Comprehensive Credit Rating Method for Civil Aviation Supply Chain Enterprises Based on Dual-Model Fusion

Wei Lin [1], Yiwen Zhang [2], Tantan Chen [1, 2],

Liang Li [1, 2], Jiehao Zheng [3], and Jing Li [3]

[1] Accounting Centre of China Aviation, Beijing, 100000, China;

[2] Antu Jinxin Commercial Factoring Limited, Tianjin, 300000, China;

[3] School of Computer Science and Technology, Civil Aviation University of China, Tianjin, 300000, China.

Abstract. Traditional enterprise credit rating models have limitations such as insufficient industry adaptability and singular risk indicators, which make them inapplicable to the credit rating of civil aviation supply chain enterprises in business financing scenarios. This paper proposes a comprehensive credit rating model that integrates financial, litigation, operational, and industry indicators. By improving the weighted random forest algorithm (introducing SMOTE data balancing and decision tree weight optimization) and combining it with Logistic regression in a dual-model fusion, a dynamic scoring card system is constructed. The model strengthens short-term liquidity indicators (with a weight of 35%) in line with the characteristics of civil aviation supply chain enterprises. Empirical results show that the model achieves an average accuracy of 94.7% in five-fold cross-validation and increases the recall rate of bad samples to 96.2%, significantly improving the precision of risk assessment. This model has been applied to the financing scenario of civil aviation supply chain and can effectively alleviate the information asymmetry between banks and enterprises, providing reliable credit decision support for financial institutions.

Keywords: Random Forest; Credit Rating; Short-term Liquidity; Civil Aviation Supply Chain.

1. Introduction

With the rapid development of the civil aviation industry, the financing needs of small and medium-sized enterprises (SMEs) in the upstream and downstream of the civil aviation supply chain are becoming increasingly urgent. Civil aviation supply chain enterprises typically apply for financing from financial institutions based on their business activities, characterized by small financing amounts and short repayment cycles[1]. However, these enterprises generally face issues such as low financial information transparency and insufficient collateral, resulting in financial institutions being reluctant or unwilling to lend. Enterprise credit rating is used to assess the willingness and ability of enterprises to fulfill their contractual obligations on time, providing objective credit information to financial institutions and helping them avoid risks.

Research on enterprise credit ratings, both domestically and internationally, has mainly focused on the following three aspects: (1) traditional financial ratio-based rating methods, which assess an enterprise's debt repayment ability through indicators such as the current ratio and debt-to-equity ratio; (2) methods that combine statistical approaches (such as Logistic regression) with machine learning models (such as Random Forest and Neural Networks) to enhance the predictive power of the model through big data analysis; and (3) credit scoring card models, which provide comprehensive scoring for financial and non-financial indicators through weight distribution. However, existing methods usually focus on the overall strength of the enterprise and bankruptcy risk, making it difficult to capture short-term solvency and industry volatility risks, and they cannot be directly applied to the credit rating of SMEs in the civil aviation supply chain[2].

Therefore, this paper proposes a targeted credit rating method based on the practical situation of small and medium-sized enterprises in the civil aviation supply chain. The specific research content includes: (1) integrating SMOTE oversampling with the weighted Random Forest method to address the data imbalance issue; (2) strengthening short-term liquidity indicators (such as current ratio, cash debt ratio), constructing industry segmentation indicators, focusing on industry-specific characteristics, and enhancing the model's adaptability across industries; and (3) combining Logistic regression with Random Forest to set the weight of indicators, improving the sensitivity of indicators for more accurate and effective credit rating.

2. Literature Review

Traditional enterprise credit rating methods primarily rely on scoring cards, such as those used by international agencies like Standard & Poor's and Moody's. The downside of these methods is that the determination of weights heavily depends on subjective human experience, and they cannot effectively eliminate the impact of information overlap on the rating results[3]. In this context, modern rating methods based on probability and machine learning models have been proposed[4-7]. These methods mainly build models based on historical data to predict the probability of enterprise default directly. Additionally, various dimensionality reduction methods are used to eliminate the impact of information overlap on the rating results, providing real-time dynamic capabilities. The disadvantage is that these methods require frequent data updates, otherwise, the ratings may lag or become unstable.

With the development of information technology, data structures are becoming more diversified, and the requirements for rating models have become more refined. Zhu Yuping and Chen Guanyu[8] proposed that using a single Logistic regression model cannot extract features from thousands of variables and emphasized the advantages of machine learning methods in credit evaluation. Ghatasheh[9] studied the effectiveness of Random Forest trees in credit risk prediction and highlighted the need to adjust the randomness and tree growth parameters to enhance classification results. Bou-Hamad[10] combined Random Forest-based methods with Bayesian model averaging to assess credit risk and predict defaults, showing higher prediction accuracy compared to other methods. Wang Haifeng[11] et al. combined data mining techniques with decision trees and fuzzy clustering algorithms to quantify the credit risk of SMEs and develop credit strategies. Feng B[12] applied a semi-supervised Random Forest regression model to predict the credit ratings of unrated companies, achieving high recognition accuracy and generalization ability. Zhouyi Gu[13] integrated the SMOTE algorithm with the XGBoost model to process imbalanced datasets. Compared to other models, the XGBoost-based scoring card model effectively improved the accuracy of credit risk assessment. However, these models have several limitations, such as linear assumption constraints, dependence on feature engineering, and limited representational ability, making them ineffective for credit rating of civil aviation supply chain enterprises.

Building on existing credit evaluation methods, this paper focuses on improving the model's applicability by combining objective data with controlled integration of subjective professional experience and industry volatility. Considering the difficulty in obtaining large numbers of high-quality sample data during modeling, and to avoid overfitting, this paper uses Logistic regression and an improved Random Forest model that require fewer parameter estimations for training.

3. Dual-Model Fusion for Enterprise Credit Rating Design

3.1 Overall Design of the Rating Method

The overall framework of the proposed credit rating method is shown in Figure 1. The core process includes:

1. Data Input and Processing Layer: Integrating financial, litigation, and industry data to construct a hierarchical indicator system and a structured feature matrix.
2. Model Fusion Layer: Joint training of Logistic regression (with strong interpretability) and improved Random Forest (with anti-overfitting capability), dynamically generating weights.

3. Evaluation and Optimization Layer: Adjusting thresholds based on enterprise types (listed/non-listed, financial/non-financial), reflecting differentiated risks, and determining positive and negative contributions of indicators.

4. Output Layer: Constructing subsystems for rating different types of enterprises, completing the indicator system construction, scoring, and outputting the credit rating (AAA-E).

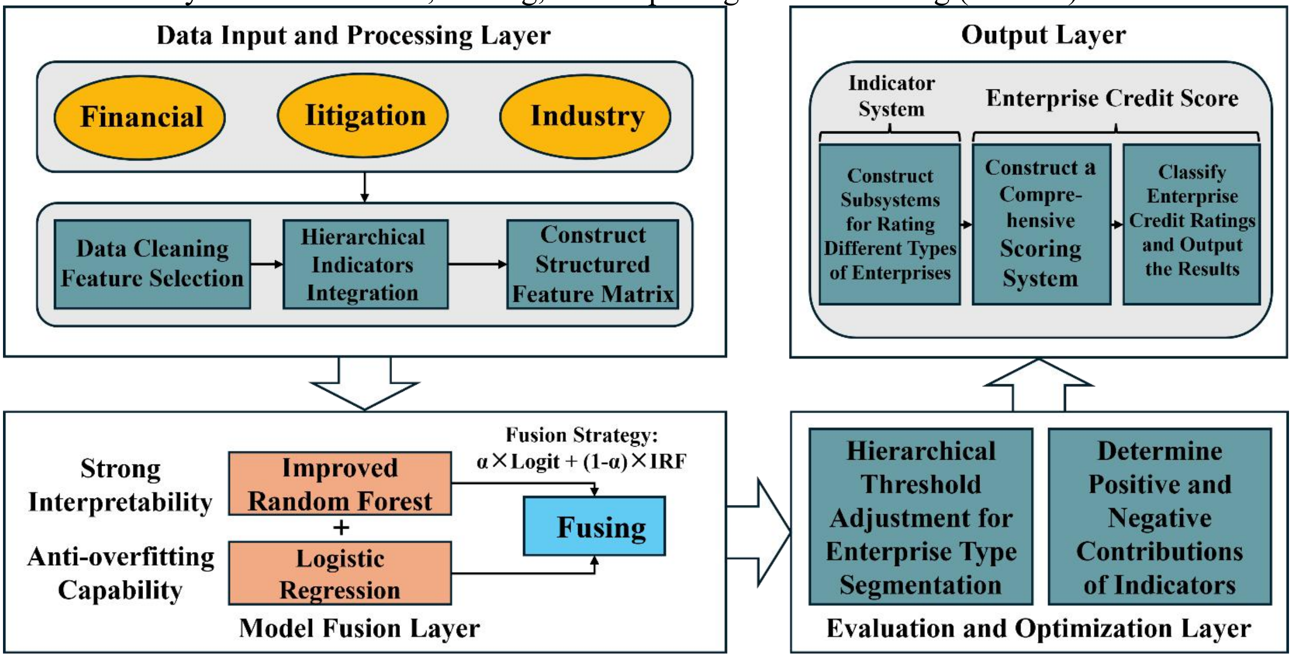

Fig. 1 Model Framework Process

3.2 Construction of the Evaluation Indicator System

3.2.1 Data Sources

To establish a targeted evaluation indicator system, this paper selects indicators from the perspectives of financial risk and business condition, while also considering the characteristic differences across industries. To overcome the limitation of insufficient sample sizes in business scenarios, this paper creatively introduces industry-specific segmentation indicators and uses a transfer learning framework to obtain experimental data from public databases and platforms.

(1) Financial Data

This paper collects financial data from 2010 to 2020 for all listed companies with A and B shares on the Shanghai and Shenzhen stock exchanges from the Wind database. The Wind database, a domestic authoritative financial data platform, covers a full range of corporate financial reports, market transactions, and other dimensional information, effectively reflecting the operational dynamics of enterprises.

(2) Litigation Data

This paper uses Python crawling technology (Scrapy framework + anti-crawling IP proxies) to scrape public litigation information from 553 listed companies. Data sources include authoritative platforms such as China Judgment Online and the National Enterprise Credit Information Publicity System.

(3) Industry Data

According to the indicators set by the National Bureau of Statistics, national economic market industries are divided into nine categories: Agriculture (agriculture, forestry, animal husbandry, and fishery); Industry; Construction; Wholesale and retail trade; Transportation, warehousing, and postal services; Accommodation and catering; Finance; Real estate; and Other industries. This paper collects industry data from 2012 to 2021.

3.2.2 Data Structure

(1) Financial Data

The financial data encompasses 43 indicators across 10 categories, including: short-term solvency (current ratio, quick ratio, cash-to-debt ratio); long-term profitability (return on equity (ROE), return on assets (ROA)); cash flow structure (net cash flow from operating activities, ratio of cash flow from investment activities); guarantee and pledge risks (pledge ratio, total external guarantees/net assets), as detailed in Table 1.

Table 1. Financial Risk Monitoring Indicator System for Enterprises

No.	Primary Indicators	Secondary Indicators
X1~X6	Assets and Liabilities	Total Assets, Current Assets, Non-current Assets, Registered Capital, Owner's Equity, Undistributed Profits
X7~X10	Profitability	Operating Revenue, Operating Profit, Total Profit, Net Profit
X11~X14	Cash Flow	Net Cash Flow from Operating Activities, Net Cash Flow from Investment Activities, Net Cash Flow from Financing Activities, Ending Balance of Cash and Cash Equivalents
X15~X16	Guarantees	Total Guarantee Balance, Stock Pledge Ratio
X17~X19	Short-term Solvency Indicators	Current Ratio, Quick Ratio, Cash Ratio
X20~X21	Long-term Solvency Indicators	Debt-to-Asset Ratio, Equity Multiplier (Total Liabilities / Equity Attributable to Shareholders of Parent Company)
X22~X27	Profitability Indicators	Return on Equity (ROE, average), Return on Assets (ROA), Core Business Ratio, Net Profit Margin, Earnings per Share (EPS - Basic), Cost-to-Profit Ratio
X28~X32	Operational Efficiency Indicators	Total Asset Turnover, Accounts Receivable Turnover, Inventory Turnover, Current Asset Turnover, Fixed Asset Turnover
X33~X39	Growth Capacity Indicators	Operating Revenue (YoY Growth), Operating Profit (YoY Growth), Total Assets (Growth from Start of Year), Shareholders' Equity (Growth from Start of Year), Net Profit (YoY Growth), Basic EPS (YoY Growth), Net Cash Flow from Operating Activities (YoY Growth)
X40~X43	Cash Flow Indicators	Net Cash Flow per Share from Operating Activities, Cash Received from Sales/Revenue, Cash Operating Index, Total Asset Cash Recovery Ratio
Y	Financial Crisis Status	Whether the company is marked as *ST or has defaulted on debt

(2) Litigation Data

After cleaning the litigation data, five core indicators are retained: number of credit defaulters and the number of enforcement actions for legal risks, the number of operational anomalies and administrative penalties for business compliance, and the number of major violations for severe illegal activities. Based on the financial status of the enterprise, companies are divided into two groups: those without financial crises and those with financial crises. The litigation indicators are then calculated for both groups.

As shown in Table 2, the median and mean values of the five indicators for the financial crisis group are much higher than for the non-financial crisis group. This indicates that the five core indicators selected in this study (such as credit defaulters, enforcement actions, operational anomalies, administrative penalties, and severe violations) are crucial for identifying financial risk in enterprises.

(3) Industry Data

Regarding the selection of industry indicators, this paper follows two main principles: (1) Industry representativeness, for example, the "capital adequacy ratio" and "non-performing loan ratio" are important for the financial industry, while "capacity utilization rate" and "inventory turnover rate" are more relevant for the manufacturing industry; (2) Risk foresight, using industry data from two years earlier (e.g., 2019 data to predict 2021 risks) to enhance the model's predictive ability. The goal is to predict if an industry is facing a crisis two years later, ensuring the evaluation method has forward-looking capabilities.

Below is a list of the industry indicators selected in this study. Based on their impact on enterprise rating, the indicators are classified into positive and negative indicators. According to industry characteristics, each industry has its own set of sub-indicators.

1) Positive Indicators: Gross domestic product (GDP) by industry sector, GDP by industry sector index (last year = 100), number of enterprises, main business income of enterprises, total assets, total profits, industry value-added ratio.

2) Negative Indicators: Number of loss-making enterprises, debt-to-asset ratio, total losses.

Table 2. Descriptive Statistics of Litigation and Operational Indicators

Enterprise Group	Indicator Name	Min	1st Qu	Median	Mean	3rd Qu	Max
Non-Financial Crisis Group	Number of Credit Defaulters	0.00	0.00	0.00	1.10	0.00	106.00
	Number of Enforcement Actions	0.00	0.00	0.00	1.60	0.00	119.00
	Number of Operational Anomalies	0.00	0.00	1.00	68.50	8.00	5000.00
	Number of Administrative Penalties	0.00	0.00	0.00	8.97	2.00	550.00
	Number of Severe Violations	0.00	0.00	0.00	0.39	0.00	29.00
Financial Crisis Group	Number of Credit Defaulters	0.00	0.00	42.50	229.63	159.25	2201.00
	Number of Enforcement Actions	0.00	1.00	30.50	120.07	92.00	782.00
	Number of Operational Anomalies	0.00	25.50	345.50	1309.40	1593.00	5000.00
	Number of Administrative Penalties	0.00	5.00	54.00	525.13	293.25	4924.00
	Number of Severe Violations	0.00	0.00	4.50	36.57	32.50	515.00

3.2.3 Data Preprocessing

(1) SMOTE-Based Data Balancing

The data collected in this study suffers from a severe imbalance between positive and negative samples. For example, in the financial data, the samples that are classified as having a financial crisis account for only 1.22% of the total samples, which is less than 5%. To prevent the fitting results from being overly optimistic (e.g., high accuracy with low recall), and to improve recall rates (True Positive Rate or TPR), this paper uses the SMOTE algorithm to perform resampling. The final result has 1224 negative samples (Y=1) and 1088 positive samples (Y=0), making the distribution of good and bad samples more balanced.

(2) Handling Missing Values

Indicators with a missing rate greater than 30% (e.g., "R&D expense ratio") are removed. For other indicators, the following methods are applied:

① Replace missing values with corresponding indicators from similar clients in the same industry or with similar scale.

② Exclude non-computable rating indicators from the rating system. For the remaining computable indicators, their weights are adjusted. The adjustment method for weight is as follows: if the original weight of a computable indicator is , the updated weight is calculated as:

$$w_i = \frac{w_i^0}{\sum_{j=1}^{p} w_j^0}, \; i = 1, \cdots, p \tag{1}$$

(3) Data Normalization

For ease of computation and processing, feature scaling is performed using the following methods to standardize indicators based on their types:

a. Positive Indicators:

$$X'_{ij} = \frac{X_{ij} - \min X_{ij}}{\max X_{ij} - \min X_{ij}} \tag{2}$$

b. Negative Indicators:

$$X'_{ij} = \frac{\max X_{ij} - X_{ij}}{\max X_{ij} - \min X_{ij}} \tag{3}$$

c. Moderate Indicators:

$$Y'_{ij} = \frac{1}{|X_{ij} - \bar{X}_{ij}|}; \quad X'_{ij} = \frac{Y'_{ij} - \min Y'_{ij}}{\max Y'_{ij} - \min Y'_{ij}} \tag{4}$$

Here, i denotes the year, and j represents the indicator.

3.2.4 Feature Extractio

For the financial data, after preprocessing, feature extraction is conducted. The 43 financial indicators are examined for multicollinearity using the condition number test. A condition number of kappa = 4.35×10^{17} indicates significant multicollinearity in the explanatory variables. The dataset is divided into 75% for training and 25% for testing in order to process the explanatory variables.

After performing Principal Component Analysis (PCA) on the 43 indicators, the first 19 principal components, which account for 90.8% of the cumulative variance, are selected as the financial evaluation indicators. These 19 indicators are renamed RC1 to RC19, which include: basic earnings per share (EPS), return on equity (ROE), cost-to-profit ratio, stock pledge ratio, total asset cash recovery rate, current ratio, operating profit, cash ratio, total asset turnover, return on assets (ROA), year-end balance of cash and cash equivalents, debt-to-equity ratio, owner's equity, fixed asset turnover, equity ratio (total liabilities/total equity of the parent company), net profit, current asset turnover, sales net profit margin, and quick ratio.

This paper uses Logistic regression modeling for the 19 selected financial indicators. The regression coefficients and corresponding p-values for each indicator are shown in Table 3. From the results in the table, it can be seen that RC1, RC5, RC7, RC8, RC18, RC12, and RC14 are not significant at the 90% confidence level, and their estimated coefficients are relatively small compared to the other indicators. Therefore, the weights of these indicators are adjusted downward.

Table 3. Regression Coefficients of Financial Indicators

No.	Estimate	Std. Error	Z-value	P-value	Significance
(Intercept)	0.44	0.09	5.04	4.58E-7	***
RC4	-2.45	0.65	-3.77	1.62E-4	***
RC16	1.00	0.22	4.55	5.26E-6	***
RC17	1.44	0.09	15.37	< 2E-16	***
RC10	0.72	0.24	2.95	3.21E-3	**
RC15	-0.46	0.17	-2.73	6.27E-3	**
RC2	-1.03	0.51	-2.02	4.33E-2	*
RC3	0.26	0.10	2.57	1.03E-2	*
RC9	0.30	0.14	2.20	2.81E-2	*
RC11	0.43	0.18	2.38	1.73E-2	*
RC13	-2.25	0.97	-2.32	2.04E-2	*
RC19	0.36	0.18	2.02	4.37E-2	*
RC1	-0.67	0.61	-1.09	2.76E-1	
RC5	0.10	0.09	1.15	2.49E-1	
RC6	-0.20	0.11	-1.83	6.76E-2	
RC7	-0.53	0.52	-1.02	3.08E-1	
RC8	-0.35	0.39	-0.90	3.70E-1	
RC12	-0.44	0.56	-0.78	4.34E-1	
RC14	0.05	0.06	0.78	4.38E-1	
RC18	-0.14	0.10	-1.40	1.59E-1	

3.2.5 Calculation of Indicator Weights

This paper constructs indicator weights based on both the Principal Component Logistic Regression model and the Random Forest model. The results of the two models are adjusted to obtain the final weights. To enhance model performance, the Random Forest model is improved as follows:

① Assign different weights to decision trees based on their classification accuracy, reducing the weight of poorly performing trees. The out-of-bag (OOB) data is used to determine the correct classification rate for each decision tree, and subsets with lower dimensions are selected to train the decision trees.

② Instead of using the mode for voting, calculate the weight of each decision tree based on its correct classification rate.

Finally, the final weights of the evaluation indicator system are obtained by jointly training and weighting the Logistic Regression and Random Forest models. In practice, Logistic Regression assigns weights based on statistical significance and linear relationships, emphasizing the direct impact of indicators on financial crises (e.g., the weight of "number of credit defaulters" is 0.443). Its advantage lies in strong interpretability, but it may overlook nonlinear feature interactions. Random Forest captures nonlinear relationships and interactions between indicators (e.g., the relationship between administrative penalties and operational anomalies) through feature importance (e.g., the importance of "Number of Credit Defaulters" is 69). The combined weights from both models balance linear and nonlinear impacts, making the model more comprehensive when dealing with complex data.

To align with the characteristics of civil aviation supply chain enterprises—short repayment cycles and small financing amounts—this paper increases the weight of short-term solvency indicators, such as the current ratio, cash-to-debt ratio, and quick ratio. By adjusting the weight ratios, short-term solvency indicators account for 35%, significantly higher than long-term profitability indicators (e.g., ROE with a weight of 12%), ensuring that the model better suits the risk assessment needs of business financing scenarios.

3.3 Design of a Comprehensive Scoring System

By integrating the financial, litigation and operational, and industry scores, this study constructs a comprehensive credit risk rating system for enterprises. The indicator framework of the system is shown in Table 4. Sub-rating subsystems have been separately developed for different enterprise types, including listed state-owned enterprises, listed non-state-owned enterprises, unlisted state-owned enterprises, internet companies, general financial enterprises and licensed financial institutions, general enterprises, and other special enterprises. Due to space limitations, these subsystems are not elaborated upon in the main text.

Table 4. Indicator Framework of Enterprise Credit Risk Rating System

	Primary System	Secondary System	Tertiary System
Enterprise Credit Risk Rating System	Financial, Litigation & Operational, Industry Indicator System	Financial Indicator System	Financial Independent Variable Indicator System
			Financial Dependent Variable Indicator System
		Litigation & Operational Indicator System	Litigation Indicator System
			Operational Indicator System
		Industry Indicator System	Nine Industry Indicator Systems
	Special Cases	Special Case Indicator System	

Based on the scoring results from the historical enterprise dataset using the new model and comparing them with the actual credit conditions, credit rating thresholds are defined such that approximately 40% of samples fall into Grade A or above, 30% into Grade B, and the remaining 30% into Grade C or below. As illustrated in Figure 2, the credit grades are determined accordingly.

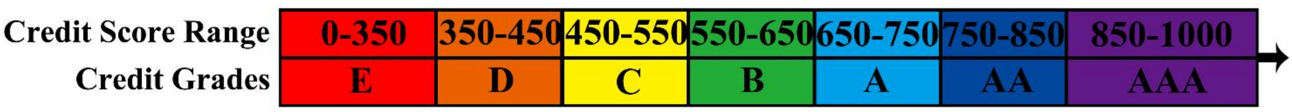

Fig. 2 Credit Scoring Grades

4. Empirical Analysis

4.1 Validation of the Indicator System

To verify the effectiveness of the evaluation indicators, a Random Forest prediction model was established in this study to predict company financial data obtained from the Wind database. The prediction results for the entire dataset are shown in Table 5, with a prediction accuracy of 94.42%. All 136 bad samples were correctly identified.

Further, five-fold cross-validation was conducted on the full sample set to evaluate the average prediction accuracy. The results are also presented in Table 5. The five verification rounds showed stable prediction performance, with an average accuracy of 94.74%, and all bad samples were successfully predicted. This indicates that the constructed evaluation indicator system achieves the goal of risk control in credit scoring.

Table 5. Prediction Results of Full Sample and Five-Fold Cross-Validation

Indicator Group	Actual \ Predicted	Non-Financial Crisis	Financial Crisis
Full Sample Prediction	Non-Financial Crisis	10399	0
	Financial Crisis	623	136
Five-Fold Cross-Validation	Non-Financial Crisis	1015	0
	Financial Crisis	65	32

4.2 Validation of the Comprehensive Scoring System

The credit scores generated by a single financial model and the comprehensive model (which includes financial, litigation, and industry indicators) were compared, as shown in Figure 3:

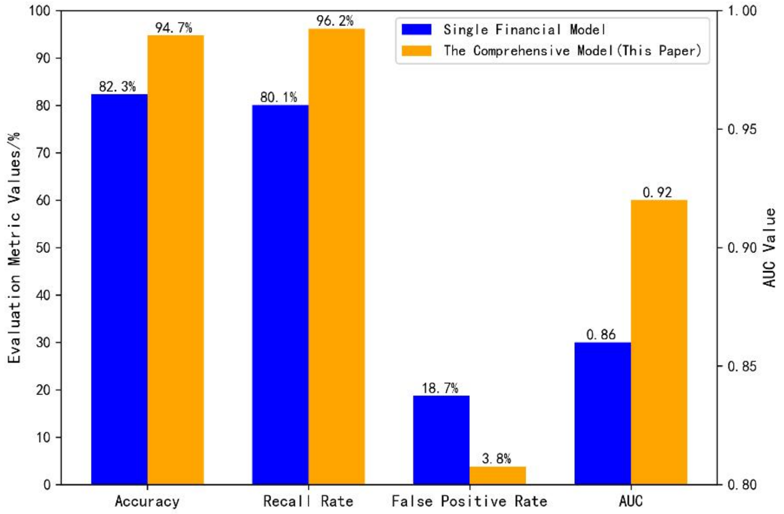

Fig. 3 Model Comparison Results

It can be observed that the comprehensive model performs well in cross-industry scenarios, with significantly higher accuracy and recall rates compared to the single financial model. In particular, the false positive rate for identifying bad samples was reduced to 3.8%, validating the effectiveness of multi-dimensional indicator fusion.

To further verify the model's effectiveness, a comparative scoring analysis was conducted using the Random Forest model, Logistic Regression model, and the fusion model. The results are shown in Figure 4:

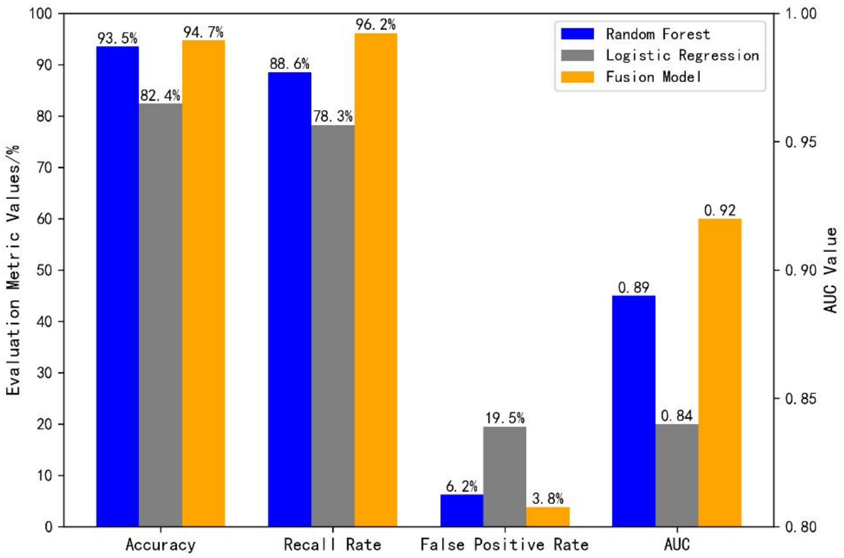

Fig. 4 Comparison of Single Models and Fusion Model

It is evident that the fusion model, by balancing nonlinear relationship capture and statistical significance, significantly outperforms the single models across all metrics. In terms of recall rate, the fusion model achieved a recall of 96.2%, which is 7.6% and 17.9% higher than the Random Forest model (88.6%) and the Logistic Regression model (78.3%), respectively. This demonstrates its superior capability to capture complex risk features. The false positive rate of the fusion model is also substantially lower than that of Random Forest (6.2%) and Logistic Regression (19.5%), confirming the enhanced ability of multi-model joint training in identifying high-risk enterprises. Furthermore, the AUC of the fusion model reaches 0.92, higher than Random Forest (0.89) and Logistic Regression (0.84), indicating superior overall classification performance and stronger robustness against noisy data.

The fusion model ensures a high recall rate while reducing false positives, helping financial institutions reduce bad debt losses and avoid the over-rejection of high-quality enterprises.

5. Conclusion and Outlook

Through the fusion of multi-dimensional indicators and improvements in weight algorithms, the credit rating model constructed in this study achieves several significant breakthroughs:

1. Enhanced Accuracy: The comprehensive model reduces the misclassification rate by 15% compared to single models.
2. Improved Industry Adaptability: By dynamically adjusting the weights of indicators specific to the civil aviation industry, the model aligns with the industry's need for short-term liquidity.
3. Stronger Interpretability: The results of Logistic Regression and Random Forest are cross-validated, enabling traceable decision-making.

The overall framework optimizes the rating process and data quality, fully leveraging data value. It holds significant practical value in business-based enterprise credit evaluation, risk control, and monitoring enterprise development. The model has already been deployed and has improved monitoring coverage and effectiveness, with no occurrences of business-based enterprise defaults or bad debts under the model's monitoring. Based on the proposed model, civil aviation supply chain enterprises can refer to the corresponding indicator scores to effectively optimize various financial ratios; meanwhile, financial institutions and related investors can forecast the credit condition of target enterprises for the upcoming year, thereby facilitating early risk warnings.

However, the model still has limitations. For example, it cannot perform dynamic credit monitoring and must be constantly updated with new data over time for calibration. Additionally, it lacks the ability to assess associated risks within the supply chain. In the future, we plan to explore the integration of temporal data and apply graph neural networks to mine correlation risks, thereby expanding the model to more credit rating scenarios and validating its generalization ability.

References

[1] Zhang J, Luo Y, et al. Enterprise credit risk portrait and evaluation from the perspective of the supply chain[J]. International Transactions in Operational Research, 2024, 31(4): 2765-2795.

[2] Yi-Qing Tian, Xi Cheng, Bo-Jing Feng. A Review of Corporate Credit Rating Calculation Models[J]. Data and Computing Frontiers, 2023, (4): 139-153.

[3] Kim K, Ahn H. A corporate credit rating model using multi-class support vector machines with an ordinal pairwise partitioning approach[J]. Computers & Operations Research, 2012, 39(8): 1800-1811.

[4] Florez-Lopez R, Ramon-Jeronimo J M. Enhancing accuracy and interpretability of ensemble strategies in credit risk assessment: a correlated-adjusted decision forest proposal[J]. Expert Systems with Applications, 2015, 42(13): 5737-5753.

[5] Shi B, Meng B, Yang H, et al. A novel approach for reducing attributes and its application to small enterprise financing ability evaluation[J]. Complexity, 2018, 2018: 1-17.

[6] Dong G, Lai K K, Yen J. Credit scorecard based on logistic regression with random coefficients[J]. Procedia Computer Science, 2010, 1(1): 2463-2468.

[7] Sun H, Guo M. Credit Risk Assessment Model of Small and Medium-Sized Enterprise Based on Logistic Regression[C]//Choy K L, Huang G Q. Proceedings of the 2015 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM). Singapore: IEEE, 2015: 1714-1717.

[8] Yu-Ping Zhou, Guan-Yu Chen. Research on Personal Credit Evaluation Based on Machine Learning Methods[J]. Financial Theory and Practice, 2019, (12): 1-8.

[9] Ghatasheh N. Business analytics using random forest trees for credit risk prediction: a comparison study[J]. International Journal of Advanced Science and Technology, 2014, 72(2014): 19-30.

[10] Bou-Hamad I. Bayesian credit ratings: A random forest alternative approach[J]. Communications in Statistics Theory and Methods, 2017, 46(15): 7289-7300.

[11] Hai-Feng Wang, Xiao-Ni Zhang, Bao-Feng Shi. P2P Online Lending Personal Credit Evaluation Model and Application Based on Fuzzy Clustering[J]. Zhejiang Finance, 2017, (10): 19-26.

[12] Feng B, Xue W. Adversarial semi-supervised learning for corporate credit ratings[J]. arXiv Preprint, 2021, 12(4):123-134.

[13] Gu Z, Lv J, Wu B, et al. Credit risk assessment of small and micro enterprise based on machine learning[J]. Heliyon, 2024, 10(5):12-21