Retraction

Retracted: Application of Machine Learning in Enterprise Risk Management

Security and Communication Networks

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] D. Jia and Z. Wu, "Application of Machine Learning in Enterprise Risk Management," *Security and Communication Networks*, vol. 2022, Article ID 4323150, 12 pages, 2022.

WILEY | Hindawi

Research Article

Application of Machine Learning in Enterprise Risk Management

Dian Jia[1] and Zhaoyang Wu [2]

[1]*School of Economics and Management, Qinghai Minzu University, Xining 810007, Qinghai, China*
[2]*School of Economics and Management, Qinghai Normal University, Xining 810016, Qinghai, China*

Correspondence should be addressed to Zhaoyang Wu; 2018130@qhnu.edu.cn

With the development of China's economic globalization, more and more enterprises have realized the importance of risk management. As a new technology, machine learning has injected new vitality into enterprise risk management. This paper analyzes the factors of enterprise risk from the perspective of risk management, explores and analyzes the application of machine learning in enterprise risk management, and points out the direction for scientific enterprise risk management. This paper first establishes a set of credit evaluation factors and then builds a model based on this evaluation factor set, which ensures the accuracy of the evaluation and also adds pertinence. With the increase of the number of samples, the accuracy of the prediction results also increases. Also, basically, with the increase of training samples and realistic samples, the accuracy rate continues to increase. In the data of the experimental results, it is shown in the process of increasing the number of samples from 400 to 600. The accuracy of the traditional method is 0.81 and 0.8316, respectively. The accuracy of the improved method is 0.842 and 0.905, respectively. This paper studies the process of enterprise risk assessment from the perspective of risk management and attempts to build an enterprise risk management model based on industry risk coefficients, which have certain practical significance.

1. Introduction

The rapid development of economy and information technology brings opportunities to enterprises but also makes them face more risks. Financial risk management is essential for most businesses to avoid losses and maximize profits. This task relies on machine learning and information-driven decision making, which makes traditional analysis methods unable to meet the growth of data volume. Due to the improvement of computing processing power and the reduction of data storage cost, the application of artificial intelligence is becoming more and more extensive. More and more machine learning methods have been used in recent years for various risk management tasks. In real life, a lot of unstructured data is analyzed and processed using cognitive techniques, including natural language processing, which uses advanced algorithms to analyze patterns in language systems. Also, this function has a significant effect in the field of risk management. Especially in the field of risk management, it is necessary to find relevant clues in the face of a large and complex information

such as contracts, documents, and legal materials. Using artificial intelligence technology to process these unstructured data will greatly improve the efficiency of problem processing.

This paper uses a combination of theoretical analysis and case studies to sort out the current situation of enterprise risk management. This provides theoretical support for the application of machine learning in enterprise risk management. The innovation of this paper is as follows. (1) The challenge of applying machine learning in the risk management industry is not how to use machine learning, but how to obtain data and how to obtain data in the correct format. Companies often have more data available than they think they have. How to obtain the required and available data with innovative ideas and structure the data is a big challenge. (2) Although the research on risk management in China started relatively late, more researchers and institutions have begun to devote themselves to the research and exploration of the theoretical aspects of risk management. Compared with countries and regions, China still has many aspects that need to be improved in theory and practice.

2. Related Work

With the development of society, more and more researchers have begun to invest in the study of enterprise risk management. Among them, Lin et al. proposed an enterprise risk management (ERM) model. The importance of incorporating pension risk into a company's ERM plan is illustrated by comparing the value of companies with managed and unmanaged pension risk with other risks in an ERM plan. ERM plans that take into account pension effects integrate risks in the operational and pension sectors, thereby achieving diversification benefits between these two sectors and within the company [1]. Abkowitz and Camp suggested that due to the breadth and depth of factors that may affect an organization's risk portfolio, it is imperative that the underlying risk assessment process to support ERM embodies a holistic and systematic approach [2]. Valanarasu and Christy introduced risk assessment and management (RAM) in ERP through advanced system engineering theory [3]. Lv and Rong proposed a security risk assessment model for cloud services as a solution using stochastic game networks. Based on graphical tools, the virtualization security risk scenarios of cloud services can be clearly described, and virtualization security risk factors can be accurately evaluated. The analysis results prove that the method has a strong ability to simulate complex and dynamic security problems in cloud services [4]. He et al. evaluated the occupational health risk of decorative paint production enterprises, discussed the applicability of occupational hazard risk index model in health risk assessment, and provided a basis for enterprise health management [5]. Buczak and Guven described a focused literature survey of machine learning (ML) and data mining (DM) methods for network analysis in support of intrusion detection. A brief tutorial description of each ML/DM method is provided. It identifies, reads, and summarizes papers representing each method based on the number of citations or the relevance of the emerging method. Since data are very important in ML/DM methods, some well-known network datasets for ML/DM are described. The complexity of ML/DM algorithms is addressed, the challenges of using ML/DM for cybersecurity are discussed, and some recommendations are provided on when to use a given method [6]. Finance analyzes enterprise risk management (ERM), risk governance systems, and own risk solvency assessment (ORSA), exploring how they have been enhanced with the addition of Solvency II. The authors surveyed chief risk officers (CROs) working in Spanish insurance companies. The findings show that Solvency II has indeed promoted ERM and improved the governance system of insurers in the European insurance industry and that ORSA's perceived value to companies is higher than cost. Clearly, the quality of ERM implementation is higher in companies that face more complex risks and have greater interdependencies, namely, larger companies, foreign insurers, and insurers with multiple lines of business. But they are not affected by legal form [7].

The above scholars have conducted in-depth research on risk management, but the theoretical research and practical research on risk management in China are still in the initial stage, and there are still various situations and problems in the specific implementation process. It can be said that there is still room for research and analysis in this direction. This article has some innovations in the topic selection. At present, scholars at home and abroad focus on the research on the concept and function of enterprise risk management in enterprises. This paper combines enterprise risk management with machine learning and innovatively applies science and technology to risk management to provide guarantee for enterprise development.

3. Application Methods of Machine Learning in Enterprise Risk Management

3.1. Enterprise Risk Management

3.1.1. Definition of Risk. Under normal circumstances, risk management refers to the scientific evaluation of the problems encountered by the enterprise according to the basic situation of the enterprise's operation and management activities, then defines whether it is a risk problem, and proposes corresponding adjustments and countermeasures [8, 9]. Risk management must run through all aspects of the enterprise, involve all employees of the enterprise, and be based on effective communication and docking. Through the implementation of effective risk management, it is ensured that the enterprise can maintain an all-round ideal state [10, 11]. Enterprise risk management begins with the formulation of strategic objectives and runs through the day-to-day business activities of an enterprise. It runs through every business process of an enterprise. The entire risk management process should comprehensively include prewarning, in-process control, and postevent supervision of risks.

3.1.2. Types of Risks. In the process of daily development, enterprises will encounter two different types of risks, namely, internal risks and external risks. Internal risks mainly refer to financial, operational, strategic, and other types of risks, while external risks refer to market and political risks [12]. The specific categories are shown in Table 1.

3.1.3. Industry Risk Factor. The development of an enterprise is not only affected by financing, industrial environment, market environment, management concepts, etc. but also by the industry prospects. Different industries have different development prospects, and the accompanying risks are also different. Next, it performs data analysis on the data of a credit company [13, 14]. The companies involved in the data are roughly divided into electronic information, power, software, materials, biomedical, electrical equipment, and energy conservation and environmental protection. An analysis of data on companies that were denied financing found that the companies that were denied loans on their credit history belonged to different industries and had different percentages. For example, software and electronic information accounts for about 6.7%, biomedicine accounts for 31% each, materials account for about 22.16%, electrical equipment accounts for 17.1%, and environmental protection and energy conservation account for 16.49%. It can be seen that the conjecture in this article is correct, and different industries come with different risks. If this factor can be

TABLE 1: Enterprise risk model.

Corporate risk	Process risk	Operational risk
		Customer satisfaction, human resources, product development, efficiency, production capacity, performance gaps, cycles, sources, channel effectiveness, partners, obsolescence/drop prices, compliance, business disruption, product service failure, environmental, health and safety, loss of trademark value
		Financial risk
		Interest rate, currency, equity, commodity, financial instrument, cash flow, opportunity cost, concentration, default, settlement, guarantee
	Decision information risk	Business decisions
		Pricing, contract commitment, assessment (management level), consistency, reporting compliance
		Financial decision
		Budgeting and planning, accounting information, financial report evaluation, taxation, pensions
		Strategic decision
		Environmental monitoring, business portfolio, valuation, organizational structure, assessment (strategic bureau), resource allocation, planning, life cycle
	Environmental risk	Market risk
		Competitors, customer needs, technological progress, industry
		Political risk
		Legal, government, regulatory

considered in the risk assessment, it can be more targeted and the accuracy can be improved.

On the other hand, enterprise development has its own lifeline, and the division methods are not the same. It is roughly divided into four stages: initial stage, growth stage, stable stage, and decline stage. Each stage has its own characteristics and also corresponds to different risks. To sum up, this paper considers two factors from the enterprise's own risk and the risk of the enterprise's life cycle and introduces a new factor industry risk coefficient. The definition in this paper is as follows:

$$T_{\text{Business}} = -M_1 * T_{\text{Self}} + M_2 * T_{\text{Stage}}. \quad (1)$$

Among them, the risks of the above-mentioned enterprises are named T_{Self}, and the risks in the enterprise life cycle are named T_{Stage}. The two parts are merged and named as the industry coefficient T_{Business}, and M_1 and M_2 are the corresponding weights, respectively.

3.2. Machine Learning

3.2.1. Classification Algorithm of Machine Learning. Machine learning (ML) is an interdisciplinary subject that integrates multi-domain knowledge such as probability theory and algorithm complexity theory. The research uses the computer to learn the corresponding human behavior according to the existing knowledge and behavior and excavates new knowledge while imitating to improve its own learning ability. It is the core of artificial intelligence, and computers achieve intelligent effects through machine learning. Machine learning has applications in various fields such as finance, the Internet, medicine, and biological robots, especially in the field of risk management. It can be said that it has spread to all branches of artificial intelligence. Machine learning can be divided into supervised learning and unsupervised learning [15–17]. Supervised learning is to have a classification label in the preclassification part and predict the data without a classification label according to the data training results with the original classification label. In the unsupervised learning mode, both the training data

and the test data have no class labels. The classic classification algorithm in machine learning is shown in Figure 1.

(1) Decision Tree. Decision tree is a typical predictive model. It is generated by the relationship between category attributes and categories and is classified according to attributes according to branch paths as rules. The classic decision tree algorithms mainly include ID3 and C4.5. C4.5 retains the advantages of ID3, optimizes the shortcomings, and is more flexible and accurate, such as replacing the gain with the gain rate and adding discretization and pruning steps.

(2) SVM. Support vector machine is mainly for linear inseparable objects. The principle is to use the non-linear mapping method to map the linearly inseparable low-dimensional input into a hyperplane high-dimensional space with the largest interval and transform it into a linearly separable case [18–20]. This type of classifier can minimize the error and maximize the margin interval while ensuring the accuracy of the classification, so it is widely used in statistical classification and regression analysis.

(3) Bayesian Classifier. A Bayesian classifier is actually a probabilistic model. The classification principle is to calculate the posterior probability according to the prior probability of the object and select the classification with the largest posterior probability as the final classification of the object [21]. Bayesian classifiers generally work on Bayesian networks. There are two types of nodes in the network. The nodes that are probabilistically dependent on each other are connected by arcs, and the nodes that are independent of each other have no dependencies. When the Bayesian classifier works, it first trains the data sample generation model and then uses the model to predict and classify the test data. Among them, Naive Bayesian is the most commonly used model, and it is also a model with relatively high accuracy.

(4) Nearest Neighbors (K-Nearest Neighbor, KNN). KNN is a very simple classification algorithm. The principle idea is to first classify the K objects that are the nearest neighbors of the predicted object in the feature space, find out the
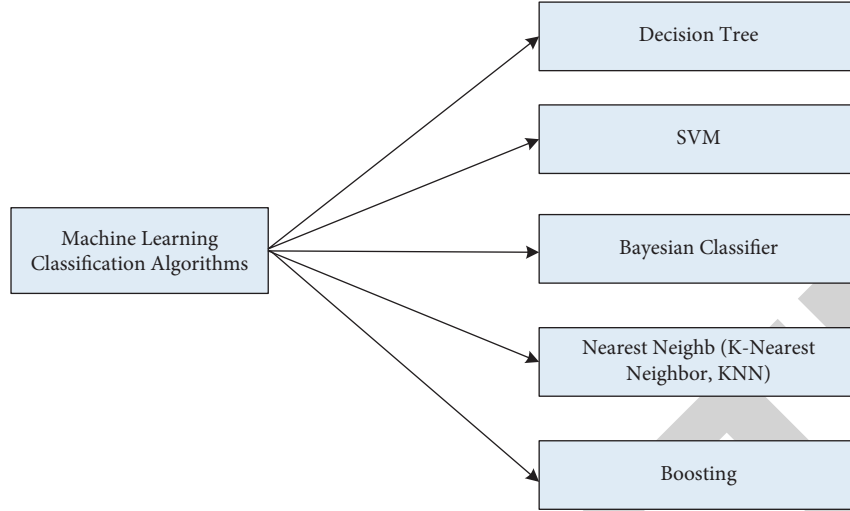
Figure 1: Machine learning classification algorithm.

category identifiers that most of them belong to, and classify the predicted object as the attribute of this category. KNN has better classification effect than other methods for the sample set to be divided with overlapping or overlapping class domains.

(5) Boosting. Boosting originated from the PAC (probably approximately correct) learning model proposed by Valiant. The idea of this algorithm is to gradually improve the accuracy of weak classification through multi-level training. It mainly obtains sample subsets by changing the sample set distribution, and then the sample subsets are trained by the weak classification algorithm to generate n subclassifiers [22]. The boosting framework algorithm performs weighted fusion of these n subclassifiers. A low weight is assigned to a good training result, and a high weight is assigned to a poor training result. In this way, those with poor training effect can get more attention in the next round of training and so on to produce a final result classifier.

3.2.2. Global Mutual Information. Mutual information (MI) is a measure used in information theory to calculate the correlation between two events. MI is usually interpreted in the context of the channel transmitting information, that is, there are three sources, a sink, and a disturbed channel. If the source sends a message u, assuming that the channel is free of noise interference, the sink receives all the information of u itself, completely eliminating the uncertainty of u. But since the channel always has interference in general, what the sink receives may be some kind of deformed h [23] caused by the interference. In fact, MI refers to "reduction in uncertainty." The essence of mutual information can be described by probability statistics. First, the probability $t(u)$ that the source sends out u is called the prior probability. Secondly, the posterior probability $t(u/h)$ can be used to represent the probability that the sink receives h and speculates that the source sends u. It defines the logarithm of the ratio of the posterior probability to the prior probability

of u as the mutual information (referred to as mutual information) of h to u, which is the following formula:

$$I(u; h) = \log \frac{t(u/h)}{t(u)}. \tag{2}$$

By extending information theory to any event, the simultaneous occurrence of event M and event N can be described by mutual information. It can be expressed as the following formula:

$$I(M; N) = \log_2 \frac{t(MN)}{t(M)t(N)} = \log_2 \frac{t(M|N)}{t(M)} = \log_2 \frac{t(N|M)}{t(N)}. \tag{3}$$

At the same time, the correctness of formula (2) is proved. Also, formula (3) can be equivalent to

$$I(M; N) = I(M) - I(M|N). \tag{4}$$

That is, MI$(M; N)$ = self-information (M) - conditional self-information $(M|N)$.

In this paper, for multiple events, in order to overcome the randomness of the mutual information $I(u_i; h_j)$, the statistical average of the mutual information $I(u_i; h_j)$ in the joint probability space $T(UH)$ is taken as a certain quantity—the average mutual information:

$$I(U, H) = \sum_{h \in H} \sum_{u \in U} T(u, h) \log \left(\frac{t(u, h)}{t_1(u)t_2(h)} \right). \tag{5}$$

The significance lies in Figure 2.

$$I(U; H) = B(U) - B(U|H),$$
$$I(U; H) = B(H) - B(H|U), \tag{6}$$
$$I(U; H) = B(U) + B(H) - B(UH).$$

According to the above explanation, MI can be extended to any event and can be for multi-event situations. In this paper, mutual information is applied to decision classification, and the correlation between attribute factors and classification can be obtained through mutual information.
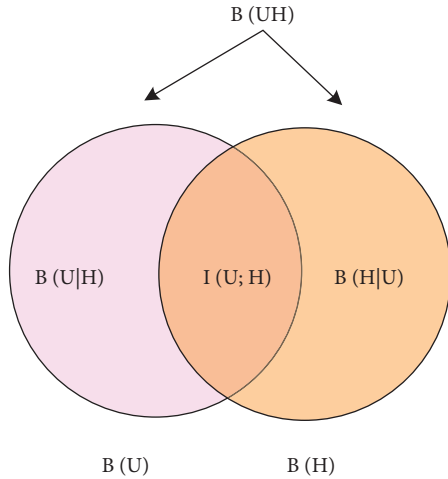
Figure 2: Mean mutual information meaning graph.

If the amount of information is larger, it means that the attribute factor has a great correlation with a certain classification category, and vice versa, which further guides the selection of attribute factors.

3.3. Risk Assessment Model Based on Machine Learning.

The above section introduces several classic classification algorithms in machine learning. In fact, some classical algorithms of machine learning have been widely introduced in risk assessment. The following introduces several risk assessment models in the field of machine learning.

3.3.1. Decision Tree Model.

The so-called decision tree is a tree-like structure generated depending on the strategy decision. It is mainly composed of non-leaf nodes, branches, and leaf nodes. Non-leaf nodes (including root nodes) correspond to attribute values, branch paths correspond to attribute selection rules, and leaf nodes represent classification class labels. The decision tree model includes a learning module and a test module, and the dataset is also divided into a learning set and a test set correspondingly. The learning module is responsible for the generation and pruning of decision trees, mining classification rules to generate decision trees. The essence of pruning is to denoise to ensure that the desired large and small-scale decision tree is generated. The prediction module applies the generated model to the test set. Because of the simple and intuitive characteristics of decision tree, it is applied to various fields, such as weather forecast, personal mortgage credit assessment, and risk assessment. The language description of decision tree is too abstract. The following will analyze the decision tree model with an example.

Some people want to use the weather forecast to predict whether people decide to travel, because the weather is good or bad largely determines whether people travel or not. Weather conditions are sunny, cloudy, and rainy; other factors are temperature, air humidity, air quality, and wind. Using the known data to build the decision tree shown in Figure 3, and decide whether the day is suitable for travel

according to the classification of the weather. The selection of splitting attributes in the decision tree generation process is the most critical part. The classic ID3 algorithm selects splitting attributes based on information gain, while C4.5 changes information gain to information gain rate and discretizes continuous values. Next, use the ID3 algorithm for modeling: traverse all attributes, calculate the information gain of all attributes, and select the root node with the largest value, which is the weather (Outlook) in the picture. Assuming that the information gain of a certain attribute Outlook is calculated, it is divided into three steps:

(1) First calculate the entropy of Outlook:

$$\text{Entropy}(\text{Outlook}) = -\sum_{e=1}^{n} p_e \log_2 p_e. \tag{7}$$

(2) It refines purity according to branch:

$$\text{Entropy}(B)_{\text{Outlook}} = -\sum_{g=1}^{l} \frac{B_g}{B} \text{Entropy}(B_g). \tag{8}$$

(3) It calculates the information gain:

$$\text{Gain}(\text{Outlook}) = Entropy(\text{Outlook}) - \text{Entropy}(B)_{\text{Outlook}}. \tag{9}$$

Branches are created based on the different values computed for the property. It repeats the first two steps until a decision tree is generated and covers all attributes.

3.3.2. Neural Network Model.

Artificial neural network is a tool to simplify complex problems by simulating the way of human thinking. It is a complex network system with self-learning and self-adaptive ability composed of a large number of basic elements (neurons) interconnected. The structure and function of each neuron are relatively simple, but the system behavior generated by the combination of large numbers of neurons is very complex. In layman's terms, an artificial neural network is composed of multiple neural layers, and a neural layer is composed of multiple neurons. The artificial neural network model is shown in Figure 4, which consists of four parts: input, weight, summation function, and output. The main research work of neural network focuses on biological prototype research, establishment of theoretical model, network model and algorithm research, artificial neural network application system, and so on.

At present, the most mature in the field of machine learning is the error backpropagation network (BP neural network). It is composed of input layer unit, hidden layer unit, and output layer unit [24]. A typical BP neural network with three-layer structure is shown in Figure 5. The mapping relationship between the input and output of BP neural network is non-linear. The subordinate nodes in each layer in the network composition are zero-connected, and the nodes between layers are guaranteed to be fully connected.
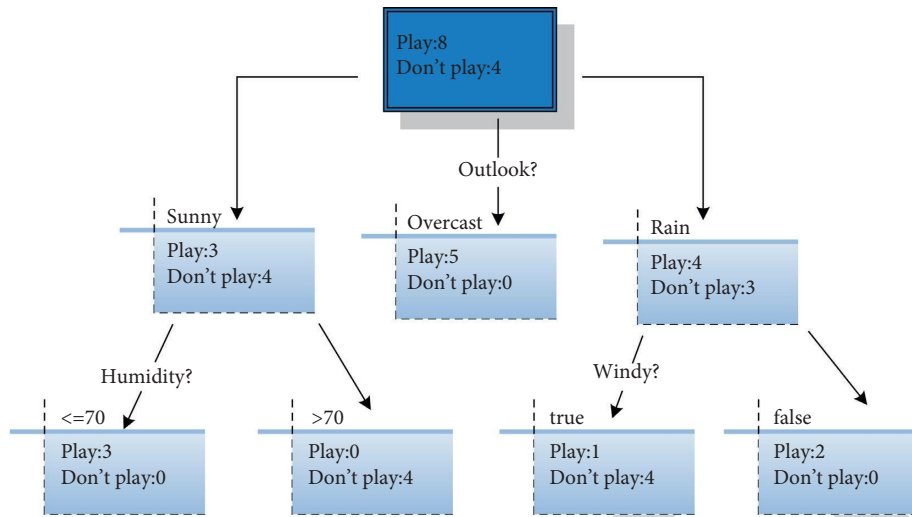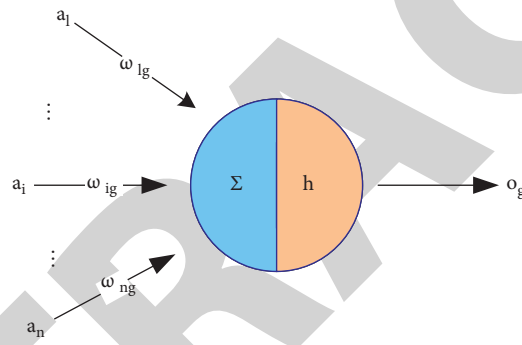
FIGURE 3: Decision tree model example.

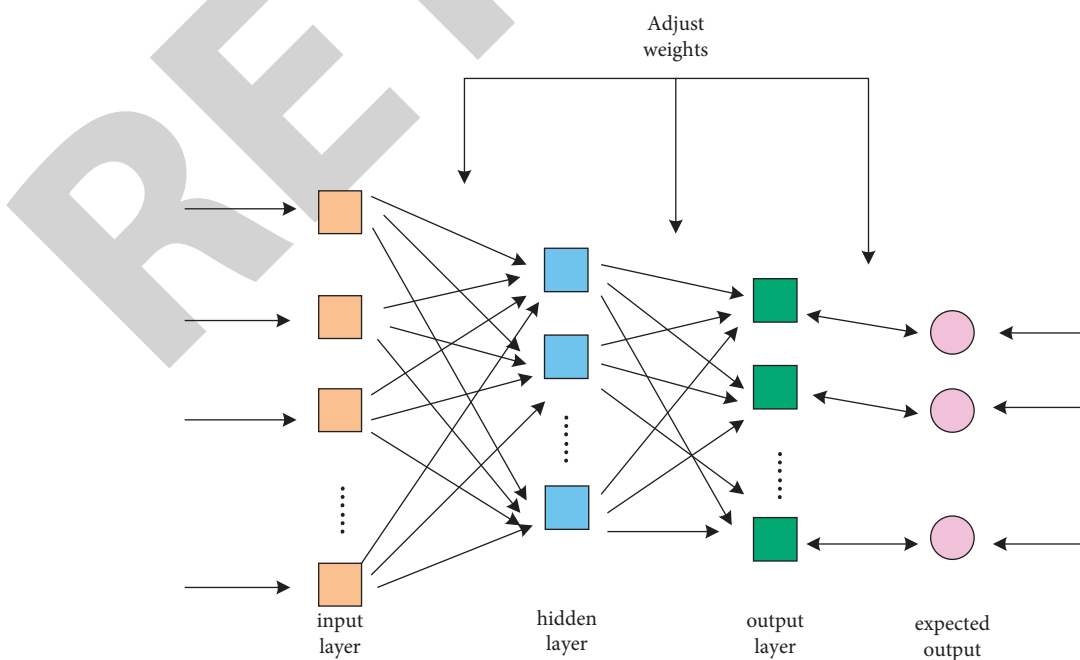FIGURE 4: Artificial neural network.

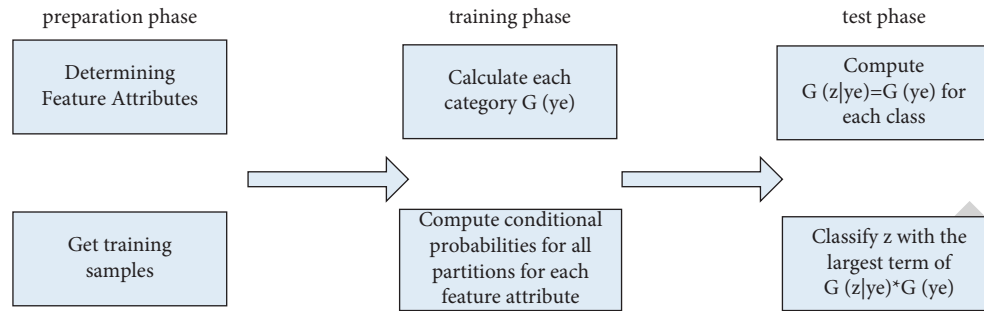
FIGURE 5: Three-layer neural network.

Figure 6: Schematic diagram of the Naive Bayes classification process.

The adaptation of the system is achieved by changing the weight values [25].

The main idea of BP is to input learning samples and establish a model through three-layer network learning to determine the relationship between input and output. In the learning phase, the BP algorithm is used to continuously adjust the weights and biases of the network. This process mainly uses the gradient descent method to continuously modify the thresholds and weights to ensure the difference between the output value and the expected value. The condition for the end of the training process is that the difference between the squared error of the output and the specified threshold is less than zero, at which time the weights and biases of the network should be saved [26]. BP neural network can approximate continuous functions infinitely, so the BP network model is widely used in pattern classification, non-linear modeling, and other fields. But the model still has shortcomings. First, the model generation process requires a large amount of data to support repeated adjustments of weights and thresholds. Second, local minimum or overfitting is easy to occur when the number of samples is large. Third, the system has weak anti-noise interference capability.

3.3.3. Naive Bayes Model. Naive Bayes classification and decision trees are the two most widely used classifiers. Compared with the decision tree model, the Naive Bayes classifier is based on the well-known Bayes theorem of probability theory. It has a solid mathematical foundation, and the advantage is that the Naive Bayes model algorithm is simple and easy to understand. Its ideological basis is that when a classification event is encountered and there is not much known information, the case with the highest probability will be judged as the final classification result [27]. In addition, the Naive Bayes classification results are relatively stable and less sensitive to missing data. Figure 6 shows a schematic diagram of the Naive Bayes model.

As can be seen from the figure, the entire Naive Bayes classification process is divided into three stages. In the preparation stage, the feature attributes are determined by the data samples and the training samples are obtained to provide input for the next stage. In the training stage, the output of the previous stage is used as input, and the training model is generated by calculating the occurrence probability

of each category and the conditional probability between attributes and categories. The last stage is to apply the training model obtained in the previous stage to the test dataset and use the item with the highest conditional probability of the attribute to be tested as the final classification category [28]. Although the Naive Bayes model has a simple principle and high accuracy, it still has shortcomings. It does not have the smallest error rate among various classification methods as in theory, but in practice there is an error rate. This is because a prerequisite of the Naive Bayes model is to assume complete independence between attributes. This assumption is too ideal, which limits its application scenarios to some extent [29].

4. Enterprise Risk Management Assessment Experiment

4.1. Experimental Design of Enterprise Risk Management

4.1.1. Experimental Data. The experimental data come from a partner credit company in the National Science and Technology Support Project. The data are extracted from the data of 56 SMESTs in the company's database from 2010 to 2015 (3 quarters/year), with a total of 1008 items. The data horizontally contain 22 attribute factors, including financial index factors and the label of whether it has passed the audit [30]. In this experiment, 80% of the experimental data (800 randomly selected) are used as the training dataset, and 20% (the remaining 208) are used as the test dataset. The specific data distribution is shown in Table 2.

4.1.2. Experimental Method

(1) Decision Tree. The algorithm process of the traditional decision tree method will not be repeated here. The decision tree program in R language can act on three sets of data, respectively [31]. This experiment uses the R package rpart, where the rpart function is used for training data, returns a model, and passes it to the predict function for prediction. The model is rpart (formula, data, weights, subset, na.action = na.rpart, method, model = FALSE, x = FALSE, y = TRUE, cost, ...).

(2) MI-DT (Decision Tree Based on Global Mutual Information). The key to the decision tree method based on global

TABLE 2: Experimental data distribution.

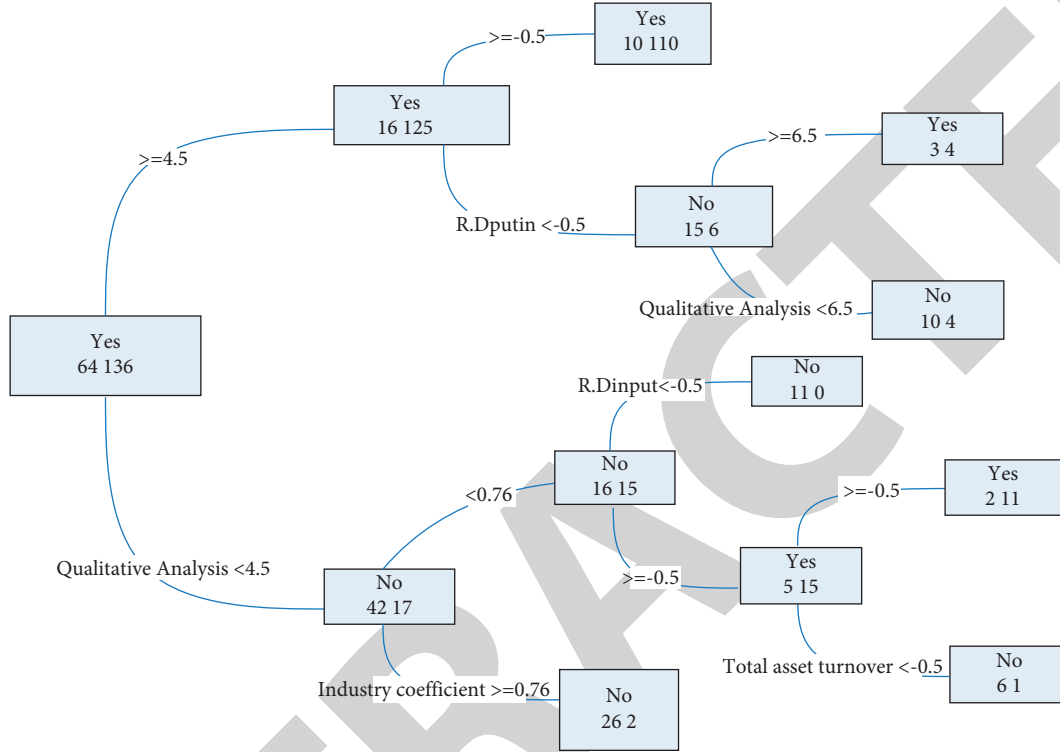|	Yes (people)	No (indivual)	Yes/no (proportion)
Sample set (1006)	675	331	2.04 : 1
Training set (802)	543	259	2.10 : 1
Test set (209)	134	75	1.79 : 1

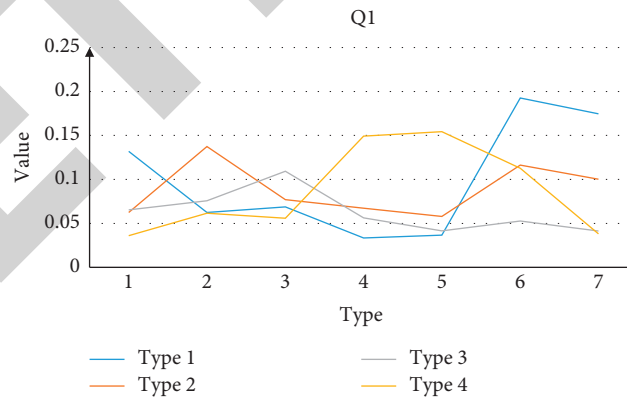FIGURE 7: Decision tree with training set of 400.

FIGURE 8: The training set refers to the MI data corresponding to 400.

mutual information lies in the selection of the splitting factor. For the 7 factors obtained above, the mutual information (MI) between the factors in the matrix Arr_new, that is, the MI dataset Q, is first calculated from the data. The diagonal line is the correlation between the factor and the class ID. Off-diagonal is the redundancy between factors with the participation of class identification.

According to the MI dataset Q, the calculation satisfies

$$g = \begin{matrix} \max, & \{Y^T Q Y\}, \\ \text{st}, & Y_i \geq 0. \end{matrix} \tag{10}$$

Among them, $Y = \{Y_i\}$, $i = 1 \ldots 11$. That is, the weight of each factor in the global state is obtained. Formally, it can be equivalently transformed into a quadratic programming problem to obtain the optimal solution $Y = [Y_i]$ that satisfies the conditions.
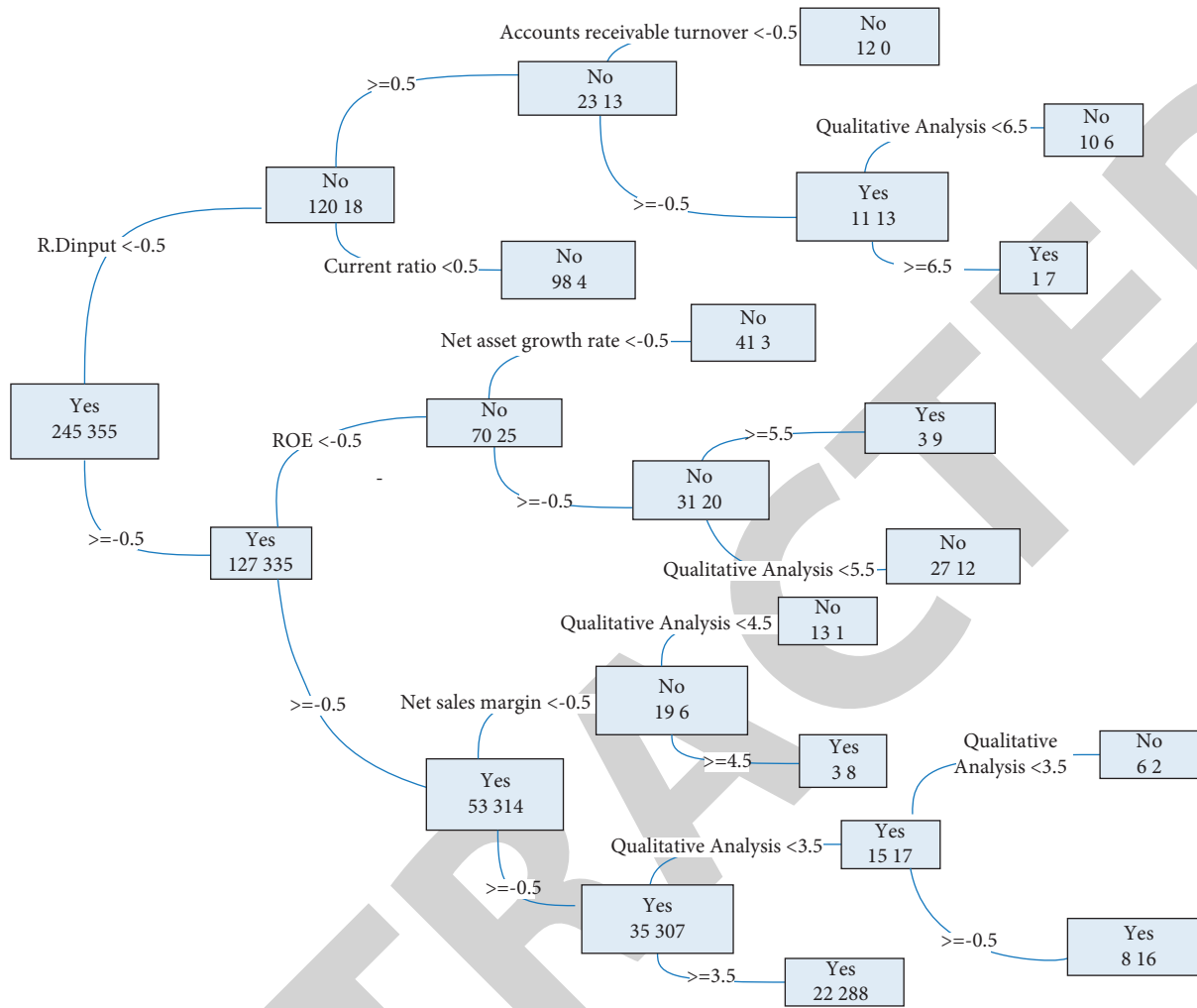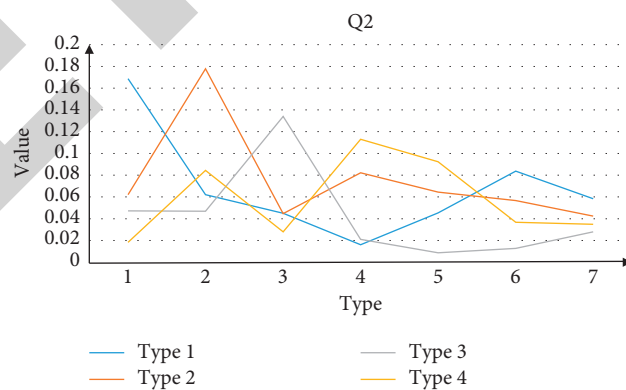
Figure 9: Decision tree with training set of 600.

Figure 10: The training set refers to MI data corresponding to 600.

According to the obtained $Y = [Y_i]$ and according to the weight in y, the new selection combination order of the attribute factors is obtained, and this selection combination is used as the attribute selection process of the decision tree. According to this order, the decision tree generation rules are obtained, and then the decision tree with the global optimal solution is generated [32].

4.2. Enterprise Risk Assessment Experiment Results. By using three sets of experimental data, the traditional decision tree model and the improved MI-DT are compared. Since the principle of the decision tree model algorithm was introduced in the previous article, the experimental results of this method are directly displayed in this experiment.
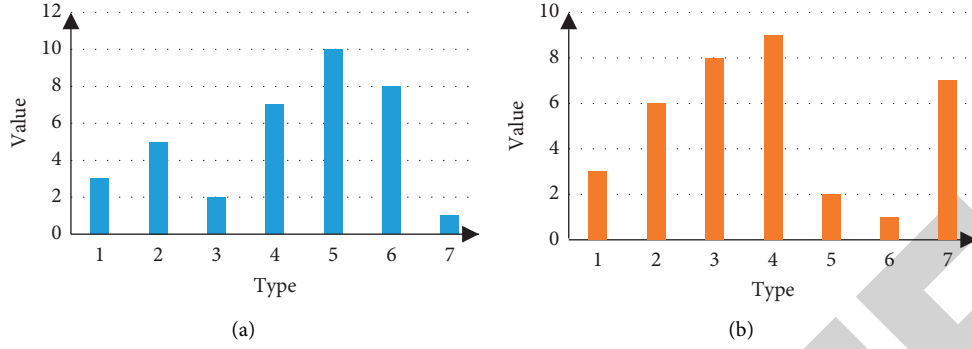
(a)

(b)

FIGURE 11: $Q1$ and $Q2$ indicator comparison charts. (a) $Q1$ index. (b) $Q2$ index.
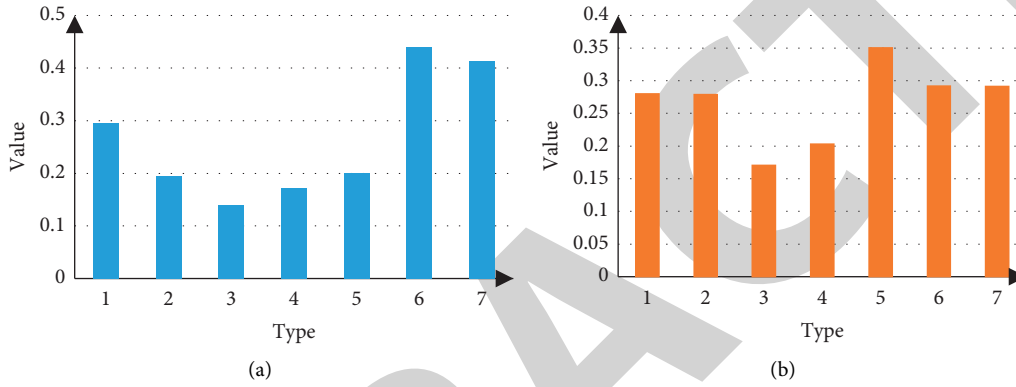

(a)

(b)

FIGURE 12: Comparison of the best factor combinations in $Q1$ and $Q2$. (a) $Q1$ best factor combination. (b) $Q2$ best factor combination.

4.2.1. Experiment 1: 400 Pieces of Training Data and 100 Pieces of Test Data

(1) Figure 7 shows the traditional decision tree method: accuracy = 0.81.

(2) Improved decision tree method: the accuracy is 0.842

It is tested according to the above experimental steps, and the MI dataset is obtained as Q_1, and Q_1 is a $4*7$ dataset, as shown in Figure 8.

The global optimal factor combination weight obtained according to $Q1$ is

$$z = [0.2953, 0.1943, 0.14, 0.1731, 0.2008, 0.4392, 0.4143]^T. \quad (11)$$

After sorting, the corresponding index order is from small to large.

$$z = [P_3 , P_4 , P_2 , P_5 , P_{10} , P_8 , P_1]. \quad (12)$$

4.2.2. Experiment 2: 600 Pieces of Training Data and 154 Pieces of Test Data

(1) Figure 9 shows the traditional decision tree method: accuracy = 0.8316.

(2) Improved decision tree method: accuracy is 0.905.

It is tested according to the above experimental steps, and the MI dataset is obtained as Q_2, as shown in Figure 10.

The global optimal factor combination weight obtained according to Q_2 is

$$z = [0.28, 0.2789, 0.1708, 0.2030, 0.3503, 0.2921, 0.2912]^T. \quad (13)$$

After sorting, the corresponding index order is from small to large.

$$z = [P_3 , P_4 , P_8 , P_9 , P_2 , P_1 , P_7 , P_6]. \quad (14)$$

4.3. Risk Assessment Experiment.
According to the above two sets of implementations, it can be seen that

(1) With the change of samples, for the traditional decision tree method, the selection of attribute factors is not the same, and the indicators also change, as shown in Figure 11. In the improved method, the classification attributes tend to be consistent as the number of samples increases. The result is an optimal set of factors. The first 4 digits of the factor combination are all $[P_3 , P_4 , P_8 , P_9]$, and the last 4 digits are all $[P_6 , P_{11} , P_5 , P_{10}]$, and the newly introduced indicator, the industry coefficient, has a large weight. This also just proves the necessity of introducing the industry risk coefficient in this paper, and the sufficient industry coefficient has a great influence on the enterprise risk assessment. The introduction of the industry coefficient strengthens the objective of the evaluation [33].
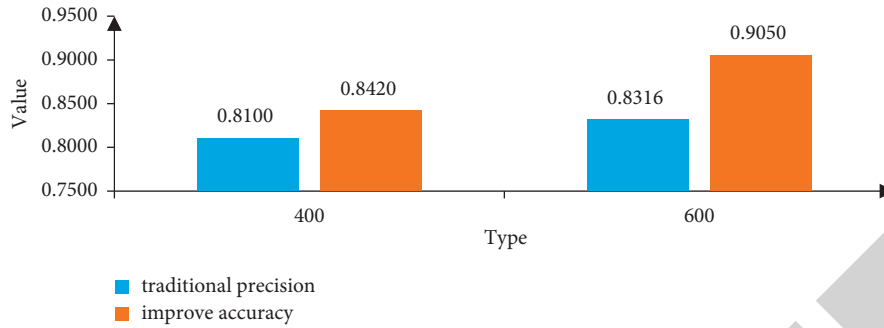
Figure 13: Comparison of the training sets of 400 and 600 methods, respectively.

It can be seen from Figure 12 that the accuracy of the prediction results varies with the number of samples. When the value is 6, the difference in the amount of data is the largest. Also, basically, with the increase of training samples and realistic samples, the accuracy rate continues to increase. In the data of the experimental results, as the number of samples increases from 400 to 600, the accuracy of the traditional method is 0.81 and 0.8316, respectively. The accuracy of the improved method is 0.842 and 0.905, respectively. From the experiment, it is found that the accuracy of the test sample decreases with the increase of the training sample size to a certain extent in the traditional decision tree method. This is because the training samples have overfitting, while the improved method of the same data samples does not have overfitting. Figure 13 shows a comparison diagram of two methods with training sets of 400 and 600, respectively.

5. Discussion

This paper studies the process of enterprise comprehensive risk assessment from the perspective of risk management and attempts to build an enterprise risk management model based on industry risk coefficients. Firstly, the ISM model is used to analyze the factors affecting the development of SMEST, and it is concluded that the most fundamental factor is R&D investment, which is in line with the characteristics of SMEST. In addition to financing problems, the most direct reason is the industry itself, which provides a direction for optimizing risk assessment. Based on the results of the first part, the new index factor of industry risk coefficient is introduced into machine learning to refine risks. This new factor is composed of two parts: the industry's own risk and the risk in the lifeline of the industry. These two parts are quantified and then weighted and integrated to obtain the measurement of the industry risk coefficient. In addition, the correlation analysis and screening of all the index factors of the index system are carried out to reduce the index factors and reduce the computational complexity.

6. Conclusions

With the increasingly complex environment in which the company operates, the number of businesses in the company is increasing day by day, the uncertainty and risks in the operation are getting higher and higher, and the application of machine learning in related fields is not mature, and it is not widely used in Chinese companies. Without establishing a sound risk management system, the risk management theory has not really played a role in the company. At present, China's theoretical and practical research on risk management is still at the initial level, and there are many problems that need to be solved urgently. This research analyzes and summarizes the actual situation of the combination of risk assessment and machine learning in risk management theory. However, based on the actual situation, there is still more room for research and analysis of this case. Although there is a detailed overview of its basic framework in the research, because of the differences between theory and practice, it should be flexibly applied in reality according to the actual situation of the enterprise. In the future operation and management of Chinese enterprises, risk management is still an important link that enterprise operators need to pay attention to. Regardless of the size of an enterprise, enterprise risks are gradually formed when the enterprise reaches a certain stage of operation. Therefore, it is one of the main goals for the future development of Chinese enterprises to build a risk management system that suits their own situation and is perfect.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that they have no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

[1] Y. Lin, R. D. Macminn, R. Tian, and J. Yu, "Pension risk management in the enterprise risk management framework," *Journal of Risk & Insurance*, vol. 84, no. S1, pp. 345–365, 2017.

[2] M. Abkowitz and J. Camp, "Structuring an enterprise risk assessment protocol: traditional practice and new methods," *Risk Management and Insurance Review*, vol. 20, no. 1, pp. 79–97, 2017.

[3] R. Valanarasu and A. Christy, "Risk assessment and management in enterprise resource planning by advanced system engineering theory," *International Journal of Business Intelligence and Data Mining*, vol. 13, no. 1-3, pp. 3–14, 2018.

[4] J. Lv and J. Rong, "Virtualization security risk assessment for enterprise cloud services based on stochastic game nets model," *IET Information Security*, vol. 12, no. 1, pp. 7–14, 2017.

[5] P. L. He, C. X. Zhao, Q. Y. Dong, S. B. Hao, P. Xu, and J. Zhang, "Application of occupational hazard risk index model in occupational health risk assessment in a decorative coating manufacturing enterprises," *Zhonghua lao dong wei sheng zhi ye bing za zhi = Zhonghua laodong weisheng zhiyebing zazhi = Chinese journal of industrial hygiene and occupational diseases*, vol. 36, no. 1, pp. 18–21, 2018.

[6] A. Buczak and E. Guven, "A survey of data mining and machine learning methods for cyber security intrusion detection," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 1153–1176, 2017.

[7] J. Finance, "Enterprise risk management and solvency II: the system of governance and the own risk and solvency assessment," *The Journal of Risk Finance*, vol. 21, no. 4, pp. 317–332, 2020.

[8] F. J. Martínez-López, Y. Li, C. Feng, and D. López-López, "Buying through social platforms," *Journal of Organizational and End User Computing*, vol. 33, no. 4, pp. 70–93, 2021.

[9] W. Li, Z. Yu, L. Zhu, P. Li, X. Yang, and S. Zhu, "Laparoscopic Roux-en-Y gastric bypass in Type 2 diabetes mellitus patient with low body mass index and gastric stromal tumor: a case report," *Zhong nan da xue xue bao. Yi xue ban = Journal of Central South University. Medical sciences*, vol. 46, no. 1, pp. 98–103, 2021.

[10] K. Mahmood, E. Shevtshenko, T. Karaulova, and T. Otto, "Risk assessment approach for a virtual enterprise of small and medium-sized enterprises," *Proceedings of the Estonian Academy of Sciences*, vol. 67, no. 1, pp. 17–27, 2018.

[11] M. Feng, "The effects of techno-stress in the role stress context applied on the proximity manager performance:conceptual development and empirical validation," *Journal of Organizational and End User Computing*, vol. 33, no. 1, pp. 1–18, 2021.

[12] H. Sun, Z. Sun, Z. Sun, and Y. Huang, "Equilibrium investment and risk control for an insurer with non-Markovian regime-switching and no-shorting constraints," *AIMS Mathematics*, vol. 5, no. 6, pp. 6996–7013, 2020.

[13] P. Wang and W. Han, "Construction of a new financial E-commerce model for small and medium-sized enterprise financing based on multiple linear logistic regression," *Journal of Organizational and End User Computing*, vol. 33, no. 6, pp. 1–18, 2021.

[14] W. Han, P. Wang, and H. Dong, "Influence of egoistic and altruistic bequest motives on the willingness to participate in reverse mortgages in China∗," *Asian Economic Journal*, vol. 34, no. 4, pp. 430–463, 2020.

[15] S. Nagi Alsubari, S. N. Deshmukh, A. Abdullah Alqarni et al., "Data analytics for the identification of fake reviews using supervised learning," *CMC-Computers, Materials & Continua*, vol. 70, no. 2, pp. 3189–3204, 2022.

[16] S. Sengan, O. I. Khalaf, P. Vidya Sagar, D. K. Sharma, L. Arokia Jesu Prabhu, and A. A. Hamad, "Secured and privacy-based IDS for healthcare systems on E-medical data using machine learning approach," *International Journal of Reliable and Quality E-Healthcare*, vol. 11, no. 3, pp. 1–11, 2021.

[17] S. Sengan, O. I. Khalaf, G. R. Rao, D. K. Sharma, K. Amarendra, and A. A. Hamad, "Security-aware routing on wireless communication for E-health records monitoring using machine learning," *International Journal of Reliable and Quality E-Healthcare*, vol. 11, no. 3, pp. 1–10, 2022.

[18] H. Elamir, "Enterprise risk management and bow ties: going beyond patient safety," *Business Process Management Journal*, vol. 26, no. 3, pp. 770–785, 2019.

[19] Z. Wan, Y. Dong, Z. Yu, H. Lv, and Z. Lv, "Semi-supervised support vector machine for digital twins based brain image fusion," *Frontiers in Neuroscience*, vol. 15, Article ID 705323, 2021.

[20] T.-Y. Kim, H. Ko, S.-H. Kim, and H. D. Kim, "Modeling of recommendation system based on emotional information and collaborative filtering," *Sensors*, vol. 21, no. 6, p. 1997, 2021.

[21] B. Jiang, J. Li, and S. Shen, "Supply Chain risk assessment and control of port enterprises: qingdao port as case study," *Asian Journal of Shipping and Logistics*, vol. 34, no. 3, pp. 198–208, 2018.

[22] J. Guo, J. Xu, and W. Liao, "Risk control of the cascading failure of multimodal transport network considering uncertain disturbance factors1," *Journal of Intelligent and Fuzzy Systems*, vol. 39, no. 1, pp. 1–12, 2020.

[23] P. Vila, A. Mota, J. Bastos, and L. Patrício, "Framework for a risk assessment model to apply in Virtual/Collaborative Enterprises," *Procedia Computer Science*, vol. 181, no. 34, pp. 612–618, 2021.

[24] Z. Xiaofeng, L. Lianghai, and X. Yabo, "Study on static risk assessment model of expressway operation companies," *IOP Conference Series: Earth and Environmental Science*, vol. 283, no. 1, p. 12051, 2019.

[25] H. Guan, F. Xu, H. Fu et al., "Research on risk control of upgraded design and development changes of laser treatment equipment after registration," *Zhongguo yi liao qi xie za zhi = Chinese journal of medical instrumentation*, vol. 44, no. 4, pp. 346–352, 2020.

[26] Y. S. Zhang, S. J. Han, J. Wang et al., "[Emission characteristics analysis and health risk assessment of unorganized VOCs in the carbon industry, zhengzhou]," *Huan jing ke xue= Huanjing kexue*, vol. 40, no. 11, pp. 4847–4855, 2019.

[27] O. M. Kolodiziev, S. M. Kyrkach, and Y. M. Ohorodnia, "Features of formation of models of the risk control on the part of client taking into account the likelihood of implementation of network schemes," *Business Inform*, vol. 12, no. 515, pp. 377–388, 2020.

[28] H. L. Dong, "Case study of risk control applying bow-tie methodology," *Korean Journal of Hazardous Materials*, vol. 8, no. 1, pp. 65–72, 2020.

[29] G. Stephen, "Book review: environmental decision making and risk management: selected essays by ian langford," *Environmental Practice*, vol. 8, no. 4, pp. 256-257, 2017.

[30] Q. Shao, M. Shao, Y. Bin, P. Zhu, and Y. Zhou, "Flight recovery method of regional multiairport based on risk control model," *Mathematical Problems in Engineering*, vol. 2020, no. 11, 18 pages, Article ID 7105381, 2020.

[31] H. Thorisson, M. Alsultan, D. Hendrickson, T. L. Polmateer, and J. H. Lamber, "Addressing schedule disruptions in business processes of advanced logistics systems," *Systems Engineering*, vol. 22, no. 1, pp. 66–79, 2019.

[32] S. Cheng, S. Jianfu, M. Alrasheedi, P. Saeidi, A. R. Mishra, and P. Rani, "Correction to: a new extended vikor approach using q-rung orthopair fuzzy sets for sustainable enterprise risk management assessment in manufacturing small and medium-sized enterprises," *International Journal of Fuzzy Systems*, vol. 23, no. 8, p. 2504, 2021.

[33] H. Keto, J. Palomki, and H. Jaakkola, "Expanding on the process perspective in software process improvement practices," *System Dynamics Review*, vol. 10, no. 2-3, pp. 175–188, 2019.