overfitting and robustness.

Second, the third-party logistics evaluation model based on random forests can identify key factors influencing logistics service quality, providing clear guidance for improvement. Feature importance analysis quantifies each indicator's contribution to the evaluation results, enabling management to prioritize resource allocation.

5.2 Research Limitations and Prospects

This study has the following limitations: First, the indicator system may not fully capture all characteristics of various types of third-party logistics services; second, although the random forest model demonstrates strong predictive power, its interpretability remains inadequate; finally, the model's application requires substantial high-quality data support, making it less feasible in scenarios where data acquisition is challenging.

Future research can be conducted in the following aspects: First, explore the application of more advanced machine learning algorithms in third-party logistics evaluation, such as deep learning and reinforcement learning; second, strengthen cross-industry and cross-regional comparative studies to enhance the model's generalizability and adaptability.

References

[1] Wang Guihua et al. Research on Credit Evaluation for Logistics Enterprises Based on Random Forests. Logistics Technology, 123–128(2023).

[2] Zhang Xiaohong et al. A Review of the Application of Intelligent Algorithms in Logistics Route Optimization. Computer Application Research, 654–659(2024).

[3] Wu Keqiang. A Review of Research on Performance Evaluation of Third-Party Logistics Enterprises. Management Practitioner, (2013).

[4] Zeng Xiangyun, Wang Qiang. Research on the Connotation and Characteristics of Third-Party Logistics Management. China Circulation Economy,(2002);