WILEY | Hindawi

Research Article

Financial Risk Early-Warning of Neusoft Group Based on Support Vector Machine

Yuxuan Dai [1] **and Chenhui Yu** [2]

[1]*School of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing 210023, China*
[2]*School of Management, Nanjing University of Posts and Telecommunications, Nanjing 210003, China*

Correspondence should be addressed to Yuxuan Dai; duinomaker@gmail.com

As an emerging development industry, the information technology industry faces the most internal and external problems without effective risk prevention measures, which requires an effective financial risk early-warning system to be established to control risks. Nowadays, the advantages of support vector machine (SVM) have gradually appeared. The research on financial risk early-warning using SVM mostly stays in dichotomy. However, the financial risk of an enterprise will not only exist in absolute risk and no risk. There are many other levels of risk categories. Therefore, this paper proposes a new financial warning idea, which extends the support vector machine dichotomous to multidivision. This article focuses on the data modeling based on the financial data of listed companies in China's A-share information technology industry and applied to the case company Neusoft Group. First, the principal component analysis method is applied to assign the weights of financial indicators, and then the efficacy coefficient method is applied to comprehensively evaluate risk classification. Finally, the classified data were input into SVM for training and testing, and the model was applied to the financial risk early warning of Neusoft Group. The research results show that the model can better predict the financial risk of Neusoft Group.

1. Introduction

Against the background of increasing economic globalization, listed companies in the information technology industry have experienced rapid development. However, under the influence of the financial crisis, trade barriers established by hegemonism, and weaker internal control, the operational and financial risks that the technology industry faces have become increasingly prominent. Therefore, it is of great significance to establish a financial risk early warning system to prevent the industry from financial crises.

Beginning in the 1930s, foreign scholars began to explore related financial early warning. The research object gradually transitioned from single variable to multivariate, and then to the application of artificial intelligence models, which gradually deepened the research on an early financial warning.

Fitzpatrick [1] was the earliest to adopt the univariate analysis technique to provide early financial warning to companies. He believed that the variable with financial early warning significance was the trend of financial ratios. Beaver [2] further investigated the univariate early warning model based on this study. The univariate model is relatively simple and easy to understand but obviously unable to reflect the overall financial status of the enterprise and has low prediction accuracy. It is likely that different financial indicators may produce different prediction results for the same company.

With the gradual emergence of the shortcomings of univariate models, scholars have realized that a single variable cannot fully reflect the overall characteristics of corporate finance, so they tend to use multivariate financial early warning models. Altman [3] first applied the multivariate discrimination method to corporate financial early warning and to build a Z-score model. The model shows strong forecasting ability in the application of corporate financial early warning. Based on this research, Altman et al.

[4] optimized and introduced a ZETA model with 7 explanatory variables, which further improved the accuracy of the early warning results. Although the accuracy of the early warning has been improved to a certain extent, there are also various disadvantages: for example, the multiple normal assumption is too rational, and the distribution of the two types of enterprises is not in line with the reality, and so on.

Subsequently, a multivariate logistic regression model was introduced and effectively overcame the shortcoming. Ohlson [5] applied the multiple regression model to early financial warning. Almansour and Arabia [6] used regression analysis to build predictive models for 22 bankrupt and nonbankrupt Jordan listed companies between 2000 and 2003. Although multiple logistic regression has made great breakthroughs, the calculation is complicated, and the choice of samples is limited by the estimation of the overall probability, which is not easy to understand and operate.

With the continuous development of information technology, artificial intelligence is gradually applied to financial warnings. For example, artificial neural networks and genetic algorithms have been used in the financial risk warning area. Artificial neural network is a parallel and decentralized processing model with dynamic characteristics. It can overcome the limitations of statistical methods, and it is an important tool for solving segmentation problems. Odom and Sharda [7] applied neural network models to financial early warning for the first time, and the sample accuracy rate obtained was more than 80%, which further improved the accuracy of financial early warning. López Iturriaga and Sanz [8], based on data from the Federal Deposit Insurance Corporation between 2002 and 2012, established a neural network model to study bank failures in the United States, predicting the probability of a crisis in the three years before the bankruptcy.

In addition to artificial neural networks, Billo et al. [9] used different entropy measures to analyze the time evolution of systemic risk in Europe, constructed a new bank crisis warning indicator, and empirical analysis showed that the entropy measure predicts the risk of the bank crisis. Pham Vo Ninh et al. [10] used accounting factors in the scoring model, market factors in the distance to default, and macroeconomic indicators to model the financial distress of Vietnamese listed companies. Empirical research finds that each factor has an impact on financial distress when considered separately. But in the comprehensive model, the influence of accounting factors is more significant than that of market factors.

The results of this prediction method show great accuracy. However, nonparametric methods also have some disadvantages. They are prone to overfitting and overlearning, and it is difficult to determine the prediction results in the prediction process.

Support vector machine (SVM) was first introduced by Vapnik, and it has been developed for a long time. Compared with other machine learning models, such as KNN classifier, Bayesian networks, artificial neural networks, decision trees, and so on, SVM has a superior generalization performance and can guarantee convergence to optimal solutions. Therefore, SVM has attracted the attention of data mining and pattern recognition [11]. Some research has also shown that SVM is superior to other supervised learning methods on classification problems [11–13]. Because of its superior performance, SVM has also been used in various fields, such as time series prediction [14, 15], business [16], water resource management [17], image processing [18], and geology [19].

SVM is a machine learning method that can solve this problem effectively. The SVM overcomes the problems of overlearning, local convergence, and dimensional disasters in traditional machine learning. It has a good learning ability and generalization ability in solving small samples, nonlinear and multidimensional pattern recognition, and it has research possibilities in the multiclassification research area of financial risk early warning. Its core idea is to establish an optimal classification surface as a decision surface. Endri et al. [20] developed four early warning system models that can predict the occurrence of delisting of Islamic stocks (ISSI) using SVM. The results show that the financial variables had predictive power for the occurrence of delisting of Islamic stocks in the ISSI index and the SVM Model 4 is the best model. Besides, some academic communities have begun to extend the research to the algorithm of SVM. Qiao and Du [21] propose a novel hybrid PSO-SVM model based enterprise financial risk early warning algorithm by converting the proposed problem into a classification problem. They found it performs better than BP neural network in this area and SVM without parameter optimization. It also began some exploration of the SVM multiple division method in other areas. Wang et al. [22] constructed an urban real estate early warning model based on a multiclass SVM. Therefore, following the principles of effectiveness, the SVM is introduced into the early warning field for the purpose of considering the financial risks the enterprise faces. The model is found to have a good early warning judgment performance and demonstrates generalization ability.

However, the existing SVM studies on financial risk stay in two categories. Unlike foreign countries where bankruptcy is used as the classification criterion for companies facing financial crises, most Chinese scholars use ST and non-ST as the classification criterion. It is difficult to make forward-looking predictions of corporate financial risks, for there is no further subdivision of the risk. It is necessary to further refine the risks, as this kind of classification is too rough. Therefore, this paper proposes a new risk early warning based on the SVM to further divide the risk classification. Before applying to SVM, the risk classification should be made first. There are a few methods that can be tried to divide the risk levels, such as the principal component method and grey evaluation. Compared to the other methods, the efficacy coefficient method is easy to understand, easy to operate, and can show considerable results. It was first applied to the performance evaluation of enterprises and gradually extended to the field of financial warning. Li et al. [23] constructed an indicator system of early warning of the real estate bubble and used the analytic hierarchy process and efficacy coefficient method to conduct an empirical analysis of the degree of real estate bubbles. It was proved to have great prediction accuracy. Therefore, it is reasonable to consider this method a risk classification tool.

In the calculation process, while using the efficacy coefficient method, the index weights should be determined.

Usually, the weighting of financial early warning indicators in the market is generally divided into the following types: one is to use the weights provided in the Standard Value of Enterprise Performance Evaluation, which is more convenient, but inflexible in assigning; the other is to use expert ratings to analyze the weights and compare them layer by layer. This method is characterized by the difficulty of obtaining expert opinions and is subjective, as it is based on the experience drawn from previous people. The experts' experience varies and their accuracy is difficult to judge. The third one is to use mathematical models to give weight. It is believed that the third method is relatively reliable and flexible, while the principal component analysis method is simple to operate and applicable. Therefore, this method is chosen to assign financial indicator weights.

In summary, this paper proposes a new idea based on SVM to do early warning of industry companies' technologies to make multisubdivision a reality. The early warning indicators will be selected first, and the PCA will be used to confirm the index weights of these indicators. Then the efficacy coefficient method is applied to determine the classification of each company per year. At last, the data will be divided into a training set and a test set to be trained and tested in SVM to confirm the accuracy of the prediction.

The remainder of this paper proposes to construct a model which integrates the principal component method, efficacy coefficient method, and the SVM to build. Section 2 introduces the basic principles of the three methods. Section 3 discusses how these methods will be made to construct the model. The financial risk warning system will be constructed step by step and applied to the case enterprise in Section 4. Section 5 concludes the research and makes some visions for the future. The research on financial risk early warning of support vector machines is deepened to a certain extent, especially the research on multicategory early-warning of risks of information technology listed companies, which provides new ideas for similar types of research.

2. The Existing Theories

The system can be constructed by the efficacy coefficient method, the PCA, and the SVM. In this section, we will introduce the basic principles.

2.1. The Support Vector Machine Theory. In the field of machine learning, SVM is a classic and effective classification model. In a binary classification task, a training sample will be as follows:

$$D = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m), \quad y_i \in \{-1, +1\}\}, \quad (1)$$

where the feature vector $x_i \in \mathbb{R}^d$ represents a sample item in the d-dimensional space and y_i represents the true label of that sample item; -1 means the sample is negative and $+1$ means the sample is positive. The goal of all classification models is to obtain a classification rule on the training set that can reasonably distinguish between positive and negative samples and then apply this rule to a set of test samples with unknown true labels in order to obtain higher classification accuracy.

In fact, in classification tasks in most scenarios, the basic SVM can rarely solve the difficulties effectively: on the one hand, the division rules that can reasonably distinguish between positive and negative samples are often nonlinear with the feature vector, so I want to follow the SVM partitioning hyperplane assumption; it is necessary to find a nonlinear to linear mapping relationship; on the other hand, it is often necessary to allow the SVM to partition errors on a small number of training samples to further reduce the risk of overfitting. In order to solve the above-mentioned problems, a basic model of SVM with "kernel function" and "soft interval" is introduced.

The kernel function is introduced to nonlinear map eigenvectors into a certain high-dimensional space. In this high-dimensional space, the training sample set can be divided by a hyperplane. Note that such a nonlinear mapping function as $\phi(\cdot)$ is generally consistent with the following equation:

$$\phi(x_i)^T \phi(x_j) = \kappa(x_i, x_j), \quad (2)$$

where $\phi(x_i)$ and $\phi(x_j)$ represent the mapped eigenvectors, $\kappa(\cdot, \cdot)$ represent a kernel function, so when calculating the inner product of eigenvectors in high-dimensional (infinite-dimensional) space, the difficulty of direct calculation can be avoided and converted into the value of the $\kappa(\cdot, \cdot)$ kernel function in the original space. Commonly used kernel functions are as follows:

$$\text{Polynomial kernel Polynomial degree } \kappa(x_i, x_j) = (x_i^T x_j)^d, \quad d \geq 1, \quad (3)$$

$$\text{Radial basis kernel band width } \kappa(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right), \quad \sigma > 0, \quad (4)$$

$$\text{Sigmoid core } \kappa(x_i, x_j) = \tanh(\beta x_i^T x_j + \theta), \quad \beta > 0, \theta < 0. \quad (5)$$

After the parameters are determined, each kernel function uniquely determines a kernel mapping function $\phi(\cdot)$ for nonlinearly mapping the feature vector to the corresponding high-dimensional space.

Soft intervals are introduced to increase the fault tolerance mechanism of SVM on some samples. With the soft interval setting, all samples on the training sample set no longer strictly meet the inequality limit $y_i(w^T\phi(x_i) + b) \geq 1$; that is, it is no longer strictly required that all positive and negative samples fall within the corresponding region specified by the division of the hyperplane and support vector, but instead gives A certain degree of slack. Let the relaxation of the samples (x_i, y_i) be ξ_i, and the new inequality constraint is converted into course $y_i(w^T\phi(x_i) + b) \geq 1 - \xi_i$ where $i = 1, 2, \ldots, m$. Such relaxation is not infinite, so in soft-spaced SVM, the sum of the relaxations of all samples is added to the minimized target term. Regularize the degree of relaxation and control the overall degree of relaxation.

In summary, the optimization goal corresponding to a basic SVM model is as follows:

$$\min_{w,b,\xi_i} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m}\xi_i$$

$$\text{s.t.} \, y_i\left(w^T\phi(x_i) + b\right) \geq 1 - \xi_i, \qquad (6)$$

$$\xi_i > 0, \quad i = 1, 2, \ldots, m.$$

Among which is the regularization coefficient of the overall relaxation degree. The larger the overall relaxation degree is, the smaller a soft interval tends to be a (hard) interval. When the maximum interval division hyperplane is found, the new sample can be classified by the following discriminant function:

$$f(x) = \text{sign}\left(w^T\phi(x) + b\right), \qquad (7)$$

where $\text{sign}(\cdot)$ is the function for the sign of a real number and $\phi(\cdot)$ is a kernel mapping function determined by a kernel function.

For the optimization objective of the SVM basic model, the Lagrangian multiplier method is used first, and the corresponding partial derivative is zero, which can be transformed into the following:

$$\max_{\alpha} \sum_{i=1}^{m}\alpha_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j \kappa(x_i, x_j)$$

$$\qquad (8)$$

$$\text{s.t.} \sum_{i=1}^{m}\alpha_i y_i = 0, \quad 0 \leq a_i \leq C,$$

and when the partial derivative is equal to zero, $w = \sum_{i=1}^{m}\alpha_i y_i \phi(x_i)$.

Note that inequality constraints can be transformed into KKT (Karush–Kuhn–Tucker) conditional constraints:

$$\begin{cases} \alpha_i \geq 0, \mu_i \geq 0, \\ y_i\left(w^T\phi(x_i) + b\right) - 1 + \xi_i \geq 0, \\ \alpha_i\left(y_i\left(w^T\phi(x_i) + b\right) - 1 + \xi_i\right) = 0, \\ \xi_i \geq 0, \quad \mu_i\xi_i = 0, \end{cases} \qquad (9)$$

where α_i, μ_i is the Lagrangian multiplier, which can be approximated by using the Sequential Minimal Optimization (SMO) algorithm on the KKT condition, and the set of support vectors $S = \{(x_s, y_s)\}$ when the inequality holds.

In practice, the calculation formula is as follows:

$$b = \frac{1}{|S|}\sum_{s \in S}\left(\frac{1}{y_s} - \sum_{i \in S}\alpha_i y_i \kappa(x_i, x_s)\right). \qquad (10)$$

Another parameter of the SVM w is generally not directly solved but is brought into the discriminant function:

$$\begin{aligned} f(x) &= \text{sign}\left(w^T\phi(x) + b\right) \\ &= \text{sign}\left(\sum_{i=1}^{m}\alpha_i y_i \phi(x_i)^T\phi(x) + b\right) \\ &= \text{sign}\left(\sum_{i=1}^{m}\alpha_i y_i \kappa(x_i, x) + b\right). \end{aligned} \qquad (11)$$

As mentioned above, SVM is designed for a two-class classification task, which can usually be extended to multiclass tasks by a one-versus-rest (OVR) or one-versus-one (OVO) strategy. Recording the number of multiclass categories of the sample as n_{class}, the two strategies are described as follows:

(i) OVR strategy: On the training sample set, train an SVM model for each type of sample. Positive samples are all samples of the current class and negative samples are all remaining samples. Therefore, a total of n_{class} SVMs need to be trained. Each SVM gives the sample the confidence that the sample belongs to the current class, and the final classification result of the sample is the category corresponding to the SVM with the highest confidence.

(ii) OVO strategy: On the training sample set, train an SVM for two or two classes. Positive samples are all samples of one class and negative samples are all samples of the other class, so you need to train $n_{\text{class}}(n_{\text{class}} - 1)/2$ SVMs; on the new test sample set, each SVM gives a sample category discrimination as a vote, and the final category of the sample is the category with the most votes.

The classification accuracy of the two strategies depends on the specific data distribution. In general, the performance is similar.

2.2. The Efficacy Coefficient Method Theory. The efficacy coefficient method is a multiobjective decision-making method. It was first proposed by Harrington in 1965. It first gives a set of indicators to be evaluated with upper and lower limits of the range of changes and then quantifies the efficacy coefficient of a single indicator based on a series of formulas. Finally, the efficacy coefficient of each group of indicators is calculated by weighting and summing so as to comprehensively evaluate the target company. The principle of this

method is easy to understand, easy to operate, and practical. The efficacy coefficient method generally has a series of processes to calculate a comprehensive score: First, the optimal and worst levels of the indicators are used to specify the satisfaction and permissible values of each indicator. Second, calculate the score of each single indicator according to the following formula:

$$\text{index efficacy coefficient} = \frac{(\text{actual value of the index} - \text{not allowed value of the index})}{(\text{satisfactory value of the index} - \text{not allowed value of the index})},$$

$$\text{index score} = \text{index efficacy coefficient} \times 40 + 60.$$

(12)

Finally, calculate the comprehensive index score

$$\text{comprehensive indicator score} = \sum \text{single indicator score} \times \text{the weight of the indicator}.$$

(13)

This article applies this method to the classification of corporate financial risk categories in the industry.

2.3. The Principle Component Theory. PCA is a commonly used unsupervised dimensionality reduction algorithm. It can extract the main components from the data and use them for analysis. It is widely used in face recognition, handwriting recognition and other fields. Here are the core ideas of PCA:

In the absence of (or without considering) the true label, let the training sample set be $D = \{x_1, x_2, \ldots, x_m\}$. The feature vector $x_i \in \mathbb{R}^d$ representing the sample is a vector in a d dimensional space. Usually the data dimensionality reduction algorithm hopes to find a lower dimensional d' $(d' \ll d)$ space $\mathbb{R}^{d'}$ and obtain a new representation of the sample set in this low-dimensional space as a "low-dimensional embedding." Generally, such an embedding allows some of the information to be lost while retaining the main information.

Out of simple assumptions, the PCA algorithm considers that the transformation from the original space to the embedded space is a (linear) coordinate transformation. We assume that the sample $x \in \mathbb{R}^d$ (which has been centralized), where the standard $\mathbb{R}^d$ orthogonal basis of the low-dimensional space referenced is taken as $W = (w_1, w_2, \ldots, w_{d'})$ and space W is taken as $\mathbb{R}^d$. The $x \in \mathbb{R}^d$ corresponding coordinate transformation result is recorded as $z \in \mathbb{R}^{d'}$, $z = W^T x$. Of course, there are many possible cases for such a low-dimensional embedding space and the corresponding standard orthogonal basis, which also correspond to different "reconstruction errors". From a thinking point of view, the core idea of PCA is to set the minimization "reconstruction error" as the optimization goal. The calculation formula for N samples is as follows:

$$W^* = \arg\min_W \sum_{i=1}^{N} \left\| WW^T x_i - x_i \right\|^2,$$

(14)

where x can be (partially) reconstructed according to the embedding result $z = W^T x$, which is denoted as a reconstruction term $WW^T x_i$, and the reconstruction error between the reconstruction term and the original sample is defined as the Euclidean distance between the vectors. In fact, minimizing the reconstruction error is equivalent to retaining as much information as possible during the low-dimensional embedding process. The solution of the optimization target $W^* = (w_1, w_2, \ldots, w_{d'})$ is called the "projection matrix," and the corresponding vectors are called "principal components" (which can be multiplied by any nonzero coefficient), which is the standard orthogonal basis of the low-dimensional embedding space.

3. Basic Framework of Proposed Early Warning Model

After a brief introduction of the three methods, the proposed method in this research should be introduced as follows. First, select the indicators that can reflect the condition of the enterprise's financial risk. In this place, the selection will rely on the existing research result, for the previous studies are already representative. Then PCA will be used to ensure the weights of each indicator, which will be used in the calculation of the marks of financial risk. Then the efficacy coefficient method will be used to calculate the marks of each year of each company to decide which extent of risk it faced or faces. Finally, according to the different risk classifications, the results will be divided into training set and test set to train and test the SVM. When the model tends to be stable and has good accuracy, it will be applied in the case company to predict the company's financial risk.

3.1. Algorithm Steps. In the following financial early warning framework, the classification of financial risk is computed.

Step 1. Construction of financial indicator system

Step 2. Determining the weight of the screened financial indicators using principal component analysis

Step 3. Applying the determined financial indicator weights to the measurement of financial risk classification by the power factor method

Step 4. Using financial indicators as *X* and risk classification as *Y*, establish a financial risk early warning model based on support vector machines

Step 5. Early warning result output

This section is divided into subheadings. It should provide a concise and precise description of the experimental results, their interpretation, and the experimental conclusions that can be drawn.

4. Case Study

To acquire a more accurate financial warning model of the case company, after knowing the process of the warning model, we selected companies in the technology industry as the model data. The model built based on the SVM will output the results of the case company and do further analysis.

4.1. Data Sources and Pretreatment of Model Data. This article selects 263 companies from 2008 to 2018 in the information technology industry listed in A-shares. First, it excludes the impact of earnings management and profit manipulation of the first public offering of listed companies. After the removal of companies no more than three years old, 182 companies remain. Taking into account the particularity of the information technology industry, in some years, many companies have deviated from the main business of the information technology industry after the listing. There are situations where the information technology industry once did not belong or once did not belong to the current information industry. Therefore, the data of industry codes I63, I64, and I65 in the data of each year were retained, and the rest were filtered out. A total of 829 sets of data of different companies in different years were left after the above-mentioned layers of screening. The 182 listed companies and stock codes involved are listed in Table 1.

4.2. Variable Selection. Based on the principles of systemicity, irrelevance, sensitivity, and practical operability, and on the characteristics of Neusoft Group's information technology industry, four aspects of a warning system are selected: profitability, operating capacity, debt repayment ability, and development ability specifically including 14 indicators, which establish the Neusoft Group risk early warning system. The financial indicator system is shown in Table 2.

4.3. The Principle Component Analysis. Before using the power factor method, it is necessary to confirm the index weight. It should go through the KMO and Bartlett's test and extract the main factors and normalize the weights of each index.

4.3.1. KMO and Bartlett's Test. First, the sample data is adaptively analyzed. Generally, KMO and Bartlett's test is used. It is usually considered that a KMO value greater than 0.6 is suitable for principal component analysis. The results of the sample data measurement in this article are shown in Table 3.

4.3.2. The Main Factor Analysis. After passing the adaptive analysis, the principal components can be determined and extracted. The number of principal components can be determined according to the principal component contribution rate, and the cumulative contribution rate can be observed. Those with eigenvalues greater than 1 can be extracted as the main component. As can be seen in Table 4, six main components are retained according to the test results of the sample data, and the cumulative contribution rate of 78.579% is close to 80% as an acceptable category.

Further, the expression of the principal components are constructed based on the extracted principal components.

From the component index matrix in Table 5, six principal component expressions can be written. The data in the table correspond to the coefficient of each indicator. The principal component comprehensive model is then calculated according to the above six principal component expressions, and the coefficients of each indicator are multiplied by the contribution rate corresponding to each indicator in each principal component and added up one by one.

The final expression is as follows:

$$F = 0.0601X_1 + 0.0081X_2 + 0.0689X_3 + 0.0642X_4 + 0.059X_5 + 0.0575X_6 + 0.0776X_7$$
$$+ 0.0704X_8 + 0.0546X_9 + 0.0706X_{10} + 0.0379X_{11} + 0.0671X_{12} + 0.0620X_{13} + 0.0526X_{14}. \tag{15}$$

According to the coefficient of each index of the principal component comprehensive model, they were normalized and the weights of each index are obtained, as shown in Table 6.

TABLE 1: 182 listed company stock codes.

000503	002401	300166	300366	002238	300043	300310	600654
000555	002405	300167	300369	002247	300044	300311	600701
000606	002410	300168	300377	002253	300047	300312	600728
000638	002421	300170	300378	002261	300050	300315	600756
000662	002425	300183	300379	002268	300051	300324	600767
000676	002439	300188	300380	002279	300052	300330	600797
000835	002447	300209	300383	002280	300059	300333	600804
000839	002464	300212	300386	002298	300074	300339	600831
000889	002467	300226	300392	002315	300075	300343	600845
000917	002474	300229	300399	002316	300085	300348	600850
000948	002517	300231	600037	002331	300096	300349	600892
000971	002530	300242	600050	002354	300098	300350	600986
000997	002544	300245	600242	002359	300104	300352	601360
002063	002555	300248	600289	002368	300113	300359	601519
002065	002558	300249	600358	002373	300150	300365	601929
002072	002609	300250	600406	002153	300017	300290	600588
002093	002619	300253	600410	002174	300020	300292	600602
002095	002624	300264	600446	002175	300025	300295	600633
002113	002642	300271	600476	002195	300031	300297	600634
002123	002649	300275	600536	002230	300033	300299	600637
002131	002657	300277	600556	002232	300036	300300	600640
002137	300002	300287	600570	002235	300038	300302	600652
002148	300010	300288	600571	603636	300235		

TABLE 2: Variable selection.

Evaluation content	Financial indicator	Formula
Profitability	Return on equity	Net profit/net assets
	Operating margin	Operating profit/net operating income
	Cost expense margin	Total operating profit/cost expense
Operating capacity	Total asset turnover	Sales revenue/average total assets
	Accounts receivable turnover rate	Net credit sales income/average balance of accounts receivable
	Inventory turnover rate	Operating cost/average inventory balance
	Return on assets	EBIT/Average total assets
	Turnover rate of current assets	Net income from main operations/average total current assets
Solvency ability	Quick ratio	Quick assets/current liabilities
	Net cash flow from operating activities/current liabilities	Net cash flow from operating activities/current liabilities
	Asset–liability ratio	Total liabilities/total assets
Growth capacity	Operating income growth rate	Growth in operating income/total operating income in the previous year
	Total asset growth rate	Total asset growth this year/total assets at the beginning of the year
	Capital accumulation rate	Increase in owner's equity for the year/capital accumulation rate of owner's equity in early

TABLE 3: KMO and Bartlett's test.

Kaiser–Meyer–Olkin measure of sampling adequacy		0.631
Bartlett's test of sphericity	Approx. Chi-square	5516.206
	Df.	91
	Sig.	0.000

4.4. Efficacy Coefficient Method

4.4.1. Determination of Early-Warning Level.
Risks can be divided into categories shown in Table 7 based on the improved efficacy coefficient method.

In the huge alert level, the company's financial situation has deteriorated sharply and is in danger of bankruptcy. Production and operation of the company are interrupted at any time, and almost all indicators have shown a negative trend.

In the major alert level, companies have some difficulties in the operation, most of the financial indicators are at the lowest value, and they are likely to face financial risks.

In the moderate alert level, the company has poor business management, some indicators have deteriorated, the overall financial situation has declined and financial risks may be faced.

TABLE 4: Main components of the characteristic value and variance contribution.

The initial factor	Initial contribution factor			Selection factor		
	Eigenvalue	Contrib. (%)	Total contrib. (%)	Eigenvalue	Contrib. (%)	Total contrib. (%)
1	3.280	23.430	23.430	3.280	23.430	23.430
2	2.378	16.989	40.419	2.378	16.989	40.419
3	1.714	12.245	52.664	1.714	12.245	52.664
4	1.570	11.213	63.877	1.570	11.213	63.877
5	1.057	7.548	71.425	1.057	7.548	71.425
6	1.002	7.154	78.579	1.002	7.154	78.579
7	0.881	6.292	84.871			
…	…	…	…			
14	0.078	0.560	100.000			

TABLE 5: Principal component extraction.

Indicators	Components					
	1	2	3	4	5	6
Quick ratio	−0.096	0.053	−0.002	0.533	0.042	−0.007
Net cash flow from operating activities/current liabilities	0.123	0.029	−0.046	0.200	−0.435	−0.074
Asset-liability ratio	0.049	−0.010	0.006	−0.508	−0.126	−0.038
Capital accumulation rate	−0.044	0.002	0.512	0.003	0.016	0.009
Total asset growth rate	−0.056	0.001	0.522	−0.023	0.014	0.001
Operating income growth rate	−0.040	0.015	0.021	0.073	0.671	−0.054
Accounts receivable turnover rate	−0.021	0.384	0.009	0.098	0.088	−0.032
Inventory turnover rate	0.001	−0.004	0.005	0.005	−0.018	0.991
Turnover rate of current assets	0.014	0.361	−0.014	−0.015	−0.061	0.016
Total asset turnover	0.003	0.403	0.008	0.017	0.000	0.005
Return on assets	0.331	0.001	−0.018	−0.067	−0.211	−0.013
Return on net assets	0.166	0.013	−0.071	0.033	0.408	0.051
Operating margin	0.336	−0.003	−0.062	−0.058	0.116	0.013
Cost expense margin	0.333	−0.012	−0.002	−0.046	−0.069	−0.015

TABLE 6: Index weight.

Indicators	Variable	Weight
Quick ratio	X_1	0.0742
Net cash flow from operating activities/current liabilities	X_2	0.0100
Asset–liability ratio	X_3	0.0850
Capital accumulation rate	X_4	0.0793
Total asset growth rate	X_5	0.0727
Operating income growth rate	X_6	0.0710
Accounts receivable turnover rate	X_7	0.0957
Inventory turnover rate	X_8	0.0869
Turnover rate of current assets	X_9	0.0673
Total asset turnover	X_{10}	0.0870
Return on assets	X_{11}	0.0468
Return on net assets	X_{12}	0.0828
Operating margin	X_{13}	0.0764
Cost expense margin	X_{14}	0.0648

In the mild alert level, enterprise operations are almost normal, while individual indicators show abnormalities and are less likely to face financial risks.

In the no alert level, the company operates well, almost all indicators perform well, and the financial situation is stable, with only a small chance of facing financial risks.

4.4.2. Determination of Risk Categories.
According to the coefficients of the indicators calculated above, corresponding to the standard values of the information technology service industry in the "Standard Values of Enterprise Performance Evaluation" issued by the country each year. The improved efficiency coefficient method above was used to calculate the comprehensive scoring of 829 groups from 2008 to 2018 to determine risk categories.

According to the results shown in Table 8 which were calculated by applying the efficiency coefficient method, 829 valid data are classified into five risk levels, so this problem is a five-class nonlinear problem, and the kernel function of the SVM can address nonlinearly effectively. However, because the calculation results are similar, the classification gap is not

TABLE 7: Early-warning level evaluation interval risk situation description.

Warning level	Evaluation interval	Risk profile
Huge alert	$0 \leq X < 0.3$	No risk
Major alert	$0.3 \leq X < 0.5$	Focus on risk
Moderate alert	$0.5 \leq X < 0.7$	Less risky
Mild alert	$0.7 \leq X < 0.85$	Greater risk
No alert	$0.85 \leq X < 1$	Significant risk

TABLE 8: Financial risk classification of 182 listed companies from 2008 to 2018.

Year	Huge alert	Major alert	Moderate alert	Mild alert	No alert
2008	1	7	10	2	–
2009	–	6	12	2	–
2010	–	2	13	6	2
2011	–	1	19	6	–
2012	–	5	17	15	–
2013	–	5	28	13	2
2014	–	5	42	26	2
2015	1	23	60	16	–
2016	2	24	101	13	–
2017	7	29	108	14	–
2018	11	51	93	26	1

large, and the sense of boundary is not strong. Since the classification level has been changed from the previous two to five, there is a problem that the classification boundary is not obvious, for there is a large amount of boundary data. In order to solve this problem, this article proposes to further filter 829 effective data and to remove the boundary data such as 0.3, 0.5, 0.7, and 0.85, which cannot accurately locate their risk category. It can remove some data that does not hinder the overall measurement result and it only affects the sample size. Removing such data will help make the classification clearer, and to a certain extent, it is conducive to the training and testing accuracy of SVM. For example, a comprehensive score with a calculated result of 0.5 is between the major alert level and the moderate alert level. It is difficult to define whether it belongs to the second risk level or the third risk level. After such screening, 713 valid data are left at last. The classification result is shown in Figure 1.

Among them, the sample data of the moderate alert is sufficient, and the data of the major alert and mild alert is acceptable, which is consistent with the situation that most enterprises are facing certain financial risks. The data for huge alerts and no alerts is small but at an acceptable level. The sample data will be divided into a training set and a test set for training and testing, respectively, by random sampling in a certain proportion. At last, the classifier will give 1 if there is a huge alert, 2 if there is a major alert, 3 if there is a moderate alert, 4 if there is a mild alert, and 5 if there is no alert. In this way, classification training is performed on the data in the training set, and classification testing is performed on the data in the test set. After repeated training and testing, the parameters are adjusted. The model tends to reach a balance point, and then the prediction effectiveness

of the model is evaluated. The model with great prediction accuracy will be applied to the financial risk early warning of Neusoft Group.

4.5. Establishment of Financial Early-Warning Model Based on Support Vector Machine.
To use SVM for financial early warning, we must first determine its kernel function and parameters. The adjustment of the kernel function and parameters will directly affect the accuracy of the model, with reference to the previous comparative studies of SVM financial early warning kernel functions. In general, the RBF kernel function has high prediction accuracy, and it can map the sample space to a higher dimensional space nonlinearly. The Sigmoid kernel function and polynomial kernel function, on the other hand, are slightly inferior. After conducting three experiments on three kinds of kernel functions, the author found that the results were basically consistent with the previous experiments, so the RBF kernel function was chosen for financial risk early warning.

In this paper, five risk dimensions from 713 sets of data were filtered, and each dimension was randomly selected according to the ratio of 8: 2, 7: 3, 6: 4. As a result, the network parameters were continuously modified, and finally the most effective financial risk early warning model is obtained. Using RBF in SVM requires setting two parameters C and gamma. C is the penalty coefficient, that is, the tolerance for error. The higher the C value, the more it is tolerant to error and prone to overfitting; the smaller the C value, the more prone to underfitting. Gamma is a parameter that comes with RBF as the kernel function, which implicitly determines the distribution of the data after it is mapped to the new feature space. The larger the gamma value, the fewer the support vectors and the smaller the gamma value, the more the support vectors, and the number of support vectors affects the number of predictions and training. According to the data set provided by the experimental design, we use the leave-one-out (LOO) validation method for cross-validation of the model and determine the optimal hyperparameters by grid search. After cross-validation, the model parameters of the support vector machine were finally determined to be C (16384×2, 16384×4, 16384×8) and gamma ($1/896/2$, $1/896$, $1/448$). Since there is a certain degree of randomness in the proposed model, the experiment is performed 100 times for each division in the same computational environment so that the experimental results of model performance can be statistically compared and tested.

4.5.1. Training and Testing of Early-Warning Models.
After the experimental steps above, the statistics of the 100 training and testing results in each case are shown in Table 9. The training and test accuracy percentage is rounded to two decimal places.

It can be known from the above three training situations that the resulting accuracy of the SVM training is stable around 85%, and the test results are stable around 73%, which is within the acceptable range. Among the five types of risks, predictions for major alert, moderate alert, and mild alert are more accurate in training and testing,
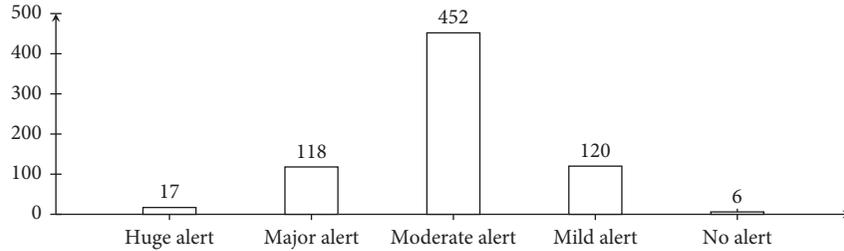
Figure 1: Risk classification.

Table 9: Financial risk warning training and testing.

Training set/Test set	Training accuracy (%)	Testing accuracy (%)
8/2	85.08	73.51
7/3	84.97	73.21
6/4	85.16	73.64

Table 10: Neusoft Group's financial risk forecast accuracy from 2008 to 2018.

Risk level	Huge alert	Major alert (%)	Moderate alert (%)	Mild alert (%)	No alert
Accuracy	−	100	85.71	100	−

with an average of 75% or more, while huge alert and no alert are not satisfactory in training and testing. Insufficient sample size makes training and testing data insufficient. There are only six sample data with no alert, and the test accuracy is almost zero. The accuracy of the huge alert samples was maintained around 50%. However, this situation is acceptable because the early warning of financial risks is mainly based on the prevention of risk categories of two, three, and four, and huge alert can also be valued in major alert areas. Therefore, it can be considered that this model is capable of being applied to the financial risk early warning of Neusoft Group.

4.6. Neusoft Group Financial Risk Early-Warning Results and Analysis. Due to the relative stability of the model, the Neusoft Group's financial indicators for a total of 11 years from 2008 to 2018 were calculated using the efficacy coefficient method to achieve a comprehensive score. The methodology is the same as above, and the classification result are divided into five risk levels.

As can be seen from the figure, Neusoft Group has had a good development momentum in the past 11 years. Except for 2014, when it reached the major alert, the rest of the years were in the moderate alert and mild alert levels.

At the same time, its financial indicator system has been input into the built SVM model to output the risk level, and its accuracy is calculated as shown in Table 10.

The financial risk of Neusoft Group from 2008 to 2018 involves three levels: major alert, moderate alert, and mild alert. Among them, the prediction accuracy of major alert and mild alert reached 100%, and the prediction accuracy of moderate alert reached 85.71%. There exists a misclassification from morderate alert to mild alert, and the total accuracy rate reaches 90.91%. It can be concluded that this model can be applied to the future financial early warning of

Neusoft Group and can play a role of financial risk early warning to a certain extent.

5. Conclusions

With the increasingly intense external and internal environment, enterprises in the information technology industry are facing serious challenges in controlling financial risks. This requires enterprises to identify and take measures to control financial risks in advance. Therefore, the necessity of rebuilding the financial early warning model of multiclass is put forward. This paper proposes an early financial warning based on the SVM model to classify the risk from two to five levels to meet the requirements. Based on sensitivity, systemicity, and operability, 14 financial indicators that can reflect the operating conditions of the information technology industry as comprehensively as possible are selected. Based on the financial data of 182 information technology companies listed in A-shares, the principal component analysis method is used to determine the index coefficients and the efficacy coefficient method is used to classify the risks. The classified data were then trained and tested in SVM for the Neusoft Group to do the test. After training and testing, the model proved to be effective for early warning and laid a solid foundation for early warning of risks.

The case study in this paper has certain implications for the financial risk early warning for other information technology companies. Neusoft has some special characteristics, but it is also common to the information technology industry. We hope that it can give some inspiration to other enterprises of the same type in terms of early warning of financial risks.

At the same time, a series of prospects for future research are also proposed. First, whether the classification of financial risks can be further distinguished in the future and whether a simpler and more suitable way can be found to

refine the risk categories. Second, it is possible to further combine financial indicators with nonfinancial indicators so that the data is more comprehensive and richer. Third, the selection of the kernel function of the support vector machine needs to be further validated, and the parameters need to be further adjusted. Fourth, the research on the multi-classification of SVM can be extended to other industries and case companies.

Data Availability

The financial statement related data used to support the findings of this study were supplied by CSMAR under license and so cannot be made freely available. Requests for access to these data should be made to https://cn gtdata com/.

Conflicts of Interest

The authors declare no conflicts of interest.

Authors' Contributions

Conceptualization was done by Y. Dai and C. Yu; methodology was developed by C. Yu; software was developed by Y. Dai; validation was performed by Y. Dai and C. Yu; formal analysis was performed by Y. Dai; investigation was done by C. Yu; resources were provided by C. Yu; data curation was performed by Y. Dai; original draft was written by C. Yu; review and editing were done by Y. Dai; visualization was performed by C. Yu; supervision was done by Y. Dai; project administration was done by Y. Dai. All authors have read and agreed to the published version of the manuscript.

References

[1] P. J. Fitzpatrick, "A comparison of ratios of successful industrial enterprises with those of failed firms," *Certified Public Account*, vol. 23, pp. 57–65, 1932.

[2] W. H. Beaver, "Financial ratios as predictors of failure," *Journal of Accounting Research*, vol. 4, pp. 71–111, 1966.

[3] E. I. Altman, "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy," *The Journal of Finance*, vol. 23, no. 4, pp. 589–609, 1968.

[4] E. I. Altman, R. G. Haldeman, and P. Narayanan, "ZETATM analysis A new model to identify bankruptcy risk of corporations," *Journal of Banking & Finance*, vol. 1, no. 1, pp. 29–54, 1977.

[5] J. A. Ohlson, "Financial ratios and the probabilistic prediction of bankruptcy," *Journal of Accounting Research*, vol. 18, no. 1, pp. 109–131, 1980.

[6] B. Almansour and S. Arabia, "Empirical model for predicting financial failure," *American Journal of Economics, Finance and Management*, vol. 1, pp. 113–124, 2015.

[7] M. D. Odom and R. Sharda, "A neural network model for bankruptcy prediction," *1990 IJCNN International Joint Conference on Neural Networks*, vol. 6, pp. 163–168, 1990.

[8] F. J. López Iturriaga and I. P. Sanz, "Bankruptcy visualization and prediction using neural networks: a study of U.S. commercial banks," *Expert Systems with Applications*, vol. 42, no. 6, pp. 2857–2869, 2015.

[9] M. Billio, R. Casarin, M. Costola, and A. Pasqualini, "An entropy-based early warning indicator for systemic risk," *Journal of International Financial Markets, Institutions and Money*, vol. 45, pp. 42–59, 2016.

[10] B. Pham Vo Ninh, T. Do Than, H. Vo, and D. Financial, "Distress and bankruptcy prediction: an appropriate model for listed firms in Vietnam," *Economic Systems*, vol. 5, pp. 132–148, 2018.

[11] J. Cervantes, F. García-Lamont, L. Rodríguez-Mazahua, and A. Lopez, "A comprehensive survey on support vector machine classification: applications, challenges and trends," *Neurocomputing*, vol. 408, pp. 189–215, 2020.

[12] J. Cervantes, F. García Lamont, A. López-Chau, L. Rodríguez Mazahua, and J. Sergio Ruíz, "Data selection based on decision tree for SVM classification on large data sets," *Applied Soft Computing*, vol. 37, pp. 787–798, 2015.

[13] X. Liang, L. Zhu, and D. S. Huang, "Multi-task ranking SVM for image cosegmentation," *Neurocomputing*, vol. 247, pp. 126–136, 2017.

[14] J. P. Karmy and S. Maldonado, "Hierarchical time series forecasting via support vector regression in the European Travel Retail Industry," *Expert Systems with Applications*, vol. 137, pp. 59–73, 2019.

[15] S. Lee, S. Lee, and M. Moon, "Hybrid change point detection for time series via support vector regression and CUSUM method," *Applied Soft Computing*, vol. 89, Article ID 106101, 2020.

[16] J. Zhang, Y. F. Teng, and W. Chen, "Support vector regression with modified firefly algorithm for stock price forecasting," *Applied Intelligence*, vol. 49, no. 5, pp. 1658–1674, 2019.

[17] M. Vafakhah and K. S. Bozchaloei, "Regional analysis of flow duration curves through support vector regression," *Water Resources Management*, vol. 34, no. 1, pp. 283–294, 2020.

[18] X. Wei, J. Du, M. Liang, and L. Ye, "Boosting deep attribute learning via support vector regression for fast moving crowd counting," *Pattern Recognition Letters*, vol. 119, pp. 12–23, 2019.

[19] X. Ding, M. Hasanipanah, H. Nikafshan Rad, and W. Zhou, "Predicting the blast-induced vibration velocity using a bagged support vector regression optimized with firefly algorithm," *Engineering with Computers*, vol. 37, pp. 2273–2284, 2020.

[20] E. Endri, K. Kasmir, and A. D. Syarif, "Delisting sharia stock prediction model based on financial information: support Vector Machine," *Decision Science Letters*, vol. 9, pp. 207–214, 2020.

[21] G. Qiao and L. Du, "Enterprise financial risk early warning method based on hybrid PSO-SVM model," *Journal of Applied Science and Engineering*, vol. 22, pp. 171–178, 2019.

[22] X. J. Wang, G. T. Zeng, K. X. Zhang, H. B. Chu, and Z. S. Chen, "Urban real estate market early warning based on support vector machine: a case study of beijing," *International Journal of Computational Intelligence Systems*, vol. 13, no. 1, pp. 153–166, 2020.

[23] J. Li, X. Jia, and H. Yuan, "Research on early warning of real estate bubble in xiamen city based on AHP-ECM," *Proceedings of the 5th International Conference on Economics, Management, Law and Education (EMLE 2019)*, vol. 110, pp. 301–205, 2019.