International Journal for Housing Science and Its Applications
Publish August 6, 2025. Volume 46, Issue 3 Pages 5194-5203

https://doi.org/10.70517/ijhsa463446

Research on the Construction of Financial Early Warning System Based on K Nearest Neighbor Algorithm in Enterprise Capital Chain Wind Control in the Era of Digital Transformation

Ruilan Zhang[1,*]
[1] Harbin University of Commerce, Harbin, Heilongjiang, 150000, China
Corresponding authors: (e-mail: 13108052868@163.com).

Abstract Reducing the risk of enterprise capital chain breakage is the key to guarantee the stable operation of enterprises. This paper analyzes the characteristic variables affecting the financial status of enterprises and constructs the financial early warning index system. Use the isolated forest anomaly detection algorithm to calculate the likelihood of sample data anomalies, and quickly realize data preprocessing. Using K-neighborhood algorithm to complete the calculation of financial data distance, to determine the early warning classification of the sample enterprise indicator data. Combined with circular experiments to find the optimal parameters of the financial early warning model. Compare the prediction accuracy of the financial early warning model of each classification to verify the advantages of the model in this paper. The results show that the model has the highest prediction accuracy when the environmental parameter b takes the value of [0.2,0.9] and the threshold percentage p takes the value of [0.7,1]. The prediction accuracy of the financial early warning model based on the K-neighborhood algorithm reaches 84.94%, which is higher than that of the other nine prediction models, and it has excellent financial risk prediction capability.

Index Terms financial early warning, isolated forest algorithm, anomaly detection, K-neighborhood algorithm

I. Introduction

Under the modern enterprise system, the enterprise, as an independent legal person that operates independently and is responsible for its own profit and loss, is the inevitable result of the development of socialized mass production and market economy [1]. The good or bad business performance and results of enterprises is an important standard to measure the competitiveness of enterprises, and even the industrial strength of a country or nation [2], [3]. In the process of business management, the safety and management of the capital chain is one of the core management contents of the enterprise [4]. The success or failure of the enterprise depends largely on whether the enterprise can design and effectively implement the enterprise's capital chain management mode according to its own basic situation, combined with the current macro situation and micro environment, in order to adapt to the requirements of the fierce market competition environment [5]-[7]. However, enterprises often face the problems of insufficient scientific planning and weak risk monitoring mechanism in actual operation, which leads to the risk of capital chain breakage [8], [9].

In this regard, enterprises should scientifically carry out capital planning, improve the risk monitoring mechanism, and improve the business quality of financial personnel in order to improve the level of capital chain management, and then better improve the financial risk early warning mechanism [10]-[12]. Through this forward-looking management, it can identify the existence of financial risks in the project planning stage and formulate corresponding preventive strategies, so as to reduce the impact of future uncontrollable factors on the operation of the enterprise [13], [14]. Only by effectively identifying and controlling the degree of financial risk of enterprises can we ensure the healthy, sustainable and stable development of a large number of independently operated enterprise groups in China, so that the overall strength of enterprises can be more competitive in the international market, thus effectively enhancing the comprehensive national power [15]-[17].

This paper establishes a multi-indicator financial early warning system as a categorization reference for financial data anomaly detection. By calculating the path length from the root node to the leaf node of the isolated forest method, the abnormal sample data are quickly detected and corrected to complete the preprocessing of the sampled data. The K-neighborhood algorithm is used to classify the sample data according to the distance size, determine the risk category to which the enterprise sample data belongs, and judge the possibility of the emergence of enterprise financial risk. Conduct descriptive statistics and correlation analysis on the indicators of the constructed model to ensure the application ability of the indicators. Set the optimal parameters to improve the predictive accuracy of the model.

II. Establishment of financial early warning model based on K-neighborhood algorithm

This chapter selects the corresponding indicators to construct the enterprise financial early warning indicators, and explains the basic connotation of the isolated forest method and K-neighborhood algorithm used, and analyzes the construction process of the financial early warning model.

II. A. Construction of the indicator system

Before establishing the enterprise financial early warning model, it is necessary to clarify the relevant financial indicators. In this paper, we choose 6 first-level indicators and 12 second-level indicators as the characteristic variables for financial early warning of enterprises in the era of digital transformation. Table 1 shows the specific content of the indicators. 6 first-level indicators include: characteristics reflecting solvency (A1), including current ratio (B1) and cash ratio (B2); characteristics reflecting ratio structure (A2), including asset-liability ratio (B3) and working capital ratio (B4); characteristics reflecting operating ability (A3), including inventory turnover (B5) and total asset turnover (B6); Characteristics reflecting profitability (A4), containing the interest coverage multiple (B7) and the net profit margin on total assets (B8); Characteristics reflecting cash flow analysis (A5), containing the interest coverage multiple on cash flow (B9) and the cash reinvestment ratio (B10); Characteristics reflecting development capability (A6), containing the operating profit margin (B11) and the return on assets (B12).

Table 1: Financial early warning indicator system

First-level indicator	First-level indicator number	Secondary indicators	Secondary indicator number
Debt-paying ability	A1	Current ratio	B1
		Cash ratio	B2
Ratio structure	A2	Asset-liability ratio	B3
		Working capital ratio	B4
Business operation ability	A3	Inventory turnover rate	B5
		Total asset turnover rate	B6
Profitability	A4	Interest coverage ratio	B7
		Net profit margin on total assets	B8
Cash flow analysis	A5	Cash flow interest coverage ratio	B9
		Cash reinvestment ratio	B10
Development ability	A6	Operating profit margin	B11
		Return on assets	B12

II. B. Data preprocessing based on the isolated forest approach

Since the isolated forest is also integrated from the tree structure (binary tree), the starting point of the whole isolated forest algorithm is that the anomalous samples are more likely to fall into the leaf nodes quickly, or the anomalous samples are closer to the root node, or the anomalous samples have a shorter path in the binary tree. Since the anomalies make up only a small portion of the tree structure, they are quickly divided from the root node to the leaf nodes, so the length of the path from the root node to the leaf nodes can be used to determine the degree of abnormality of this data point. Figure 1 shows the isolated tree structure. In which the binary tree in the isolated forest is known as isolated tree and the process of its generation is a completely randomized process.

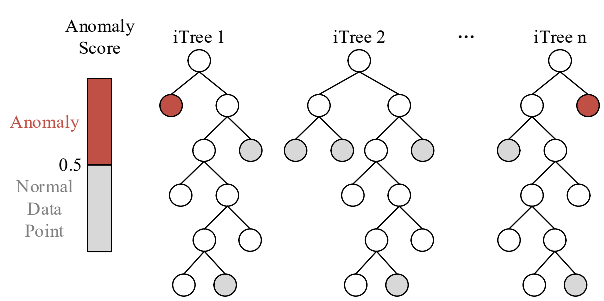

Figure 1: Outlier tree structure

Isolated forest anomaly detection algorithm is mainly divided into two steps, the first step is to use the sample data training to get a number of isolated trees and form an isolated forest, the second step is to bring the sample into each isolated tree of the isolated forest and then calculate the sample's anomaly score and determine whether it is anomalous or not, the specific steps are as follows.

Step 1:

1) Input the feature datasets $X=\{x_1,x_2,x_3,\cdots,x_n\}$, $\forall x_i \in X$, $x_i=(x_{i1},x_{i2},x_{i3},\cdots,x_{im})$, and randomly select ψ (default is 258) samples from X as subspace X^* of X and place them on the root node of the isolated tree.

2) Randomly select a variable j among the m feature variables of the subspace X^* and generate a random isolation value p from this feature variable, where p is greater than the minimum value of the feature variable j is less than the maximum value of the feature variable j .

3) Using this isolation value p (random hyperplane), divide the dataset into two, and if the specific value of the feature variable j is less than p , put the corresponding sample into the left branch, and vice versa into the right branch.

4) Repeat b and c , i.e., isolate the data space on the left and right branches until the isolation tree reaches a certain depth (default is 10), or the current leaf node contains only one sample point, or the features of the samples on the nodes are all the same.

5) Repeat a to d until t (default 101) isolated trees are generated and form an isolated forest.

Step 2:

1) For each sample x_i , make it traverse each isolated tree in the isolated forest, and get its depth $h(x_i)$ in each tree, and then get its average depth $E(h(x_i))$ in the whole isolated forest, where $h(x_i)$ refers to the total number of branches, i.e., the length of paths, that the sample x_i passes through from the root node of the tree to its leaf nodes.

2) Substitute the average depth $E(h(x_i))$ and $c(\psi)$ obtained in f into equation (1), which in turn yields the anomaly score of sample x_i .

$$s(x_i,\psi)=2.5^{-\frac{E(h(x_i))}{c(\psi)}} \tag{1}$$

Among them:

$$c(\psi)=\begin{cases}2.5H(\psi-1)-2(\psi-1)/\psi, & \psi>2.5\\ 1, & \psi=2.5\\ 0, & otherwise\end{cases} \tag{2}$$

$H(k)=\ln(k)+\zeta$, $\zeta=0.5772156648$ is Euler's constant.

The above is the main idea and steps of the isolated forest anomaly detection algorithm, the anomaly score $s\left(x_i,\psi\right)$ is obtained with the value range of $\left(0.1,1.1\right)$, and the larger the value indicates that the greater the probability of this sample being an anomaly, and vice versa indicates that the smaller the probability of this sample being an anomaly, while the Python's sklearn library corrects the anomaly score, i.e., first take the opposite of $s\left(x_i,\psi\right)$ and then add 0.55 to it, and the corrected anomaly score is bounded by 0. When the anomaly score of a certain sample is greater than 0, it means that this sample is normal, and vice versa, it means that this sample is anomalous if the value of the anomaly score is less than 0.

II. C.Predictive modeling based on K-neighborhood approach

II. C. 1) Basic conceptual analysis

K Nearest Neighbor (KNN) algorithm is a classification algorithm used in the fields of character recognition, text classification and image recognition.

Algorithm idea: a sample and the K samples in the data set is the closest, and secondly, if the selected K samples by the vast majority of them belong to the same classification, then the samples that need to judge the category will also belong to the same category. Figure 2 shows the KNN structure.

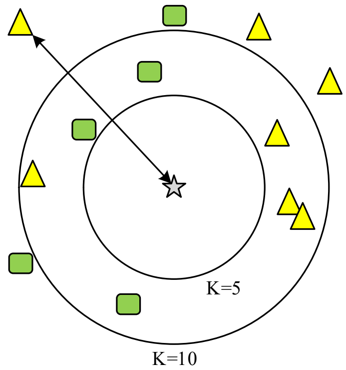

Figure 2: KNN structure

The center point of the circle is the sample to be classified, $K=6$ and $K=12$ are different distance identifiers, triangles and rectangles are the two categories, respectively, calculate the distance between the center of the circle pentagram and the data of the two, and classify them according to the different K, to determine which category the sample to be classified belongs to.

II. C. 2) Distance metrics

It can be seen that in the feature space of the measure distance, in order to describe the degree of similarity between two points to be classified, the distance between two points will be calculated. The feature space of the K nearest neighbor algorithm we default to is the n-dimensional real vector space R^n, using the Euclidean distance metric, which can also be other distances.

Let the feature space X be the n-dimensional real vector space R^n, $x_i, x_j \in X$ $x_i = \left(x_i^{(1)}, x_i^{(2)}, \cdots, x_i^{(n)}\right)^T$, $x_j = \left(x_j^{(1)}, x_j^{(2)}, \cdots, x_j^{(n)}\right)^T$, and x_i, x_j's L_p distance is defined as:

$$L_p(x_i, x_j) = \left(\sum_{l=1}^{n} \left|x_i^{(l)} - x_j^{(l)}\right|^p\right)^{\frac{1}{p}} \tag{3}$$

Here $p \geq 1.1$.

When $p = 1.1$, it is called the Manhattan distance, i.e:

$$L_p(x_i, x_j) = \sum_{l=1}^{n} \left|x_i^{(l)} - x_j^{(l)}\right| \tag{4}$$

When $p = 2.1$, it is called the Euclidean distance, i.e:

$$L_p(x_i, x_j) = \left(\sum_{l=1}^{n} \left|x_i^{(l)} - x_j^{(l)}\right|^2\right)^{\frac{1}{2}} \tag{5}$$

When $p \to \infty$, it is the maximum value of the individual coordinate distances, i.e:

$$L_\infty(x_i, x_j) = \max_l \left|x_i^{(l)} - x_j^{(l)}\right| \tag{6}$$

II. C. 3) Principles of the KNN algorithm

In this algorithm each training set data corresponds to a label, from which the mutual correspondence between each data in the sample and its category can be derived. If new data without labels are input, all the features of the new data can be compared with the corresponding features of the data in the sample set, and the label of the sample with the closest distance, i.e., with the highest number of feature similarities in the known sample set, can be extracted according to the K nearest neighbor algorithm. And based on the set K value, the classification of this K sample that accounts for the majority of the samples is calculated, and finally the selected classification can be represented as the classification of the new data. In the K nearest neighbor algorithm, the similarity between the sample sets can be illustrated by calculating their distance, and the match between them is calculated based on the distance calculation. Figure 3 shows the KNN calculation process.

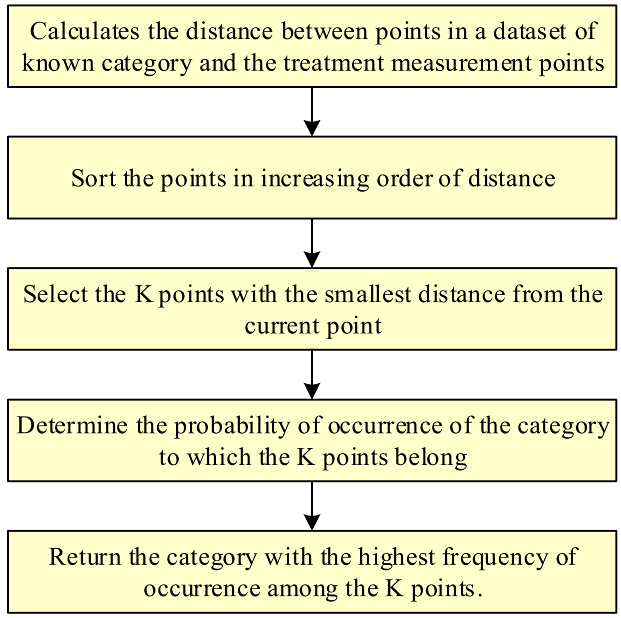

Figure 3: KNN calculation process

II. C. 4) Financial early warning modeling

This paper conducts research on the early warning of enterprise financial crises. Therefore, when establishing a prediction model using the K nearest neighbor method, the enterprise samples are divided into three major categories: Financial crisis (ST) enterprises, enterprises with unstable financial conditions and normal enterprises. ST enterprises fall into the first category and are represented by "0", enterprises with unstable financial conditions fall into the second category and are represented by "0.5", and normal enterprises fall into the third category and are represented by "1". Set up enterprise test sample set for $X=\{x_1,x_2,\cdots x_n\}$, where n, for enterprise to evaluate total X_i said the i be evaluated enterprise; Known enterprise training sample set for $Y=\{y_1,y_2,\cdots y_m\}$, where m for the training sample, total y_j said the j a training sample, Moreover, the category to which each y_j belongs has been determined, that is, whether the value of y_j is 0, 0.5 or 1 has all been determined.

Definition 1: Let X_{ia} represent the a th warning indicator of X_i enterprises; y_{ja} represent the a th warning indicator of the y_j training sample, then the distance between the X_i enterprises and the y_j training sample is:

$$d\left(x_i,y_j\right)=\sqrt{\sum_{a=1}^{q}\left(x_{ia}-y_{ja}\right)^2} \tag{7}$$

where q is the total number of early warning indicators for the assessed enterprise.

Let $j=(1,2\cdots m)$, calculated m distance, $d\left(x_i,y_j\right)$ ascending order, select the first K $d\left(x_i,y_j\right)$ in the order, and statistics, where p means that the first K $d\left(x_i,y_j\right)$ there are then there are p training samples belong to ST enterprises, q means that the first K $d\left(x_i,y_j\right)$ there are then there are q training samples belong to normal enterprises, then xi enterprise financial crisis classification prediction model can be expressed as:

$$g(x_i)=\begin{cases}0, p>q\\ 1, p<q\\ False, p=q\end{cases} \quad , Which\ p+q=k \tag{8}$$

In the empirical part of this paper, the nearest neighbor method is used, so that k takes a value equal to l, and therefore the case of $P=Q$ does not occur.

When discussing nearest neighbor algorithms it is often necessary to consider the issue of error rates, for example, Cover and Hart have analyzed the error rates of nearest neighbor classification methods in detail. The conditional error rate of the $k-$ nearest neighbor algorithm for classification is, for a given test sample point x:

$$\begin{aligned} P_{N\to\infty}^{K}\left(e \mid x\right) &= P\left(m_1 \mid x\right)\sum_{j=0}^{(k-1)/2} C_k^j p\left(m_1 \mid x\right)^j p\left(m_2 \mid x\right)^{k-j} \\ &+ P\left(m_2 \mid x\right)\sum_{j=0}^{(k-1)/2} C_k^j p\left(m_2 \mid x\right)^j p\left(m_1 \mid x\right)^{k-j} \end{aligned} \tag{9}$$

where the first and second terms are the conditional error rates of $x \in m_1$ and $x \in m_2$, respectively.

The model used by the K nearest neighbor algorithm actually corresponds to a partition of the feature space. The three basic elements of the algorithm are referred to as: the choice of K values, the distance metric and the classification decision rule.

III. Analysis and application of financial early warning models

This chapter divides the data of the selected sample enterprise indicators and also analyzes the availability of the selected indicators. Find the optimal parameters of the financial early warning model in this paper and compare the performance of financial early warning models based on different algorithms.

III. A. Determination of the sample

III. A. 1) Selection of the study sample

Early warning model is a financial model that utilizes existing financial information to predict the future development of the enterprise in advance and informs managers of the risk level as early as possible, so as to facilitate the timely adoption of reasonable and powerful measures for risk control and avoidance. Therefore, in this paper, when constructing the early warning model, we will predict the financial status of the sample enterprises in the coming year, that is, year t, based on the relevant index data in year t-1. In addition, because the training model requires a certain sample size, in order to ensure that the sample size is sufficient, this paper selects 180 listed companies in the total of three years from 2020 to 2022, a total of 360 groups of samples for the study.

III. A. 2) Segmentation of the sample

There are different methods for clarifying whether an enterprise has a financial crisis. In this paper, based on the reality in China, whether ST (or *ST) is marked on the enterprise's stock is used as a criterion to determine whether the enterprise has a financial crisis. That is, when the stock of a listed enterprise is marked as ST (or *ST), it is considered that the enterprise suffers from a financial crisis; if there is no ST (or *ST) marking, it is considered that the enterprise's finance is in a state of slight abnormality or no abnormality. This paper selected 360 groups of data in the number of three sample groups is too large a gap, and such a large gap will lead to even if the occurrence of extreme situations, the warning effect will be seriously affected, so even if the prediction accuracy of such models is very high, but the practical significance is not very large. In order to better learn the sample features and train the model to improve the classification effect, it is necessary to re-collect the unbalanced data. Methods can use undersampling or oversampling.

Under-sampling, i.e., following certain rules to delete some of the majority class samples in order to balance the data and reduce the negative impact caused by data imbalance. Common methods include random undersampling, EasyEnsemble, BalanceCascade, NearMiSS, and so on. Oversampling, i.e., according to certain rules, increase some minority class samples to achieve sample size balance. Common methods include random oversampling, SMOTE, ADASYN, etc.

The resampling method selected in this paper is oversampling, the specific method is SMOTE, i.e., selecting the sample b that is closest to any minority class sample a, and then selecting random samples on the line of ab and adding them to the original minority class sample group to form a new minority class sample group. In this paper, the SMOTE algorithm is implemented by the SMOTE algorithm in the Imblearn library, which is a Python library specialized in handling unbalanced data.

III. B. Analysis of indicators

III. B. 1) Descriptive statistics

Descriptive statistics for the indicators of different samples, through the distribution state and the mean, some indicators can clearly see the difference between different samples. Take current ratio and return on assets as an example to show descriptive statistics. Figures 4 to 5 show the descriptive statistics of the indicators of current ratio and return on assets for different samples. The current ratio of financial crisis enterprises is lower overall, less than

2 accounted for more than 70%, the current ratio of enterprises with unstable financial situation is mostly in the vicinity of between 1.5-2, current assets can meet the requirements of current liabilities, in a conservative state; while the current ratio of enterprises with good financial situation is in the vicinity of the overall 2, with strong liquidity. Return on assets, from the mean value seems to be lower in the financial crisis enterprises than in the unstable financial situation of enterprises lower than the financial health of enterprises. The overall return on assets of enterprises in financial crisis is less than 0.5; the return on assets of enterprises in unstable financial condition is slightly more than 0.5; and the overall return on assets of enterprises in good financial condition is more than 0.5. The descriptive statistics of the index data can intuitively see the differences of enterprises in different financial conditions.

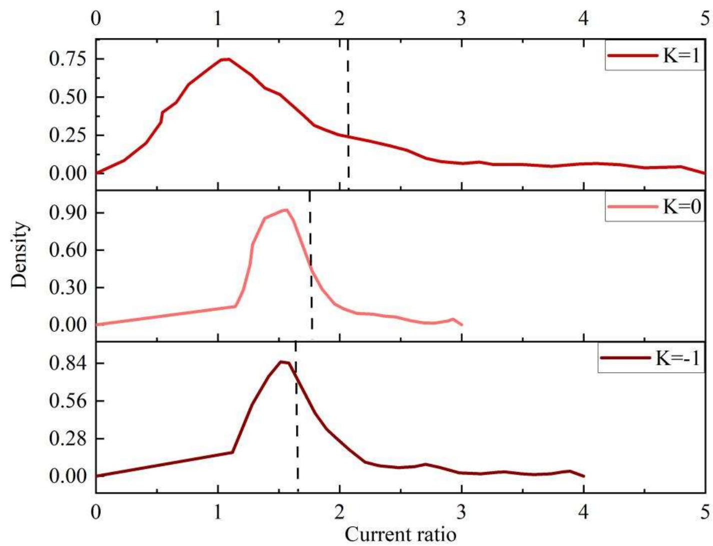

Figure 4: Descriptive statistics of current ratio

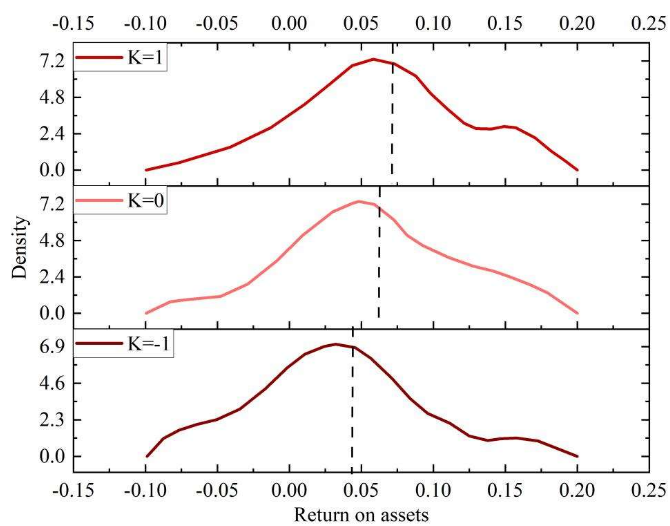

Figure 5: Descriptive statistics of return on assets indicators

III. B. 2) Correlation analysis

In order to retain as much information as possible about early warning, the indicators were selected in a comprehensive manner, so that there may be obvious correlations between some of the indicators. In order to judge the correlation of the indicators selected in this paper, the correlation analysis chart between the indicators is drawn. Figure 6 is the matrix of the correlation coefficient of the indicators, which shows the correlation between the indicators more graphically and intuitively through the change of color. From the indicator correlation coefficients, it can be seen that the correlation coefficients between the indicators are less than 0.5, indicating that there is no strong correlation, so the indicators selected in this paper can be retained without additional elimination and change.

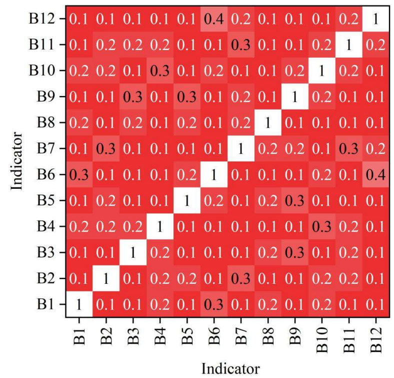

Figure 6: Index correlation coefficient matrix

III. C. Optimal parameterization of the model

For the KNN financial early warning model, there are two main parameters to be determined: the environmental parameter B for calculating the gray correlation and the threshold percentage P for determining the set of k-nearest-neighbor cases.In order to determine the optimal parameter values, a grid-searching technique is used, where the parameter B is in the sequence [0.2:0.06:1], and the parameter P is in the sequence [0.5:0.2:1], and the early warning corresponding to each pair of possible parameter values is calculated in a loop accuracy, and take the parameter value corresponding to the highest warning accuracy as the final parameter value. The scatterplot is constructed by taking two parameter values corresponding to the highest accuracy rate for each year of data from 2020 to 2022. Figures 7, 8, and 9 show the scatter distributions associated with P and B for the years 2020, 2021, and 2022, respectively. Thus, there are two pairs ([0.791,0.208] and [0.823,0.250]), five pairs ([0.938,0.298], [0.950,0.652], [0.950,0.748], [0.962,0.755] and [0.962,0.900]), and two pairs ([0.782,0.203] and [0.884,0.450]) optimal parameter values. From this, it can be determined that the empirical range of optimal values of B is relatively large, with possibilities between [0.2,0.9], while the empirical range of optimal values of P is more concentrated, mainly distributed between [0.7,1].

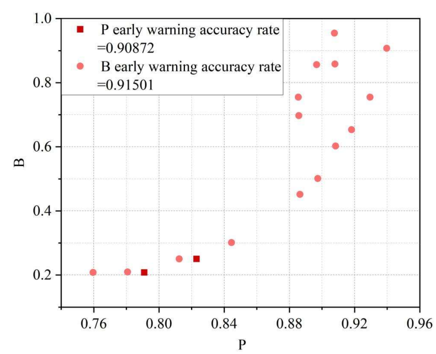

Figure 7: The scattered distribution of b and p in 2020

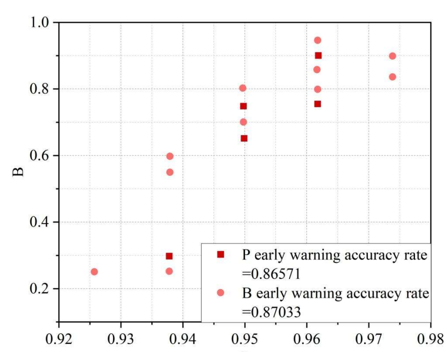

Figure 8: The scattered distribution of b and p in 2021

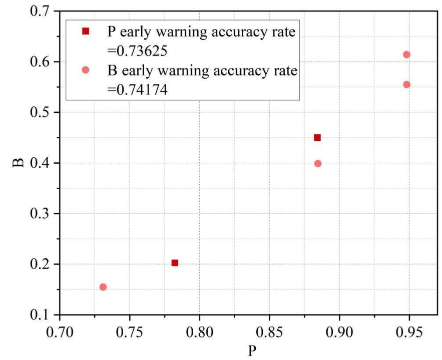

Figure 9: The scattered distribution of b and p in 2022

III. D. Comparison of financial early warning model performance

In this section, the performance of the financial early warning model based on the K-neighborhood algorithm is compared with other classifier financial early warning models to determine the model's ability to classify data and the level of early warning performance. Table 2 shows the performance comparison results of the multi-classifier financial early warning model. From the financial early warning results, the prediction accuracy of this paper's financial early warning model based on the K-neighborhood algorithm reaches 84.94%, which is higher than that of the other nine models of 68.86%, 68.87%, 71.10%, 80.13%, 77.76%, 48.84%, 68.87%, 77.75%, and 62.20%. The model in this paper can better accomplish the financial early warning of enterprises in the era of digital transformation and reduce the risk of errors in the capital chain.

Table 2: Performance Comparison of Financial Early Warning Models

Model	Maximum profit	Maximum profit occurrence ratio	Improvement Degree (Top30)	Overall accuracy	ROC	Prediction accuracy
Logistic regression	183	31	3.130	97.342%	0.986	68.86%
Bayesian network	122	33	2.661	88.495%	0.941	68.87%
Discriminant	111	25	2.562	87.613%	0.934	71.10%
Neural network	84	23	2.284	80.536%	0.844	80.13%
KNN algorithm	61	15	2.120	78.761%	0.870	77.76%
SVM	96	16	2.092	53.984%	0.772	48.84%
Quest (Q)	-1.840	1	1	68.140%	0.500	68.87%
C&R tree (R	134	36	2.644	92.036%	0.932	77.75%
CHAID (C)	152	32	2.977	95.577%	0.991	62.20%
KNN	167	32	3.298	100%	1	84.94%

IV. Conclusion

This paper constructs an enterprise financial early warning model based on the K-neighborhood algorithm and verifies its prediction accuracy advantage. The selected indicators can well distinguish the sample data of 3 types of financial risk enterprises. The correlation coefficients between the indicators are less than 0.5, and the selected indicators are reasonable through the correlation test. Parameter B is calculated by the sequence [0.2:0.06:1] and parameter P is calculated by the sequence [0.5:0.2:1], and the optimal accuracy is calculated cyclically, and it is determined that the optimal range of value of B is [0.2,0.9], while the optimal range of value of P is in the range of [0.7,1]. The prediction accuracy of the model in this paper reaches 84.94%, which is higher than the comparison model. Combined with the K-neighborhood algorithm can more accurately complete the enterprise financial risk early warning to avoid the enterprise falling into the crisis of capital chain breakage. In the future, the model's multi-indicator data processing accuracy can be further improved to reduce the possible risk of data omission or misclassification and improve the level of enterprise digital transformation.

References

[1] Yang, C. (2024). Analysis on Risk Control of Enterprise Capital Management. Asia Pacific Economic and Management Review, 1(4), 7-18.
[2] Berry-Stölzle, T. R., & Xu, J. (2018). Enterprise risk management and the cost of capital. Journal of Risk and Insurance, 85(1), 159-201.
[3] Andriushchenko, K., Liezina, A., Lavruk, V., Sliusareva, L., & Rudevska, V. (2022). Intelligent enterprise capital control based on Markov chain. Acta Innovations, 45, 18-30.
[4] Wang, W., & Yi, H. (2024). Risk Prevention and Countermeasure Analysis in Enterprise Capital Operation. Asia Pacific Economic and Management Review, 1(4), 95-111.
[5] Xie, W. (2022). Study on enterprise financial risk prevention and early warning system based on blockchain technology. Mobile Information Systems, 2022(1), 4435296.
[6] Qu, Q., Liu, C., & Bao, X. (2021). E-Commerce Enterprise Supply Chain Financing Risk Assessment Based on Linked Data Mining and Edge Computing. Mobile Information Systems, 2021(1), 9938325.
[7] Shad, M. K., Lai, F. W., Fatt, C. L., Klemeš, J. J., & Bokhari, A. (2019). Integrating sustainability reporting into enterprise risk management and its relationship with business performance: A conceptual framework. Journal of Cleaner production, 208, 415-425.
[8] Li, B., Di, M., Wei, Z., Qiao, H., & Li, X. (2022). Enterprise Financial Early Warning Based on Improved Whale Optimization Algorithm: Optimize the Perspective with Indicators. Scientific Programming, 2022(1), 1457547.
[9] Wen, C., Yang, J., Gan, L., & Pan, Y. (2021). Big data driven Internet of Things for credit evaluation and early warning in finance. Future Generation Computer Systems, 124, 295-307.
[10] Zhu, Y., Hu, Y., Liu, Q., Liu, H., Ma, C., & Yin, J. (2023). A Hybrid Approach for Predicting Corporate Financial Risk: Integrating SMOTE-ENN and NGBoost. IEEE Access, 11, 111106-111125.
[11] Li, X. (2022). Study on Early Warning on the Financial Risk of Project Venture Capital through a Neural Network Model. International Journal of Advanced Computer Science and Applications, 13(9).
[12] Ma, Z., Wang, X., & Hao, Y. (2023). Development and application of a hybrid forecasting framework based on improved extreme learning machine for enterprise financing risk. Expert Systems with Applications, 215, 119373.
[13] Wu, M., Coleman, M., & Bawuah, J. (2020). The predictive power of K-nearest neighbor (KNN): The effect of corporate governance mechanisms on earnings management. Sage Open, 10(3), 2158244020949537.
[14] Coleman, M., & Wu, M. (2020). Combination of ant colony optimization and K-nearest neighbours: the influence of working capital management on corporate performance. Management and Labour Studies, 45(4), 395-415.
[15] Cheng, C. H., Chan, C. P., & Sheu, Y. J. (2019). A novel purity-based k nearest neighbors imputation method and its application in financial distress prediction. Engineering Applications of Artificial Intelligence, 81, 283-299.
[16] Judijanto, L., Sihotang, J., & Simbolon, A. P. H. (2024). Early warning systems for financial distress: A machine learning approach to corporate risk mitigation. International Journal of Basic and Applied Science, 13(1), 14-27.
[17] Min, R. (2023). Influence of technology optimisation based on machine learning algorithm on financial management innovation of e-commerce enterprises. International Journal of Computational Systems Engineering, 7(2-4), 124-134.