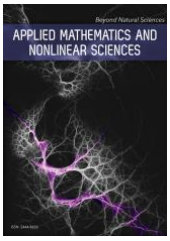
sciendo Applied Mathematics and Nonlinear Sciences, 9(1) (2024) 1-17



Applied Mathematics and Nonlinear Sciences

https://www.sciendo.com

Research on the Application of Data Mining Techniques in Early Warning Models for Financial Management

Wei Su[1,†]

1. TaiYuan Tourism Vocational College, Taiyuan, Shanxi, 030006, China.

Submission Info

Communicated by Z. Sabir
Received March 7, 2024
Accepted May 9, 2024
Available online June 3, 2024

Abstract

In the process of accelerating enterprise development in China, enterprises are prone to financial management crises while grasping development opportunities. This paper is oriented towards enterprise financial management data and proposes an adaptive neuro-fuzzy inference system to reason about risky data. The ANFIS model is integrated into K-mean clustering to provide early warning of enterprise risk by dividing different financial management risk indicators. The validation method uses both the comparison of similar algorithms and the practical test of A enterprise. Compared to other analysis models, the accuracy of the early warning model constructed by the K-Means algorithm is as high as 99.64%, which is suitable for enterprise financial management early warning. Using the model built in this paper to cluster the types of enterprise financial risk, the second cluster centers clusters of enterprises in financial difficulties, a number of indicators show negative results, and the net sales interest rate is as low as -94.23. This indicator is applied to the financial crisis of Company A's financial situation analysis and found to be consistent with the characteristics of the second cluster center. This verifies the accuracy of the financial early warning model constructed in this paper.

Keywords: ANFIS model; K-mean clustering; Risk early warning; Financial management.
AMS 2010 codes: 68T05

†Corresponding author.
Email address: 13613485479@163.com

ISSN 2444-8656
https://doi.org/10.2478/amns-2024-1301

1 Introduction

After years of development, the profitability level of enterprises has been improved year by year. Still, with the continuous promotion of China's supply-side structural reform, especially by the impact of the new Crown pneumonia epidemic, the competitive pressure on enterprises is increasing, and the financial risks they face are also becoming more and more prominent [1-2].

Financial management is the foundation of the enterprise and is the core part of the enterprise activities. Doing a good job of enterprise financial management early warning can effectively avoid financial losses and reduce enterprise risk at the same time, it can also be for the prosperity of the enterprise to broaden the road [3-5]. Financial management early warning refers to the marketing project, financial statements, and other data, the use of comparative analysis, ratio analysis, factor analysis, and other analytical methods, all the data occurring in the enterprise activities to analyze the pre-judgment of enterprise activities in the presence of risk, data charts and graphs to show to the enterprise managers [6-7].

Financial management early warning can not only pre-judge the financial risk but also business managers to provide an effective basis for business decision-making and resource allocation to amend the direction of business operations [8-9]. With the development and growth of enterprises, financial data is increasing, and it is a difficult problem to find its relevance in many financial data and predict the potential problems of enterprise operations [10].

Currently, widely used data mining technology can be automatically searched from the huge amount of data hidden in which has a special relationship between the data, which has the advantages of short time, high accuracy, and can be applied to the financial management of the risk early warning, to realize the financial management of accurate and effective early warning [11]. Literature [12] designed a financial risk prediction framework composed of a k-means++ algorithm and optimized radial basis function neural network as the core logic and based on simulation experiments to test the framework, confirming the feasibility of the proposed risk prediction framework. Literature [13] proposes a financial risk early warning model based on the theory of kernel fuzzy dual support vector machine (KFT), which has better performance and stability than the VM risk early warning model. Literature [14] combines data mining technology, least absolute shrinkage, and selection operator (LASSO) selection strategy to build an enterprise financial risk prediction system with high prediction accuracy, which effectively improves the accuracy and reasonableness of the prediction of enterprise financial risk. Based on empirical analysis, literature [15] reveals the association between enterprise size, diversification background, and the industry in which the enterprise is located and enterprise risk management and points out that the higher the financial leverage, the lower the degree of enterprise risk management application.

The occurrence of enterprise financial risk belongs to a long-term process, which is formed by the influence of multiple internal and external factors. In the face of increasingly complex competition in the market environment, the financial risk management of enterprises is under increasing pressure. Literature [16] explains that enterprise risk management has the role of reducing the company's operating risks and improving the company's economic efficiency, and cites seven risky operations of the company as a framework for building an enterprise risk management analytical framework, in order to make clear what kind of impacts different types of enterprise risks have on the operation of the company and market competition. Literature [17] explored the association between operational capability and sustainable risk management in Palestinian insurance firms and pointed out that the strength of ERM capability significantly affects the profitability of firms. Literature [18] innovatively used corporate return on equity as an indicator of a firm's operational capability, revealing the significant impact of risk management on the operational effectiveness of Sri Lanka's diversified

industries. Literature [19] examined the association between organizational culture and enterprise risk management and found that a firm's organizational culture determines the style of risk management, which in turn is determined by the firm's leading managers.

This paper firstly uses the adaptive neural system, analyzes enterprise financial statements, and, according to the results, analyzes the factors leading to enterprise financial management risk, establishes an enterprise financial management early warning indicator system, compares the accuracy of the K-Means model with other models, and verifies the scientificity of the enterprise financial management early warning model constructed in this paper. The descriptive statistics of the overall sample of indicators, according to standard deviation, will be analyzed to analyze the main factors that cause difficulties in enterprise financial management. Then, using the hybrid fuzzy clustering algorithm on 6949 enterprise samples for cluster analysis, it was found that the existence of financial management risks of enterprises are located near the second cluster center. To test the accuracy of the model constructed in this paper, it is substituted into the financial indicators of enterprise A as a specific case to analyze.

2 Enterprise financial management risk early warning model construction

2.1 Adaptive Neuro-Fuzzy Inference System (ANFIS)

In essence, a fuzzy inference system is an application model that uses fuzzy logic-related theories to achieve nonlinear mapping. Due to the difference in theoretical basis, fuzzy inference systems can be divided into three categories: pure fuzzy logic type, T-S type, and Mamdani type. The pure fuzzy logic type has fuzzy sets of inputs and outputs, which cannot be directly applied to social reality, and thus its application is limited. The Mamdani type improves the pure fuzzy logic type by adding fuzzy generators and fuzzy eliminators. In contrast, the T-S type improves the conclusions of the posterior of the fuzzy rules to exact values. The fuzzy set limitation of the pure fuzzy logic type is broken by these two improved types, making them the mainstream of fuzzy inference systems. Financial management early warning can benefit greatly from the use of ANFIS.

1) Mamdani-type fuzzy inference system

Mamdani-type fuzzy inference system well synthesizes fuzzy solutions in complex systems and decision making problems. Mamdani-type fuzzy inference algorithms are used to take small operators as T -module to deal with intersecting operations of different fuzzy sets, such as rules:

$$R: \text{If } x \text{ is } A \text{ then } y \text{ is } B \tag{1}$$

Where: x is the input linguistic variable. A is the fuzzy set of inference antecedents. y is the output linguistic variable. B is the posterior of the fuzzy rule. Denote the fuzzy relationship by Rc:

$$R_c = A \times B = \int_{X \times Y} \mu_A(x) \wedge \mu_B(y) f(x, y) \tag{2}$$

When x is A', if the fuzzy synthesis operation is defined as "take big-small," the conclusion of the fuzzy reasoning is as follows:

$$B' = A' * Rc = \int_Y \vee \left(\mu_A(x) \wedge \left(\mu_A(x) \wedge \mu_B(y) \right) \right) / y \tag{3}$$

In Mamdan-type fuzzy inference systems, each rule's inference results in the distributional affiliation function or discrete fuzzy set of variables. After synthesizing the outputs of multiple rules, each rule's output is defuzzed to obtain the desired output of the actual problem.

2) T-S type fuzzy inference system

Sugeno-type fuzzy inference combines defuzzification with fuzzy inference, and its output is an exact quantity. The form of Sugeno fuzzy rules determines this. The Sugeno-type fuzzy inference algorithm is the most commonly used fuzzy inference algorithm. The Sugeno-type fuzzy inference algorithm is similar to the Mamdani type. A typical zero-order Sugeno-type fuzzy rule has the following form:

$$\text{If } x \text{ is } A \text{ and } y \text{ is } B \text{ then } z = k \tag{4}$$

Where: x and y are input linguistic variables. A and B are fuzzy sets of inference antecedents z is the output linguistic variable. k is a constant. A more general first-order Sugeno modeling rule is of the form:

$$\text{If } x \text{ is } A \text{ and } y \text{ is } B \text{ then } z = px + qy + r \tag{5}$$

Where: x and y are linguistic variables. A and B are fuzzy sets of inference antecedents. z is the output linguistic variable. p, q , r are constants. For a T-S type fuzzy inference system consisting of n rules, each of its regulations has the following form:

$$R_i : \text{If } x \text{ is } A_i \text{ and } y \text{ is } B_i \text{ then } z = z_i \quad (i = 1, 2, \ldots, n) \tag{6}$$

The total output of the system is:

$$y = \sum_{i=1}^{n} \mu_{Ai}(x)\mu_{Bi}(y)z_i \Big/ \sum_{i=1}^{n} \mu_{Ai}(x)\mu_{Bi}(y) \tag{7}$$

Adaptive Neuro-Fuzzy Inference System (ANFIS), also known as Adaptive Network Fuzzy Inference System, belongs to the typical T-S type.ANFIS combines fuzzy logic and neural networks organically in a new fuzzy inference system structure, which has the function of approximating arbitrary linear and nonlinear functions with arbitrary accuracy. If the system has m input quantity $\{a_1, a_2, \cdots, a_n\}$ and the single output quantity can be represented by a set of n fuzzy If-Then rules, the ANFIS structure is shown in Fig. 1.

Layer 1: Fuzzification of input variables. The output of this layer is the degree of affiliation of the fuzzy set, where the transfer function of each node can be expressed as:

$$O_{1,j} = \mu_{Ai}^{k}(x_i) \tag{8}$$

Where, n is the number of fuzzy rules, m is the number of input data variables, $i = 1, 2, \cdots, m, k = 1, 2 \cdots, n, j = 1, 2, \cdots, n, O_{1,j}$, represents the output, and the affiliation function is usually used as a Gaussian function with the following formula:

$$\mu_{Ai}^{k}(x_i) = \exp\left(-\frac{1}{2}\left(\frac{x_i - c_i^k}{a_i^k} \right)^2 \right) \tag{9}$$

Where the set of parameters consisting of all $\left\{a_i^k, c_i^k\right\}$ is the conditional parameter.

Layer 2: Calculate the affiliation of the conditional part, usually by multiplication:

$$O_{2,k} = w_k = \prod_{i=1}^{m} \mu_{Ai}^k \left(x_i\right) \tag{10}$$

Where $k = 1, 2, \cdots, n$.

Layer 3: Normalize the affiliation of each rule:

$$O_{3,k} = \hat{w}_k = \frac{w_k}{\sum_{l=1}^{n} w_l} \tag{11}$$

Layer 4: The transfer function for each node is a linear function that computes the output of each rule:

$$O_{4,k} = \hat{w}_k \times f_k = \hat{w}_k \left(d_0^k + d_1^k x_1 + \cdots + d_m^k x_m\right) \tag{12}$$

The set of parameters consisting of all $d_0^k, d_1^k, \cdots, d_m^k$ is called the conclusion parameter.

Layer 5: Calculate the sum of all rule outputs:

$$O_5 = \sum_{k=1}^{n} \hat{w}_k f_k \tag{13}$$

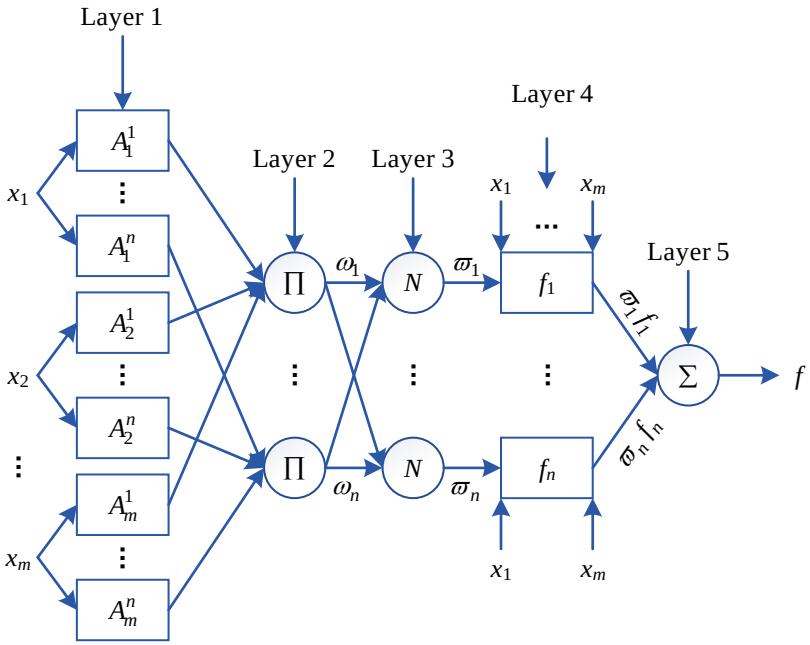


Figure 1. The ANFIS structure with m-dimension input

2.2 Hybrid fuzzy clustering algorithm construction

2.2.1 K-mean clustering algorithm

The k-mean clustering algorithm in the distance measure can choose a variety of forms, such as the Euclidean distance, Manhattan distance, Minkowski distance, etc. The use of different distance measures is only in the calculation of distance in that step of the calculation method. The rest of the algorithmic steps are the same. Which distance method to use in practical applications should be selected according to the characteristics of the actual application. Applicable to the cluster analysis of financial management early warning models.

The metric used in the K-means clustering algorithm to optimize the location of the center of mass is the intra-cluster error variance. For a specified k cluster, the smaller the distance between points within the cluster, the correspondingly smaller the SSE, and the better the clustering.

Let the data set be $A=\{a_1,a_2,\ldots a_i\}$, where a_1 and $a_2\ldots a_i$ denote the i data records in the data set, respectively. The initial center of mass is randomly generated for the first time, denoted by $K_1=\{K_1,K_2,\ldots Kj_1\}$, where Kj denotes the j th center of mass and Kj_1 denotes the point where Kj is located for the first time. After the center of mass is generated, traverse from a_1, a_2 to a_i to calculate their respective distances to $K1_1,K2_1,\cdots Kj_1$, and determine a_i which distance from Kj_1 is the closest to divide a_i into clusters centered on that Kj_1. Calculate SSE_1 for the first time as in equation (14), which is the sum of the squares of the distances a_i from the center of mass Kj_1 that one is at:

$$SSE_1=\sum_{m-1}^{j}\sum_{a_{m1}\in Km_1}\left\|a_{m1}-Km_1\right\|^2 \tag{14}$$

Where SSE_1 denotes the first SSE, m denotes the variable representing the number of centers of mass at the time of programming, which takes integer values from 1 to j, a_{m1} denotes all the data belonging to the m th center of mass after the first center of mass determination, Km_1 denotes the m th center of mass determined for the first time, and $a_{m1}\in Km_1$ denotes that a_{m1} is classified into clusters centered on the center of Km_1.

The mean value of the points within each cluster is calculated, and it is taken as the new center of mass K_2:

$$K_2=\{K1_2,K2_2,\ldots Kj_2\} \tag{15}$$

$$Kj_2=mean(a_{m1}) \tag{16}$$

The $mean()$ in Eq. (16) represents the mean value function, by which the center of mass K_2 of the second time can be determined, and the data in A are reassigned to the new cluster according to the principle of closest distance to calculate the SSE_2 of the second time:

$$SSE_2=\sum_{m-1}^{j}\sum_{a_{m2}\in Km_2}\left\|a_{m2}-Km_2\right\|^2 \tag{17}$$

Calculate the gap ΔSSE_2 between SSE_2 and SSE_1:

$$\Delta SSE_2 = \left| SSE_1 - SSE_2 \right| \tag{18}$$

To determine whether ΔSSE_2 meets the error requirement, set δ as a set positive number, which is the interval value for error judgment, and determine it according to the actual situation. If ΔSSE_2 is less than δ, that is:

$$\Delta SSE_2 < \delta \tag{19}$$

Then, the clustering error is considered to satisfy the accuracy requirement, and the algorithm stops. If it does not meet the criteria, then calculate the mean value of the points in each cluster with K_2 as the center of mass to form a new center of mass K_3, according to the principle of closest distance to the data in A and then assign it to the new cluster did not calculate the third time SSE_3, calculate ΔSSE_3 and determine whether to terminate or loop again until the end of the algorithm.

The K-mean clustering algorithm's inputs and outputs can be found in Table 1 below.

Table 1. Inputs and outputs of the K-mean clustering algorithm

Type	Content
Input	Data set A, number of clusters k
Output	k cluster centroids

The flowchart of the K-mean clustering algorithm is shown in Figure 2 below.

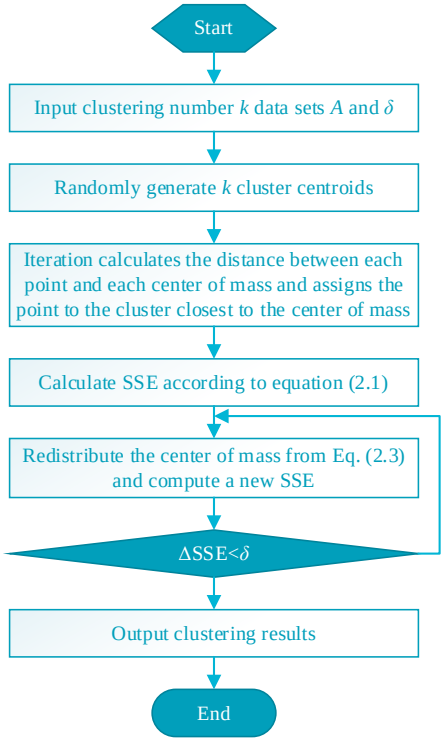


Figure 2. Flowchart of K-mean clustering algorithm

2.2.2 Fuzzy Clustering Algorithm

First, recall the fuzzy clustering algorithm, represented by the fuzzy C-means algorithm. The FCM algorithm is a data division method that gives the objective function of clustering, and then is based on iterative optimization of the objective function. The clustering center v is calculated by the objective function. The results of sample point attribution in the clustering process are estimated by calculating the distance from each data point to the center point v, and then analyzing and calculating the degree of affiliation of the end to the clustering center region module, which can be expressed by solving the objective function for a specific value of the regional center. Optimizing the weighted inertia criterion involves expressing the objective function as:

$$\begin{gathered} J_{FCM}(U,V)=\sum_{i=1}^{C}\sum_{j=1}^{N}\mu_{ij}^{s}d_{ij}^{2} \\ d_{ij}^{2}=\left\|x_j-v_i\right\|^2 \\ s.t.\mu_{ij}\in[0,1],\sum_{i=1}^{C}\mu_{ij}=1,1\le j\le N \end{gathered} \tag{20}$$

Equation (20) above, C denotes the number of clusters. $v_i=\left(v_{a1},v_{i2},\ldots,v_{ak}\right)$ denotes the center point of the i rd cluster class. Where the fuzzy affiliation is represented by μ_{ij}, which specifically indicates the meaning of the fuzzy affiliation of the j th object belonging to the i th class. The similarity fuzzy index δ in the clustering calculation process must satisfy $\delta>1, x_j$ denotes as the j th sample point in the data set and N denotes as the size of the data set.

Weighted k-means is proposed as an extension of k-means by adding the calculation of variable weights through an iterative process. Equation (21) is the objective function of WKM:

$$J_{WKM}\left(U,V,W_m\right)=\sum_{i=1}^{C}\sum_{j=1}^{N}\mu_{ij}^{s}d_{ij}^{2}\sum_{l=1}^{M}w_{lj}^{\beta} \tag{21}$$

Where $\beta>1$ or $\beta<0$, β is a power parameter of feature weights. However, it has been shown that it cannot achieve good results in the clustering process of high-dimensional data.

2.2.3 Hybrid fuzzy clustering algorithm

Although the commonly used FCM algorithm and K-means algorithm can solve the problem of determining the clustering center better, they do not guarantee the best clustering results in the complicated real world. Niros and Tsekouras2012 proposed the HFCM algorithm. The objective of this algorithm is to achieve the following:

$$J=\theta\sum_{j=1}^{n}\sum_{i=1}^{c}u_{ij}d_{ij}+(1-\theta)\sum_{j=1}^{n}\sum_{i=1}^{c}u_{ij}^{2}d_{ij}^{2} \tag{22}$$

Where $\theta\in[0,1]$ is the degree of affiliation of the result to the clear domain, the larger θ is, the higher the degree of affiliation to the clear domain, the faster the convergence speed. The algorithm utilizes the alternating optimization algorithm, which is solved as follows:

$$u_{ij} = \frac{2+(c-2)\theta}{2(1-\theta)} \cdot \frac{1}{\sum_{k=1}^{c} \frac{d_{ij}^2}{d_{kj}^2} - \frac{\theta}{2(1-\theta)}} \tag{23}$$

$$v_i = \frac{\sum_{j=1}^{n} \left[\theta u_{ij} - (1-\theta)^2 u_{ij}^2 \right] x_k}{\sum_{j=1}^{n} \left[\theta u_{ij} - (1-\theta)^2 u_{ij}^2 \right]} \tag{24}$$

FCM is similar to the specific steps. The HFCM algorithm has greater adaptability than the traditional FCM algorithm by adjusting parameter θ.

Step 1: Initialize the affiliation matrix U with a random number between 0 and 1.

Step 2: Calculate c clustering centers $v_i, i = 1, \ldots, c$.

Step 3: Calculate the objective function according to Eq. Terminate the iteration when the value of the objective function or the change in the objective function is less than some determined threshold.

Step 4: Calculate the new U matrix. Return to step 2.

2.3 Combined model construction based on hybrid fuzzy clustering and ANFIS

For n set of m-dimensional sample data pairs $\left(X_1, X_2, \cdots, X_n\right)$, a $(m-1)$-input single-output model is built, and the ith set of data can be described as $X_i = \left(x_i(1), x_i(2), \cdots, x_i(m)\right)$. Where the first $(m-1)$ dimensions are used as inputs, and the last 1 dimension is used as an output. Cluster the n $(m-1)$-dimensional data pairs using HFCM (hybrid clustering method) to obtain c $(m-1)$-dimensional cluster centers where the p th can be represented as $c_p = \left(c_p(1), c_p(2), \cdots, c_p(m-1)\right), p = 1, 2, \cdots, c$. By c clustering centers c rules can be produced where the p th rule is described as follows:

$$\text{IF } X_i \text{ is } u_{pt}, \text{ then } y \text{ is } f_{pi} \tag{25}$$

Where $X_i = \left(x_i(1), x_i(2), \cdots, x_i(m-1)\right)$ is the input variable for the i nd set of samples and y is the corresponding output:

$$\mu_{p,i} = \exp\left(-\left\| x_i - c_p \right\|^2 / \sigma_i^2 \right) \tag{26}$$

Is the affiliation of the i st data sample for the p nd rule. The final output is:

$$y_i = \frac{\sum_{p=1}^{c} u_{p,i} f_{p,i}}{\sum_{p=1}^{c} u_{p,i}} \tag{27}$$

Where $f_{p,i} = k_{p,1}x_i(1) + k_{p,2}x_i(2) + \cdots + k_{p,m}x_i(m-1) + k_{p,m}$ is the output of rule p.

3 Enterprise financial management risk early warning results and analysis

3.1 Comparative analysis of the performance of different models

To evaluate the financial management early warning model that utilizes hybrid fuzzy clustering and ANFIS proposed in this paper. Using the Logistic model, SVM model, BP neural network, XGBoost model, and this paper's model to test the enterprise financial management early warning, respectively, in the selection of indicators are selected to optimize the indicator group, the same samples as well as using the same method to deal with the unbalanced data, and will be divided into the training set and the test set according to the ratio of 4:1, and finally, get based on the Logistic model The prediction results of corporate financial risk identification models based on Logistic model, SVM model, BP neural network, XGBoost, and K-Means clustering type are finally obtained. Table 2 shows the comparison of the five models.

The prediction accuracy of the enterprise financial management early warning model constructed using the K-Means algorithm is the highest, and the overall accuracy of prediction is as high as 99.64%. The BP neural network model is the second highest, with an accuracy of 94.56%, and the worst prediction is the SVM model, with a prediction accuracy of only 79.82%. Overall, the K-Means model is optimal in terms of TPR, FPR, Accuracy, f1-score, and AUC values for corporate financial management risk warning. Compared with the prediction accuracy of the Logistic model, which is most widely used in commercial banks, the K-Means model improves by 12.3% in prediction accuracy. Compared with the XGBoost model, its prediction accuracy improves by 7.82%, through the above comparisons, we can conclude that the K-Means model is applicable in the early warning of corporate financial management.

Table 2. Comparison of 5 Models

	TPR	FPR	Accuracy	f1-score	AUC
Logistic	0.8214	0.1938	0.8734	0.7834	0.9278
SVM	0.9116	0.2317	0.7982	0.8813	0.8341
BP Neural Networks	0.7742	0.1764	0.9328	0.9456	0.9268
XGBoost	0.9423	0.2579	0.9270	0.9182	0.6541
K-Means	0.9923	0.0013	0.9964	0.9964	0.9987

3.2 Analysis of enterprise financial management risk factors

The analysis of enterprise financial management risk factors on the basis of this paper will be based on the comprehensiveness, availability, and scientific basis, respectively, from the macro-factors, non-financial features, and financial features selected 19 indicators, as the basis for subsequent screening, Table 3 for the enterprise financial management risk factor indicators.

Table 3. Index of Corporate Financial Risk Factors

Index System		Index Symbol
Solvency	Current ratio	X1
	Quick ratio	X2
	Gearing ratio	X3
Profitability	Net sales margin	X4
	Return on net assets	X5
	Return on Total Assets	X6
Operating Capability	Fixed Assets Turnover Ratio	X7
	Inventory Turnover Ratio	X8
Development Capacity	Total Assets Growth Rate	X9
	R&D Expense Growth Rate	X10
Macroeconomic	GDP Growth Rate	X11
	Inflation Rate	X12
	Unemployment Rate	X13
Industry Environment	Industry Concentration	X14
	Industry Sales Growth Rate	X15
	Industry Lerner Index	X16
Other Non-Financial Indicators	Number of Directors	X17
	Percentage of outstanding shares	X18
	TONE2	X19

The sample for this paper is selected from 6439 enterprises in China, divided into normal and financial high-risk samples. The normal sample is the enterprises with low financial risk in 2019-2023, and the high-risk sample is the enterprises with high financial risk in 2019-2023. Sample characteristics are extracted according to Table 3, and 6349 enterprise samples are statistically analyzed. Table 4 shows the descriptive statistics of the overall sample.

In terms of macroeconomics, the mean values of GDP growth rate, inflation rate, and unemployment rate are 3.58%, 1.46%, and 1.95%, respectively, and the standard deviations are 0.32, 0.36 and 0.39, respectively, which indicate that the macroeconomics environment is relatively stable during 2019-2023. In terms of corporate financial factors, the mean values of quick ratio and current ratio are 1.61% and 2.07%, respectively. Their corresponding standard deviations are 2.02 and 3.03, which shows that the short-term solvency of the enterprises in the sample is roughly the same. Still, there is a huge difference in terms of the growth capacity, with standard deviations of all types of indicators above double digits.

Table 4. Descriptive statistics for the overall sample

Statistic	Average	Std.	Min	Medium	MAX
X1	1.61	2.02	-8.38	21.13	45.94
X2	2.07	3.03	-131.57	47.34	55.06
X3	7.73	0.31	-3.03	13.84	54.37
X4	2.65	480.45	-96.28	55.44	102.2
X5	25.34	32.76	-47.42	66.52	25.96
X6	73.09	19.6	-115.4	18.75	55.04
X7	22.39	18.36	-19.08	15.94	85.85
X8	68.81	9.26	-57.24	42.11	20.21
X9	22.15	913.69	-7.89	66.25	58.47
X10	45.01	824.82	-76.77	59.53	54.43
X11	3.58	0.32	-114.42	61.16	114.9
X12	1.46	0.36	-16.44	43.37	38.66
X13	1.95	0.39	-22.12	23.39	24.27
X14	2.14	1.13	-81.43	30.33	110.98
X15	83.42	22.1	-52.22	33.15	5.02
X16	22.54	6.39	-110.06	14.6	39.7
X17	92.34	5.66	-45.8	42.94	22.34
X18	60.32	4.97	2.75	39.53	63.22
X19	8.23	1.24	-113.11	53.75	40.43

3.3 Cluster analysis of corporate financial management risks

The financial management risk early warning model proposed in this paper is used to cluster and analyze the dataset. The second step involves recording cluster clusters and cluster centers and analyzing the financial management risk status of each type of cluster center and cluster cluster. Existing financial distress-related studies generally only classify the financial status of enterprises into ST and non-ST categories. Some scholars have conducted empirical studies on the degree of financial distress of forestry enterprises and communication enterprises. The results show that the division of financial distress into three to five categories is more in line with the actual financial status of listed companies. Considering that the division into five categories may lead to the results by avoiding the impact of only one or two enterprises in certain categories, which is not conducive to the analysis of the clustering results, this paper formulates the financial situation of the sample enterprises into three categories and then adjusted according to the clustering results, which can avoid the impact of extreme values as much as possible. After using SPSS26.0 software for cluster analysis, Table 5 shows the initial clustering results. The majority of the 6439 enterprise samples are grouped in the first clustering center with 3346 enterprises, followed by the third clustering center with 1968 enterprises, and lastly, the second clustering center with 1125 enterprises.

Table 5. Initial clustering results

Cluster Center	First Cluster Center	Second Cluster Center	Third Cluster Center
Number of clusters	3346	1125	1968

The number of cases clustered in the first clustering center is the largest, and the number of cases in the third clustering center is the second largest. Most ST companies are classified into the second clustering center, which indicates that the sample data in the second category are facing the most serious financial difficulties. The analysis of clustering centers and the degree of distress discrimination is shown in Table 6.

The difference between the three clustering centers is obvious: the first clustering center, no matter the solvency, profitability, operating ability, development ability, or the industry environment in which it is located, as well as the trend of the corporate annual report disclosure of the sentiment, all show a higher or better level, the average value of the quick ratio and the current ratio is 3.383% and 3.339%, respectively, which indicates that the enterprise has a better short-term solvency. The net sales margin and return on net assets reached 2.61% and 9.404%, respectively, indicating that the enterprises have strong profitability. Overall, the enterprises in the clustering and the first clustering center belong to enterprises with better financial status, and the possibility of falling into financial difficulties is low. The indicators of the third clustering center lag behind the first clustering center as a whole. Still, they are better than the second clustering center, in which the profitability shows a higher level, the net sales interest rate of 4.177%, while the debt servicing ability and the development ability of the performance of people with low incomes, so that the concentration of the clustering center of the enterprise's financial position is medium, the possibility of financial difficulties is low, but still need to be vigilant about the occurrence of financial problems. The second clustering center of the indicators than the first clustering center and the third clustering center have a large gap, profitability, development capacity indicators are produced a negative result, the net sales interest rate of enterprises is as low as -94.23% shows that the second clustering center of the enterprises is clustered in the financial difficulties of the enterprise.

Table 6. Initial clustering results

Index		First Cluster Center	Second Cluster Center	Third Cluster Center
Solvency	X1	3.383	2.959	1.66
	X2	3.399	0.161	0.814
	X3	1.727	-41.54	2.714
Profitability	X4	2.61	-94.23	4.177
	X5	9.404	-2.297	2.113
	X6	2.112	-8.677	0.686
Operating Capability	X7	2.259	-60.299	2.923
	X8	0.64	-30.864	2.485
Development Capacity	X9	3.279	4.532	2.652
	X10	2.47	-56.374	1.839
Macroeconomic	X11	4.054	2.965	1.397
	X12	0.838	7.846	2.02
	X13	0.676	2.523	0.577
Industry Environment	X14	0.768	-76.023	0.585
	X15	3.258	-5.692	0.851
	X16	6.036	-68.265	3.212
Other Non-Financial Indicators	X17	4.151	3.355	3.229
	X18	3.637	-8.487	1.897
	X19	2.559	1.434	0.798

Through the financial management risk model established in this paper, the enterprise can be identified for financial distress. If the clustering results are clustered with the first clustering center, it means that the enterprise's financial situation is better, and the possibility of falling into financial distress is lower. Clustering results are clustered in the second cluster center, indicating that the enterprise has fallen into financial difficulties and needs to analyze the weakness of the performance of the financial system by adjusting the strategy or the implementation of targeted measures to improve the financial situation of the enterprise to avoid bankruptcy. Cluster results clustered in the center of the fifth cluster indicate that the enterprise's financial situation is average. However, the possibility of financial distress is lower but more likely than the first cluster of enterprises in financial distress. Hence, we still need to be vigilant about the occurrence of financial distress.

3.4 Case Study of Early Warning Models for Corporate Financial Management

This paper selects Company A, which experienced a financial crisis from 2019 to 2023, as a case study. The indicator data from 2019 to 2023 are chosen to identify financial distress based on the enterprise financial risk factor indicators constructed in this paper. Table 7 shows the data for 19 financial distress identification indicators from 2019 to 2023 for Company A. The level of net sales interest rate of the enterprise from 2019 is low at 0.531%. The sales interest rate of the enterprise is negative from 2020 to 2023, which indicates that Company A is in a loss position for a long period, with a minimum of -100.435 in 2021.

Table 7. Company A Financial Distress Identifying Indicator Data

Index		2019	2020	2021	2022	2023
Solvency	X1	0.477	0.859	1.255	0.965	0.719
	X2	0.746	0.6864	0.863	0.578	0.453
	X3	29.836	65.466	78.578	100.678	98.567
Profitability	X4	0.531	-19.333	-100.435	-59.754	-78.345
	X5	3.439	15.257	-262.978	-324.861	-55.223
	X6	6.092	18.55	0.359	-32.156	-68.478
Operating Capability	X7	8.141	1.433	0.456	0.413	0.815
	X8	0.713	0.456	0.892	0.789	0.349
Development Capacity	X9	-24.682	-11.731	-35.819	-57.321	-43.899
	X10	3.535	98.356	69.227	-43.621	83.876
Macroeconomic	X11	5.356	7.738	6.324	6.297	5.245
	X12	3.634	3.976	2.038	1.834	3.117
	X13	4.226	3.234	3.003	3.007	4.810
Industry Environment	X14	0.234	0.873	0.239	0.549	0.287
	X15	13.542	14.345	18.981	19.334	15.023
	X16	0.346	0.123	0.496	0.389	0.128
Other Non-Financial Indicators	X17	8.000	8.000	8.000	8.000	8.000
	X18	76.646	88.231	90.222	93.128	93.642
	X19	0.822	0.235	0.451	0.314	0.878

The data of the financial distress identification index system of Company A from 2019 to 2023 are inputted into the model, and Table 8 shows the model's output results.

The results show that the clustering result in 2019 is the second clustering center, and the distance to the clustering center is 126.458. And Company A was in financial distress in 2019, and the distance to the clustering center of financial distress is very close. The clustering result in 2020 is the third clustering center, and the distance to the clustering center is 515.657. And although Company A has not been included in the classification of financial distress, it is still not in the classification of normal financial status. In 2021, the clustering result is the second clustering center, and the distance to the clustering center is 125.789, in this year, company A is in financial distress again. 2022, the clustering result is the second clustering center, and the distance to the clustering center is 198.674. However, the distance to the financial distress clustering center is enlarged, and the distance to the clustering center is 198.674. However, the distance to the financial distress clustering center is enlarged. The distance to the financial distress clustering center is enlarged, and the distance to the financial distress clustering center is enlarged. In 2022, the clustering result is the second clustering center; the distance from the clustering center is 198.674. Although the distance from the financial distress clustering center is enlarged, but still unsuccessful in getting out of trouble. The 2023 clustering result is the fourth clustering center, and the distance from the clustering center is 340.762. The distance from the financial distress clustering center is further enlarged, but still not out of trouble. The result is similar to the first financial distress review, the first out-of-distress review and the second financial distress review of Company A analyzed in the previous section, which further supports the accuracy of the financial distress identification model.

Table 8. Model output results

Year	Identification results	Cluster center distance
2019	Second Cluster Center	126.458
2020	Third Cluster Center	515.657
2021	Second Cluster Center	125.789
2022	Second Cluster Center	198.674
2023	Second Cluster Center	340.762

4 Conclusion

Based on digital mining technology, this paper utilizes an adaptive neuro-fuzzy inference system and hybrid fuzzy clustering algorithm to construct an enterprise financial management early warning model, and then applies the model built in this paper to specific enterprise financial management analysis, and draws the following conclusions:

1) Overall, the K-Means model is optimal in the TPR, FPR, Accuracy, f1-score and AUC values in the enterprise financial management risk warning.

2) Macroeconomic level factors are more stable and have less impact on corporate financial management. In terms of corporate financial factors, the mean values of the quick ratio and current ratio are 1.61% and 2.07%, respectively. Their corresponding standard deviations are 2.02 and 3.03, which shows that the short-term solvency of the enterprises in the sample is more or less the same. Still, there is a huge difference in growth capacity, and the standard deviations of all kinds of indexes are in the more than two digits.

3) Using the model constructed in this paper to generate clustering results, the 6349 enterprises are divided into 3 classes, and most ST companies are divided into the second clustering center, which indicates that the sample data in the second class faces the most serious financial

difficulties. The profitability and growth capacity produce negative results, with the net sales margin as low as -94.23.

4) The data of Company A's financial distress identification index system from 2019 to 2023 are inputted into the model. The model results show that Company A is closest to the center of the second clustering, and the distance from the center of the clustering in 2019 is 126.458. The distance expanded to 340.762 in 2023, but it is still not completely out of the woods, which is in line with the fact of Company A's financial distress during this period and illustrates the fact that the paper The financial management early warning model constructed has a scientific nature.

References

[1] Sewchurran, K., Dekker, J., & Mcdonogh, J. (2019). Experiences of embedding long-term thinking in an environment of short-termism and sub-par business performance: investing in intangibles for sustainable growth. Journal of Business Ethics, 157(4), 997-1041.

[2] Anbinder, J. (2018). Selling the world: public relations and the global expansion of general motors, 1922€"1940. Business History Review, 92.

[3] Dounavi, L. E., Dermitzakis, E., Chatzistelios, G., & Kirytopoulos, K. (2022). Project management for corporate events: a set of tools to manage risk and increase quality outcomes. Sustainability, 14.

[4] Fadun, O. S. (2018). Risk factors and risk management strategies as correlates of profitability and survival of selected small and medium enterprises (smes) in nigeria. Journal of Management Studies(2).

[5] Erasmo, Giambona, John, R., Graham, & Campbell, et al. (2018). The theory and practice of corporate risk management: evidence from the field. Financial Management.

[6] Hillemann, J., Verbeke, A., & Oh, W. Y. (2019). Regional integration, multinational enterprise strategy and the impact of country level risk: the case of the emu. British Journal of Management.

[7] Christopher, J., & Sarens, G. (2018). Diffusion of corporate risk-management characteristics: perspectives of chief audit executives through a survey approach : diffusion of risk-management characteristics. Australian Journal of Public Administration, 77, 427-441.

[8] Henderson, C., Jia, S., & Mattioli, C. (2020). Supervisory bank risk early warning modeling: an examiner's rst line of defense. Journal of Credit Risk, 16(3), 75-100.

[9] Kim, Sungjae, F., Chance, & Don, M. (2018). An empirical analysis of corporate currency risk management policies and practices. Pacific-Basin finance journal.

[10] Lu, H., Liu, X., & Falkenberg, L. (2022). Investigating the impact of corporate social responsibility (csr) on risk management practices:. Business & Society, 61(2), 496-534.

[11] Tang, X., Li, S., Tan, M., & Shi, W. (2020). Incorporating textual and management factors into financial distress prediction: a comparative study of machine learning methods. Journal of Forecasting(4).

[12] Lv, D., Wu, C., & Dong, L. (2020). A k-means++-improved radial basis function neural network model for corporate financial crisis early warning: an empirical model validation for chinese listed companies. Journal of Risk Model Validation, 14(3).

[13] Huang, X., & Guo, F. (2020). A kernel fuzzy twin svm model for early warning systems of extreme financial risks. International Journal of Finance & Economics.

[14] Huang, J., Wang, H., & Kochenberger, G. (2017). Distressed chinese firm prediction with discretized data. Management Decision, 55(5), 786-807.

[15] Lechner, P., & Gatzert, N. (2018). Determinants and value of enterprise risk management: empirical evidence from germany. The European journal of finance, 24(10-12), 867-887.

[16] Savic, P. E., Velickovic, M., Voza, D., Ivkovi, I., & Zuzana Virglerová. (2020). The impact of enterprise risk management on the performance of companies in transition countries: serbia case study. Journal of Operational Risk, 14(4), 105-132.

[17] Shaheen, R., Aa, M., Rjoub, H., & Abualrub, A. (2020). Investigation of the pillars of sustainability risk management as an extension of enterprise risk management on palestinian insurance firms' profitability. Sustainability, 12.

[18] Sax, J., & Andersen, T. J. (2019). Making risk management strategic: integrating enterprise risk management with strategic planning. European Management Review, 16(3).

[19] Chen, J., Jiao, L., & Harrison, G. (2019). Organisational culture and enterprise risk management: the australian not-for-profit context. Australian Journal of Public Administration(7).

About the Author

Wei Su (1985.07-), female, Han nationality, native to Xiaoyi Shanxi, master, lecturer, research direction: financial management, cost accounting, management accounting.