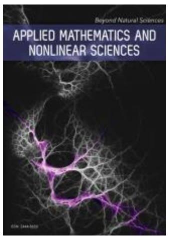
sciendo Applied Mathematics and Nonlinear Sciences, 9(1) (2024) 1-20



Applied Mathematics and Nonlinear Sciences

https://www.sciendo.com

Research on Corporate Governance and Internal Audit Wisdom Building under Financial Sharing Model Based on Logistic Modeling

Hongli Wu[1,†]

1. Guangdong Innovative Technical College, Dongguan, Guangdong, 523000, China.

Submission Info

Communicated by Z. Sabir
Received April 9, 2024
Accepted July 3, 2024
Available online August 5, 2024

Abstract

In recent years, the assessment of corporate financial risk has become increasingly significant for banks. Therefore, studying models for corporate financial risk assessment holds substantial practical importance. This paper combines the Logistic model and the Lasso model based on their basic principles to construct an improved Lasso-Logistic regression model. Immediately after that, this paper selects 15 representative indexes from the four aspects of the enterprise's profitability, solvency, operating ability, and growth ability as the indexes to respond to the company's financial situation and extracts 4 public factors after factor analysis and analyzes them using the Lasso-Logistic regression model designed in this paper with these 4 public factors as the variables. The results show that the coefficients of public factors F1, F2, F3, and F4 are -2.9513, -1.8347, -1.9659 and -2.2714, respectively, and the coefficients of the four public factors are negative, and the classification accuracy of the Lasso-Logistic combination model in this paper is 89.46%, the misclassification rate of the first category is 6.21%, and the F_1 score, R_{S2} score and AUC values are overall better than the two single models of Lasso and Logistic, the Lasso-Logistic model designed in this paper can well help enterprises assess their own financial risk and make targeted decisions.

Keywords: Logistic; Lasso; Regression model; Financial risk; Factor analysis.
AMS 2010 codes: 00A71

†Corresponding author.
Email address: whlpwf009@163.com

ISSN 2444-8656
https://doi.org/10.2478/amns-2024-1920

1 Introduction

Since the end of the twentieth century, there has been a wave of multinational operations and mergers in enterprises, and now the operating environment of enterprises is deeply affected by mergers, globalization, and other activities, which objectively requires that enterprises must improve the operational efficiency and unify their business standards in order to be in a favorable competitive position in the market environment [1-4]. Along with the rapid changes in the operating environment of enterprises, many enterprises in the financial aspects of some common problems, such as high financial costs, business, financial departments, the information communication barriers between the full use of the value of financial information, etc. [5-7]. In this environment, companies urgently need to seek an efficient financial management model, so as to give full play to the role of financial management in enterprise development [8-10]. In the early days, some companies separated their financial functions and set up new organizations to deal with financial operations, and Ford's shareholders, as well as executives, began to establish financial shared service centers in order to solve the company's crisis and save costs. The financial shared service model is different in financial processes and business processes compared to the traditional management model [11-13]. Enterprises under the financial sharing model can unify the operation process of subsidiaries and can extract financial data of subsidiaries at any time, compared to the financial management of enterprises under the financial sharing model, which has a higher level and efficiency. In terms of financial processes, enterprises under the financial sharing model centralize the accounting work, thus strengthening the financial control of subsidiaries. From the point of view of the role of enterprise internal audit [14-15]. As an extension of the internal control system, an enterprise internal audit can supervise and appraise the routine business and key processes of enterprise operation and management. Enterprise internal audit institutions and personnel rely on professional technical methods to ensure the establishment of internal constraints [16-18]. In the continuous expansion of the scope of business operations, the business involved is more and more extensive, the enterprise control chain is getting longer and longer, and the business risk is also getting bigger and bigger. The scope of internal audit has been expanded from financial income and expenditure to all aspects of enterprise production, operation, and management activities, and the role of internal audit in the continuous expansion of enterprises is also growing. Through the financial sharing mode of corporate governance and internal audit to build a Logistic model, to analyze the characteristics of the wisdom mode on its impact, and in this way to make corporate governance and internal audit prediction model [19].

Literature [20] found that the correlation between the establishment of a finance committee and financial performance is not very strong, but if the investment is in a highly independent finance committee, the performance is the opposite and helps to reduce the cost of debt in companies with small audit committees. Literature [21] analyzed New Zealand institutional investment directors' perceptions of the role of audit committees in investment companies and found that there are differences in their perceptions and that audit committees can have an impact on corporate governance and change. Literature [22] used a critical literature review and a multi-theoretical governance perspective to examine the limitations of independent financial statement design, where auditors can make judgments and develop a sound audit plan based on audit risk. Literature [23] mainly investigates the model of enterprises building cloud accounting financial shared management in the context of big data, which effectively improves the financial control ability, reduces the cost of enterprise operation, and enhances the competitiveness level of enterprise financial management. Literature [24] Based on the LightGBM algorithm model, the multi-dimensional data of the enterprise financial market is recognized and calculated to predict the financial risk, and the results show that the model has a more accurate prediction performance and can effectively determine the user's default risk. Literature [25], in order to improve the modern enterprise management system, designed a decision tree-based enterprise financial risk analysis and prediction model that tries to avoid risks in advance, which provides an effective reference for enterprise financial risk control methods. Literature [26] proposes an innovative method of enterprise financial audit centralized control model in the context of Industry 4.0 to reduce the cost of enterprise financial audit centralized control while improving efficiency and sound internal control. Literature [27] describes how to securely and efficiently share Chief Information Office cloud resources, including cluster isolation, automation, and identity and access, providing ideas for enterprise user management and data migration. Literature [28] aims to explore the relationship between some corporate governance mechanisms and financial reporting stability; conservatism reduces management involvement and is significantly related to equity concentration, but there is no significant relationship between shareholder directors and conservative accounting. Literature [29] reveals the trend of change in corporate governance and property relations in the digital society because digital assets, human capital, and others are growing in innovation, and the role of the human factor must be strengthened in order to achieve greater breakthroughs in technology.

An improved Lasso-Logistic combined regression model for predicting financial risk in corporate governance is created through the combination of Logistic and Lasso single regression models in this paper. Immediately after that, this paper selects 15 indicators from the four aspects of corporate profitability, solvency, operating capacity, and growth capacity to represent the financial situation of enterprises and uses factor analysis to extract four public factors as variables from these 15 indicators. Taking the public data of the annual financial audit reports of 40 enterprises of different sizes in China from 2021 to 2023 as the original data, the Lasso-Logistic regression model designed in this paper was used to analyze and explore the relationship between the four public factors and the financial risk of enterprises. This paper demonstrates the superior performance of its model by comparing the Lasso-Logistic model to commonly used single models such as Lasso, Logistic regression, and SVM in terms of classification effects.

2 Method

2.1 Basic models for financial risk assessment

2.1.1 Logistic regression model

1) Logistic distribution function

Suppose there exists a continuous variable y_i^* related to the magnitude of the likelihood of an event occurring, with a range of values from negative infinity to positive infinity. This event occurs when the value of this variable exceeds a critical point θ (one might as well set $\theta = 0$), i.e., there is:

When $y_i^* > \theta$, the event occurs, $y_i = 1$. In all other cases, the event does not occur, $y_i = 0$.

If it is assumed that there is a linear relationship between the dependent variable y_i^* and the independent variable x_i, the error term ε_i has a logistic distribution, i.e.:

$$y_i^* = \alpha + \beta x_i + \varepsilon_i \tag{1}$$

From Eq. (1), it can be obtained:

$$\begin{aligned} P\left(y_i=1 \mid x_i\right) &= P\left(\alpha+\beta x_i+\varepsilon_i>0\right) \\ &= P\left(\varepsilon_i>-\alpha-\beta x_i\right) \end{aligned} \tag{2}$$

Since the Logistic distribution is symmetric, equation (2) can be rewritten as:

$$\begin{aligned} P\left(y_i=1 \mid x_i\right) &= P\left(\varepsilon_i \leq \alpha+\beta x_i\right) \\ &= F\left(\alpha+\beta x_i\right) \end{aligned} \tag{3}$$

Where F is the distribution function of ε_i. When ε_i obeys a standard Logistic distribution with mean 0 and variance $\pi^2/3$, the cumulative distribution function of ε_i is:

$$\begin{aligned} P\left(y_i=1 \mid x_i\right) &= P\left(\varepsilon_i \leq \alpha+\beta x_i\right) \\ &= \frac{1}{1+\exp\left(-\alpha-\beta x_i\right)} \end{aligned} \tag{4}$$

This cumulative distribution function is known as the Logistic distribution function, and its graph is an S-shaped curve, which is a centrosymmetric graph that grows faster near the center point and slower at both ends.

2) Binomial Logistic Regression Model

We can construct a logistic regression model using the logistic function by considering ε_i in Eq. (4) as a linear function with respect to the independent variable x, i.e:

$$\varepsilon_i=\alpha+\beta x_i \tag{5}$$

Where α is called the regression intercept, β_i is called the regression coefficient, and x_i is the independent variable. At this point, equation (4) can be rewritten as:

$$P\left(y_i=1 \mid x_i\right)=\frac{\exp\left(\alpha+\beta x_i\right)}{1+\exp\left(\alpha+\beta x_i\right)} \tag{6}$$

$$P\left(y_i=0 \mid x_i\right)=\frac{1}{1+\exp\left(\alpha+\beta x_i\right)} \tag{7}$$

Given the input data x_i, two conditional probability values $P\left(y_i=1 \mid x_i\right)$ and $P\left(y_i=0 \mid x_i\right)$ can be obtained according to Eqs. (6) and (7), respectively. Subsequently, the magnitude of the values $P\left(y_i=1 \mid x_i\right)$ and $P\left(y_i=0 \mid x_i\right)$ are compared to categorize x_i into the category with the larger probability value. The regression coefficients and independent variables are expanded and expressed in the form of vectors, denoted as β, x, i.e. $\beta=\left(\alpha, \beta_1, \beta_2, \ldots, \beta_p\right)^T$, $x=\left(1, x_1, x_2, \ldots, x_n\right)$. At this point, the binomial Logistic regression model can be expressed as:

$$P(Y=1 \mid x)=\frac{\exp\left(\beta^T x\right)}{1+\exp\left(\beta^T x\right)} \tag{8}$$

$$P(Y=0 \mid x)=\frac{1}{1+\exp\left(\beta^{T} x\right)} \tag{9}$$

If the probability of an event occurring is denoted as P and the probability of the event not occurring is $1-P$, then the odds of the event occurring (the ratio of the likelihood of the event occurring to the probability of the event not occurring) is $P/(1-P)$ and the logarithmic odds of the event are:

$$\log it(P)=\log \frac{P}{1-P} \tag{10}$$

From Eqs. (8) and (9), the logarithmic odds of outputting $Y=1$ are:

$$\log \frac{P(Y=1 \mid x)}{1-P(Y=1 \mid x)}=\beta^{T} x \tag{11}$$

This indicates that the log odds of output $Y=1$ can be expressed as a linear function of input x, i.e., the log odds of output $Y=1$ is modeled by a linear function of input x, i.e., the Logistic regression model [30].

3) Parameter Estimation of Binomial Logistic Regression Model

The estimation of the parameters of the logistic model is different from the estimation of the familiar linear regression model because the two models obey different distributions. One is the normal distribution, and the other is the binomial distribution. So, we do not take the least squares method as our reference method, and in this paper, we choose the maximum likelihood estimation method to estimate the parameters.

Setting:

$$P(Y=1 \mid x)=\pi(x), P(Y=0 \mid x)=1-\pi(x) \tag{12}$$

Then its likelihood function is:

$$\prod_{i=1}^{N}\left[\pi\left(x_{i}\right)\right]^{y_{i}}\left[1-\pi\left(x_{i}\right)\right]^{1-y_{i}} \tag{13}$$

The log-likelihood function is:

$$\begin{aligned} L(P) &=\sum_{i=1}^{N}\left[y_{i} \log \pi\left(x_{i}\right)+\left(1-y_{i}\right) \log \left(1-\pi\left(x_{i}\right)\right)\right] \\ &=\sum_{i=1}^{N}\left[y_{i} \log \frac{\pi\left(x_{i}\right)}{1-\pi\left(x_{i}\right)}+\log \left(1-\pi\left(x_{i}\right)\right)\right] \\ &=\sum_{i=1}^{N}\left[y_{i}\left(\beta_{i} x_{i}\right)-\log \left(1+\exp \left(\beta x_{i}\right)\right)\right] \end{aligned} \tag{14}$$

An estimate of the parameter β is obtained by taking the maximum value for $L(P)$.

4) Advantages and Disadvantages of Logistic Models

The logistic model is developed using linear regression. Compared to a simple multiple linear regression model, the Logistic model has greater superiority. First, the assumption setting of the model is that the probability of the event is between (0, 1), and the cumulative probability release obeys the Logit distribution. Secondly, the explanatory variables of the model do not have strict requirements for continuous or categorical distributions and can abide by any distribution. Finally, the dependent variable of the model can be classified into two or more categories, which can accurately classify the credit status and solve practical problems. If there is some correlation between the factors and the outcome of an event, Logistic regression can predict and reconstruct these relationships as an equation. Logistic regression equations show the probability of an event occurring.

Logistic models have the following disadvantages:

(1) It cannot screen factors. There are more factors affecting the outcome of a real event, and accurate prediction of the outcome requires screening the most influential few factors out of many, while simple use of the model requires human screening of the factors, which is highly subjective.

(2) Easy underfitting. Without artificial screening of the factors, the data of dozens or even hundreds of alternative factors will most likely result in underfitting and failing to produce results. In fact, the results of the event may be affected by more factors. As the research is carried out, the factors will continue to increase. Logistic model, as a simple model can not adapt to the demand.

2.1.2 Lasso regression models

Lasso is a compressed variable, a penalty function designed for high-dimensional data [31]. Since Lasso is a penalized likelihood method, the method has a fairly general generality and can be applied to a wide variety of models. Lasso has the property of soft thresholding, which allows for thresholding and shrinkage of the coefficients. A single regularization parameter simultaneously controls these two properties. Due to the thresholding property, Lasso produces sparse representations, and through this property, the regularization parameter apparently determines a compromise between bias and variance. The effective coefficients may be removed if the parameter value is too large. This can cause a significant bias. If the parameter value is too low, useless coefficients will persist. This can also lead to a considerable difference.

When the independent variable is larger than the sample size, the parameter estimation will appear overfitting problem that needs a kind of complexity regularization to solve this problem, namely L_1 -penalty regularization. It is characterized by compressing the coefficients and thus achieving dimensionality reduction, which is the role of variable screening. Most models built are more concise and easy to interpret. In addition Lasso is not only applicable to continuous dependent variables but also to discrete ones, i.e., it can be used for linear models as well as binary classification models.

The basic idea of the Lasso method is to add L_1 penalties on the basis of least squares, as in equation (15):

$$\min \sum_{i=1}^{n}\left(y_i - \alpha - \sum_{j=1} c_j \hat{\beta}_j x_{ij}\right)^2 \quad s.t. \sum_{j=1}\left|\beta_j\right| \le \lambda \tag{15}$$

For everything λ, there is $\hat{\alpha} = \bar{y}$. Assuming $\bar{y} = 0$, the estimated coefficients can be expressed in the form of a penalty function as in equation (16):

$$\beta = \operatorname{argmin}\left\{\sum_{i=1}^{n}\left[\left(y_i - \sum_{j=1}\beta_j x_{ij}\right)^2 + \lambda\sum_{j=1}\left|\beta_j\right|\right]\right\} \tag{16}$$

The reconciliation parameter λ in Eq. (15) affects the penalty of the parameters in Eq. An increase in this value leads to an increase in the penalty, which finally leads to a decrease in the independent variables retained in the model, and a corresponding reduction in this value leads to a decrease in the penalty, which leads to an increase in the independent variables retained in the model. Thus, we can find that factor screening as a key aspect of the Lasso model is greatly influenced by the reconciliation parameter λ.

Lasso regression is generally operated by dividing the data into a training set and a test set and operating them separately, using the training set data to determine the correlation coefficients to build the model and then using the test set data to test it and find the model that minimizes the error between the predicted value and its real value for the prediction set data, at which point the model can be considered to be constructed as the final version.

The disadvantage of the Lasso model is that it requires a high amount of data. The Lasso model will perform better with a large amount of data. However, in practice, publicly available data tends to be less, which also makes the final amount of data available less, so the model results are less accurate.

2.2 Improved Lasso-Logistic Regression Models

2.2.1 Lasso-Logistic Regression Modeling Framework

After a detailed introduction of the Logistic regression model and Lasso's basic theory introduction, based on which we proposed an improved Lasso-Logistic regression model, Figure 1 shows the structure of the model.

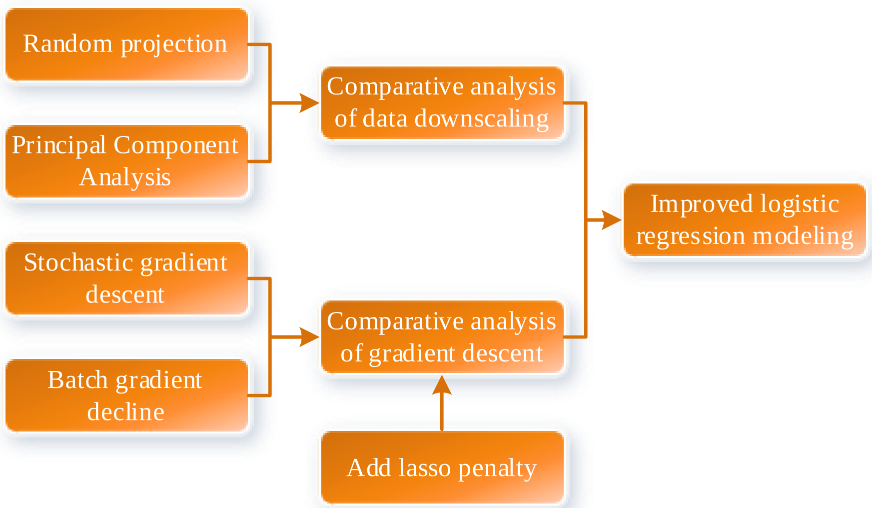


Figure 1. Lasso-logistic regression model structure

We use a combination of stochastic projection and stochastic gradient descent to achieve the optimization of the logistic regression model. However, the features of stochastic projection dimensionality reduction still have room for further optimization. There are a small number of redundant variables, so we add Lasso on the basis of stochastic gradient descent for further feature screening to improve the accuracy of the model.

The first step is to build the analysis architecture for data dimensionality reduction, and we then choose the best dimensionality reduction method by comparing and analyzing the computing speed of traditional principal component analysis and stochastic projection. Then we select logistic regression as the classifier, and in the process of solving the optimal problem by logistic regression, we respectively use stochastic gradient descent and batch gradient descent as the parameter updating method, further compare the convergence speed, and select the parameter updating method with the fastest convergence speed. Finally, lasso regularization is added on the basis of gradient descent to further reduce dimensionality and remove redundant features.

2.2.2 Lasso-Logistic Regression Model Solution

1) Stochastic Gradient Descent for Solving Logistic Regression Models

When the sample size is large, the complexity of using batch gradient descent is high. Therefore, in the classification model of this paper, the stochastic gradient descent is used to solve the logistic regression model, and the stochastic gradient descent of logistic regression takes the form of the process of partial derivation of the loss function:

$$\theta_{j+1} = \theta_j - \alpha\left(h_\theta\left(x^i\right) - y^i\right)x^i \tag{17}$$

Unlike solving the Logistic regression model, Eq. (17) is updated with parameters one at a time by randomly selecting a sample from all the samples.

2) Proximal Gradient Descent for Solving the Lasso Problem

For non-smoothing techniques, proximal gradient descent is an important method for solving non-smoothing problems and one of the solutions for solving the Lasso problem. Consider such a function problem:

$$\min_{x \in R^d} f(x) = g(x) + h(x) \tag{18}$$

Where $g(x)$ and $h(x)$ are both convex functions, but $g(x)$ is smooth and $h(x)$ is non-smooth. The most typical example of this type of problem is the least squares method based on the L1 paradigm, with the problem being:

$$\min_{x \in R^d} \frac{1}{2}\|Ax - b\|^2 + \lambda\|x\|_1 \tag{19}$$

Proximal gradient descent is the most appropriate method for solving such nonsmooth optimization problems because of their fast theoretical convergence rate and strong practical performance. The classical subgradient algorithm reaches convergence only at an error of $O(1/\sqrt{k})$ after k iteration, but the proximal gradient descent has an error of only $O(1/k)$, i.e., the proximal gradient descent for nonsmooth convex optimization achieves the same optimal convergence speed as the accelerated gradient method.

The proximal gradient method requires the computation of neighboring operators at each iteration. See equation (20):

$$y = \underset{x \in R^d}{\arg\min} \frac{L}{2}\|x-y\|^2 + h(x) \tag{20}$$

Where L is a *Lipschitz* -constant of the g -gradient, i.e., there exists a constant $L > 0$ such that:

$$\|g'(x) - g'(x)\| \le L\|x' - x\| \tag{21}$$

This is a standard assumption for differentiable optimization, and if g is second-order differentiable, we base the Taylor expansion around x on Eq. (22):

$$g(y) \approx g(x) + \langle g'(x), y - x\rangle + \frac{\mu}{2}\|y-x\|^2 \tag{22}$$

Rearranging the above equations gives:

$$g(y) \approx \frac{L}{2}\left\| x - \left(x_k - \frac{1}{L}\nabla g\left(x_k\right)\right)\right\|_2^2 + \varphi\left(x_k\right) \tag{23}$$

Where $\varphi\left(x_k\right)$ is a constant that can be neglected. It follows from Eq. (23) that the minimum value of $g(x)$ is obtained in Eq. (24):

$$x_{k+1} = x_k - \frac{1}{L}\nabla g\left(x_k\right) \tag{24}$$

Each gradient descent can be updated according to Eq. (24), and when the non-smooth penalty term L1 is added, the minimax idea above is analogized to the case where the penalty term is added:

$$x_{k+1} = \underset{x \in R^d}{\arg\min} \frac{L}{2}\left\| x - \left(x_k - \frac{1}{L}\nabla g\left(x_k\right)\right)\right\|_2^2 + \lambda\|x\|_1 \tag{25}$$

The problem of solving this objective function is transformed into that of considering both the gradient descent iterative problem and the depreciation of Lasso, such that the proximal gradient descent problem of solving Lasso can be interpreted as one can first compute Eq. (24), consider x_{k+1} as z , and then solve Eq. (26):

$$\underset{x \in R^d}{\arg\min} \frac{L}{2}\|x-z\|_2^2 + \lambda\|x\|_1 \tag{26}$$

For this equation, we can solve it by soft thresholding. Geometrically, Fig. 2 shows the image of the solution by soft thresholding.

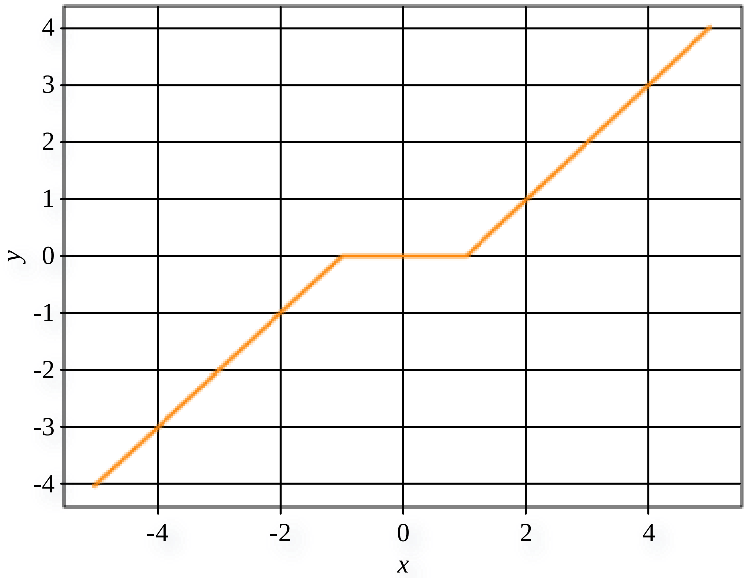


Figure 2. Soft threshold solution

As can be seen from Figure 2, for variables with little influence, soft thresholding can compress their coefficients to 0, which in turn can achieve the effect of a screening variable. In a theoretical sense, soft thresholding is generally solving the following problem:

$$\underset{x\in R^d}{\arg\min}\|X-B\|_2^2+\lambda\|X\|_1 \tag{27}$$

From a paradigm point of view, we can split this equation into:

$$\begin{aligned} F(X)&=\|X-B\|_2^2+\lambda\|X\|_1 \\ &=\left[(x_1-b_1)^2+\lambda|x_1|\right]+\cdots\left[(x_N-b_N)^2+\lambda|x_N|\right] \end{aligned} \tag{28}$$

After splitting, Eq. can be transformed into solving a small problem, and there is no interaction between different x's, see Eq. (29):

$$f(x)=(x-b)^2+\lambda|x| \tag{29}$$

The derivative of the function can be obtained:

$$\frac{df(x)}{dx}=2(x-b)+\lambda\,\mathrm{sgn}(x) \tag{30}$$

Where $\mathrm{sgn}(x)$ is the sign function, and we assume that the partial derivative is zero:

$$x=b-\frac{\lambda}{2}\mathrm{sgn}(x) \tag{31}$$

For b, there exist three different cases of solutions, and the minimum of the function takes values in this state; see equation (32):

$$\arg\min f(x)=\begin{cases} b+\lambda/2, & \mathrm{b}<-\lambda/2 \\ 0, & |b|\leq\lambda/2 \\ b-\lambda/2, & \mathrm{b}>\lambda/2 \end{cases} \tag{32}$$

We can find that $f(x)$ and $F(x)$ are only different for variable b and independent variable x. The function expressions are still approximately the same, so $F(x)$ can also be expressed by the above solution. Meanwhile, by analogy to Eq. (26), just multiplying a $L/2$ in the first term equation, we can find its closed-form solution in Eq. (33):

$$x_{k+1}^{i}=\begin{cases} z^{i}-\lambda/L, \lambda/L<z^{i} \\ 0, \left|z^{i}\right|\leq\lambda/L \\ z^{i}+\lambda/L, \lambda/L>z^{i} \end{cases} \tag{33}$$

Where ${x_{k+1}}^{i}$ and z^{i} are the i th components of x_{k+1} and z respectively. From the above, it can be obtained that by proximal gradient descent, we can minimize Lasso and other Lasso-based paradigms to reach the solution quickly.

3) Stochastic proximal gradient descent to solve Lasso-logistic model parameters

Logistic regression with the addition of the L1 regular term is effective in preventing overfitting problems and has good generalization performance for many uncorrelated feature selections. Unregularized logistic regression is an unconstrained optimization problem with a continuously differentiable objective function, and it can be effectively solved using standard optimization methods such as gradient descent.

However, the addition of Lasso will make the computation more complex and turn the problem into a convex optimization problem for solving non-smooth planes, which can be solved using the proximal gradient descent model introduced in the previous section. However, to improve the solution speed, this section focuses on combining stochastic gradient descent and proximal gradient descent methods to solve the Lasso-logistic regression model. The problem to be solved by proximal gradient descent is Eq. (34):

$$\min_{x\in R^{d}} f(x)=g(x)+h(x) \tag{34}$$

Where $g(x)$ is a smooth convex function and $h(x)$ is a relatively simple convex function, but not minimizable. In machine learning, we also often encounter optimization problems of this form. Given a column of data $(a1,b1),\ldots,(an,bn)$, we assume:

$$g_{i}(x)=\log\left(1+\exp\left(-b_{i}x^{T}a_{i}\right)\right), h(x)=\lambda\|x\|_{1} \tag{35}$$

To solve this problem, our conventional approach is to use proximal gradient descent, which can be described as equation (36):

$$x_{k+1}=prox_{\eta_{k}h}\left(x_{k}-\eta_{k}\nabla g\left(x_{k}\right)\right) \tag{36}$$

Where $prox$ is the approximation operator, and the solution of the approximation operator generally needs to be transformed into Eq. (37):

$$prox_{\eta h}(y) = \arg\min_{x\in R^d}\left\{\frac{1}{2}\|x-y\|^2 + \eta\|x\|_1\right\} \tag{37}$$

The stochastic transformation of proximal gradient descent is stochastic proximal gradient descent (SPGD), in which, at each iteration $k = 1, 2, \ldots$, we randomly select i_k from $\{1, 2, \ldots, n\}$ to update the parameters in the form of Eq. (38):

$$x_{k+1} = prox_{\eta_k h}\left(x_k - \eta_k \nabla g_{i_k}\left(x_k\right)\right) \tag{38}$$

The advantage of stochastic proximal gradient descent over proximal gradient descent is that at each iteration, stochastic forced gradient descent only needs to compute a single gradient $\nabla g_{i_i}\left(x_k\right)$, in contrast to proximal gradient descent, which uses all the samples N each time, and thus the computational cost of stochastic proximal gradient descent is $1/N$ of the cost of proximal gradient descent.

3 Result and Discussion

3.1 Selection and screening of financial risk Assessment model variables

3.1.1 Descriptive statistical analysis

In this paper, we select the public data of annual financial audit reports of 40 enterprises of different sizes in China from 2021-2023. All the data is from Wind Financial Terminal. The ratio of 8:2 is used to construct the training set and test set.

The financial status of a company mainly depends on its profitability, solvency, operating ability, and development ability. When the company's financial condition is good, the profitability level is good, the debt service level is high, and it is in a state of benign development. If the company's financial situation is not good, it will affect its profitability level and solvency accordingly. Therefore, this paper decides to select representative indicators from the four aspects of profitability, solvency, operating ability, and growth ability to reflect the company's financial situation, mainly including the following indicators:

1) Profitability: Profitability indicators reflect the ability of the enterprise to obtain profits in a certain period, mainly including gross operating profit margin, return on total assets, return on net assets, retained earnings to total assets ratio, and other indicators.

2) Solvency: solvency is an important symbol to ensure the ability of enterprises to repay debts on maturity, whether the enterprise can pay cash and debt repayment is the key to the survival and healthy development of the enterprise. The financial risk will increase when the enterprise's solvency is weak. Mainly includes the gearing ratio, interest coverage multiple, cash flow to current liabilities ratio, working capital to total assets ratio, and other indicators.

3) Operating capacity: operating capacity is mainly reflecting the turnover rate and turnover speed of the enterprise for assets. Primarily includes accounts receivable turnover, working capital turnover, total asset turnover, net sales rate, and other indicators.

4) Development ability: development ability reflects the future development potential and prospects of the enterprise. It mainly includes growth rate indicators such as total assets growth rate, net profit growth rate, net assets growth rate, and so on.

After selecting the above indicators, descriptive statistics on the distribution of sample data were analyzed. Table 1 shows the results of the descriptive statistical analysis of the distribution of sample data.

Table 1. Sample index data distribution statistics table

Index	Mean	SD	Minimum	Median	Maximum
Gross operating margin(%)	9.68	128.53	-43.71	8.94	67.85
Return on total assets(%)	3.11	5.68	-42.34	2.63	25.79
Net assets income rate(%)	7.23	19.16	-12.93	6.88	22.67
The ratio of retained earnings to total assets	0.39	12.33	-3.78	0.26	2.33
Ratio of liabilities(%)	6.48	13.72	-1.46	4.49	15.68
Interest coverage ratio	13.25	44.59	2.46	11.37	19.74
The ratio of cash flow to current liabilities	1.37	5.93	0.11	1.29	3.61
The ratio of working capital to total assets	0.94	3.22	0.13	0.79	1.28
Turnover of account receivable(%)	12.63	27.83	0.39	9.75	18.91
Turnover of working capital(%)	10.58	33.54	0.67	10.11	17.63
Circulating rate of total assets(%)	1.75	7.61	0.23	1.59	4.79
Net cash on sales(%)	6.32	13.95	1.06	5.83	10.21
Growth rate of total assets(%)	2.91	21.78	-44.57	2.62	23.56
Net profit growth rate(%)	1.67	7.24	-31.46	0.73	4.09
Growth rate of net assets(%)	0.82	17.65	-56.91	0.77	47.86

3.1.2 Factor analysis

The objective of this paper is to perform factor analysis on 15 selected financial indicators, extract the public factors, and build a Lasso-Logistic regression model using the public factors as variables. Before factor analysis is performed on the sample data, the KMO test [32] is first performed on the sample data to detect whether the selected financial indicators are suitable for factor analysis. The KMO test is used to compare the relative magnitude of the simple correlation coefficients and partial correlation coefficients among the observed variables, and its value varies from 0 to 1. When the sum of squares of the simple correlation coefficients among all the variables is much larger than the sum of squares of the partial correlation coefficients, The closer the KMO value is to 1, the stronger the correlation between the variables and the more suitable the original variables are for factor analysis. When the sum of the squares of the simple correlation coefficients of all variables is close to 0, the KMO value is close to 0. As the KMO value approaches 0, the correlation between the variables becomes weaker, and the original variables become less suitable for factor analysis. Usually, the relationship between KMO value and suitability for factor analysis is judged according to the following criteria: KMO value above 0.9 indicates very good effect, between 0.8 and 0.9 indicates good effect, between 0.7 and 0.8 indicates average effect, between 0.6 and 0.7 indicates poor effect, between 0.5 and 0.6 indicates very poor impact, and below 0.5 is not suitable for factor analysis. The test of the variable data of this paper shows that the KMO statistic of this paper is 0.874, and the significance of the test is 0.000. Judging from the above criteria, the sample data of this paper is more suitable for factor analysis, and it is possible to achieve better results.

Factor analysis was conducted on the 15 financial indicators selected above, and Table 2 shows the eigenvalues, contribution rate, and cumulative contribution rate of each factor. Table 2 shows that the eigenvalues of the first 4 public factors are greater than 1. The cumulative contribution rate reaches 84.6654%, which indicates that if the first 4 public factors are selected, they can reflect 84.6654% of the information contained in the original 15 financial indicators, and if the first 5 public factors are selected, the cumulative contribution rate is 87.2305%. By adding one factor, the contribution rate it can provide is only 2.5651%. Therefore, considering the overall efficiency, the first four public factors are selected in this paper.

Table 2. Characteristic value and contribution rate of financial index

Serial number	The sum of squares of unrotated factor loads		
	Eigenvalue	Contribution(%)	Cumulative contribution(%)
1	4.3497	34.7936	34.7936
2	2.7169	23.6352	58.4288
3	1.5408	15.9827	74.4115
4	1.2135	10.2539	84.6654
5	0.8943	2.5651	87.2305

These 4 public factors are used to replace the original 15 financial indicators, and these 4 public factors are denoted as F1, F2, F3, and F4. In order to provide a reasonable interpretation of the 4 public factors, it is necessary to obtain the correlation coefficients between these 15 financial indicators and the 4 public factors. In this paper, an orthogonal rotated variance maximization method was used for the conversion, and Table 3 shows the factor loading matrix obtained.

From the factor loading matrix, it can be found that:

1) Public factor F1 has a large loading on four indicators, such as gross operating margin, return on total assets, return on net assets, and retained earnings to total assets ratio, with loading coefficients of 0.743, 0.641, 0.869, and -0.617, respectively, which can be interpreted by these four indicators, and public factor F1 reflects information on the profitability of the enterprise.

2) Public factor F2 has a large load on four indicators, including gearing ratio, interest coverage multiple, cash flow to current liabilities ratio, and working capital to total assets ratio, with load factors of 0.854, 0.899, 0.791, and 0.872, respectively, which these four indicators can interpret, and public factor F2 reflects information on the solvency of the enterprise.

3) The public factor F3 has a large load on six indicators, including accounts receivable turnover rate, working capital turnover rate, total asset turnover rate, total asset growth rate, net profit growth rate, and net asset growth rate, with load coefficients of -0.853, 0.792, 0.619, 0.684, -0.547 and 0.611 respectively, which can be explained by these six indicators, and the public factor F3 reflects the information on the operation and growth ability of the enterprise.

4) The public factor F4 has a large loading on the indicator of the net present rate of sales, with a loading coefficient of 0.842, which can be interpreted from this indicator, and the public factor F4 reflects the information on the cash flow of the enterprise.

Table 3. Factor load matrix

Index	Loading coefficient			
	F1	F2	F3	F4
Gross operating margin	0.743	0.279	-0.061	0.145
Return on total assets	0.641	0.527	0.054	0.068
Net assets income rate	0.869	-0.231	0.202	-0.365
The ratio of retained earnings to total assets	-0.617	-0.037	0.293	0.214
Ratio of liabilities	0.236	0.854	-0.047	-0.239
Interest coverage ratio	0.042	0.899	-0.076	0.187
The ratio of cash flow to current liabilities	0.201	0.791	0.098	-0.034
The ratio of working capital to total assets	0.403	0.872	0.259	-0.380
Turnover of account receivable	0.017	0.132	-0.853	0.307
Turnover of working capital	0.274	-0.031	0.792	-0.026
Circulating rate of total assets	0.129	0.139	0.619	-0.281
Net cash on sales	0.356	0.047	0.381	0.842
Growth rate of total assets	0.251	0.169	0.684	0.130
Net profit growth rate	-0.087	-0.218	-0.547	0.022
Growth rate of net assets	-0.180	0.376	0.611	-0.178

3.2 Results of the Company’s financial risk assessment and comparative analysis

3.2.1 Model Output Results

Taking the public factors F1, F2, F3, and F4 extracted in part 3.1 of this paper, which reflect the information of enterprises in various aspects, as variables, and the public data of annual financial audit reports of 40 enterprises of different sizes within China in 2021-2023 as raw data, the analysis is carried out using the Lasso-Logistic regression model designed in this paper, and Table 4 is the output of the Lasso-Logistic regression model output results.

In the output results of the Lasso-Logistic regression model, public factors F1, F2, F3, and F4 are significant at the level of $p=0.05$, which indicates that each public factor in the model has a strong explanatory ability for whether or not a financial crisis will occur. The coefficients of public factors F1, F2, F3, and F4 are -2.9513, -1.8347, -1.9659, and -2.2714, respectively, and the coefficients of the four public factors are all negative, which indicates that the stronger the profitability, solvency, operation and growth of an enterprise, and the larger the cash flow, the better the enterprise’s operating condition is, and thus the lower the possibility of financial risk. Therefore, the outputs of the Lasso-Logistic regression model in this paper can help enterprises assess their financial risks more effectively and take better-targeted countermeasures.

Table 4. Lasso-Logistic regression model output results

Factor	B	S.E	Wald	Df	Sig	Exp(B)
F1	-2.9513	0.4932	21.356	1	0.0000	2.3714
F2	-1.8347	0.4675	7.839	1	0.0000	0.0361
F3	-1.9659	0.5861	18.524	1	0.0001	29.5387
F4	-2.2714	0.7928	11.732	1	0.0003	0.0219
Constant	-3.0671	0.3879	2.647	1	0.2784	0.1983

3.2.2 Comparative analysis of models

In order to prove the advantages of the Lasso-Logistic combination model designed in this paper, this paper compares the Lasso-Logistic combination model with the commonly used single models such as Lasso, Logistic regression, and SVM in terms of classification effects, and Table 5 shows the comparison results of the classification effects of each model.

First, the accuracy of the combined Lasso-Logistic model is 89.46%, which is 5.39 percentage points higher than the Lasso model and 14.55 percentage points higher than the Logistic model. Secondly, in the first category misclassification rate, the Lasso-Logistic combination model has the good classification performance of the Lasso model due to the integration of the advantages of the Lasso model, so the Lasso-Logistic combination model has the same first category misclassification rate as the Lasso model, i.e., 6.21%. Finally, taken together, the Lasso-Logistic combined model is overall better than the two single models, Lasso and Logistic, in terms of F1 scores, RS2 scores, and AUC values, with the Lasso-Logistic combined model having an AUC value of 89.03%, which is 2.21 percentage points higher than the Lasso model, and 14.03 percentage points higher than the Logistic model by 14.07 percentage points, which can indicate that the combination model has the best classification performance.

In summary, through the comparative empirical analysis of Lasso, Logistic, SVM, and other single classification models and the Lasso-Logistic combination model constructed in this paper on corporate financial data, it can be seen that the combination model is better than the other single models in classification performance, and due to the combination of the advantages of the single regression models such as Lasso, Logistic and other single regression models, the combination model can reasonably explain the output results of the model, and the explainable ability has been significantly improved.

Table 5. Classification evaluation results of different models

Model	ACC	Err_1	F_1 score	RS_2 score	AUC
SVM	0.8043	0.2359	0.7938	0.7452	0.8195
Lasso	0.8407	0.0621	0.8310	0.7956	0.8682
Logistic	0.7491	0.2283	0.7361	0.6984	0.7496
Lasso-Logistic	0.8946	0.0621	0.8674	0.8129	0.8903

4 Corporate Governance and Internal Audit Construction Combined with Logistic Modeling

The Lasso-Logistic combination model in this paper can predict the financial risk of the company well, and based on the analysis results, this paper proposes that internal auditing under the demand of corporate governance provides a new path of improvement.

4.1 Emphasis on enhancing the independence of internal audit

To better meet the needs of corporate governance, attention should be paid to enhancing the independence of the company's internal audit and establishing a more independent, objective, and impartial internal audit mechanism. Clarifying the regulatory framework for internal auditing is a first step for the company to ensure that the audit team exercises its independent functions in accordance with the regulations. Establish a regulatory framework for independence, clarify the relationship between the audit team and other departments, and prevent potential conflicts of interest. The second step is to strengthen the professional conduct and ethics of internal audit team members through training and education so that they better understand and comply with the principles of independence. It is also necessary to establish a clear code of ethics to guide the audit team to always adhere to an independent, objective, and fair audit attitude. The third step is that the company should provide the audit team with sufficient resources and support to ensure that the internal audit team has the right to independently formulate the audit plan and report and clearly express independent opinions in the audit report. As a fourth step, the company should conduct regular assessments of the level of independence, which can be done by means of internal and external independence reviews and evaluations by independent third-party organizations, in order to effectively maintain and enhance the independence of the internal audit team.

4.2 Improved internal audit risk management capacity

Effective risk management capability is conducive to ensuring that the audit team assesses the various risks faced by the company in a more comprehensive and in-depth manner and provides the company with timely and accurate risk information, thereby improving the company's operational soundness and ability to cope with uncertainty. The improvement path is as follows: first, strengthening the risk management of internal audit requires the establishment of a perfect risk identification mechanism that covers all aspects of the company's risks. The audit team needs to ensure a comprehensive understanding of potential risks that the company may face by conducting regular risk studies and collecting and analyzing internal and external information. Secondly, the audit team should establish clear risk assessment criteria to ensure consistency and comparability of assessment, including clear criteria for probability, degree of impact, and prioritization of risks, so as to more accurately identify and assess risks. At the same time, it is necessary to utilize modern technological means, such as data analysis and artificial intelligence, to improve the audit team's ability to process and analyze large amounts of data, which is conducive to identifying potential risks more quickly and accurately. Again, to strengthen the risk management of internal audit, it is necessary to establish a risk response mechanism, clarify the measures and responsible persons for all kinds of risks, and ensure that all types of risk response measures are implemented in a timely manner. Finally, risk management is a dynamic process. The audit team should continuously monitor and update the risk assessment. Through timely adjustments of risk assessment standards and feedback on the actual response effect, we can ensure the continued adaptability and effectiveness of the risk management system.

4.3 Regular development of corporate internal audit reports

To meet the company's governance requirements, it is crucial to establish a mechanism for regularly developing the company's internal audit report. Internal audit reports not only provide the company's senior management with an assessment of the company's operations and internal controls but also serve as an important means of reporting the status of the company's internal audit to the board of directors and other stakeholders. The audit team must first clarify the scope and content of the audit report by consulting with the company's senior management, the board of directors, and other relevant parties. The audit report is required to evaluate key business processes and internal controls, identify issues and recommendations, and follow up on previous audit recommendations. Establishing a clear timetable for the regular development of the company's internal audit report is necessary. Usually, internal audit reports can be prepared on a quarterly or semi-annual basis to ensure that audit information is provided to relevant parties in a timely and regular manner. To ensure consistency and readability of internal audit reports, it is recommended that a standardized report format be developed. The format should be clear and concise, including sections such as a summary, key findings, recommendations, and action plans so that readers can quickly understand the audit findings and recommended measures. Transparency and truthfulness should be the focus of the report, and conclusions and recommendations should be faithfully documented with reasonable explanations and contextualization.

5 Conclusion

This paper combines the basic principles of Lasso and Logistic models to construct a Lasso-Logistic combined regression model and uses this model to regress information on enterprises and enterprise financial risk.

1) Based on 15 financial indicators, this paper extracts 4 public factors, which reflect the information on enterprise profitability, information on enterprise solvency, information on enterprise operation and growth ability, and information on enterprise cash flow.

2) In the output of the Lasso-Logistic regression model, public factors F1, F2, F3, and F4 are significant at the level of p=0.05. The coefficients of public factors F1, F2, F3, and F4 are -2.9513, -1.8347, -1.9659, and -2.2714, respectively. The coefficients of the four public factors are negative, which indicates that corporate The coefficients of the four common factors are all negative, which suggests that there is a significant negative correlation between profitability, solvency, operation, growth ability, cash flow, and financial risk of enterprises.

3) The classification accuracy of the combined Lasso-Logistic model is 89.46%, which is 5.39 and 14.55 percentage points higher than that of the Lasso and Logistic models, respectively. The combined Lasso-Logistic model is the most effective model for classification overall, with F1 scores, RS2 scores, and AUC values compared to the two single models, Lasso and Logistic.

References

[1] Bi, G. (2021). Research on financial stochastic dynamic model of energy market based on mcmc simulation. International journal of global energy issues(5/6), 43.

[2] Gao, X., Zhang, P., Huang, G., Jiang, H., & Zhang, Z. (2020). Financial information prediction and information sharing supervision based on trend assessment and neural network. Soft Computing, 24(2).

[3] Bum-Jin, & Park. (2019). Audit committees and managerial influence on audit quality: 'voluntary' versus 'mandatory' approach to corporate governance. Australian Accounting Review.

[4] Bunget, O. C., Mate, D., Dumitrescu, A. C., Bogdan, O., & Burc, V. (2020). The link between board structure, audit, and performance for corporate sustainability. Sustainability, 12.

[5] Elgammal, W., El-Kassar, A. N., & Messarra, L. C. (2018). Corporate ethics, governance and social responsibility in mena countries. Management Decision, 56(1), 273-291.

[6] Dobija, D., & Pulawska, K. (2023). Linking diversity on audit committees and financial reporting quality in poland. European journal of international management.

[7] Hassanein, A., & Kokel, A. (2022). Corporate cash hoarding and corporate governance mechanisms: evidence from borsa istanbul. Asia-Pacific Journal of Accounting & Economics, 29.

[8] Spooner, M. (2020). Technologies of governance in context: four global windows into neoliberalism and audit culture in higher education:. Qualitative Inquiry, 26(7), 743-747.

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

[10] Roussy, Melanie, Rodrigue, & Michelle. (2018). Internal audit: is the 'third line of defense' effective as a form of governance? an exploratory study of the impression management techniques chief audit executives use in their annual accountability to the audit committee. Journal of Business Ethics.

[11] John, K., Li, Y., & Pang, J. (2017). Does corporate governance matter more for high financial slack firms?. Management Science, 63.

[12] Pham, A. V., Pham, M. H., & Truong, C. (2021). Ceo cultural heritage and the pricing of audit services. Journal of Business Finance & Accounting.

[13] Oncioiu, I., Petrescu, A. G., Florentina-Raluca Bîlcan, Petrescu, M., Melinda Timea Fülöp, & Topor, D. I. (2020). The influence of corporate governance systems on a company's market value. Sustainability, 12.

[14] Gad, J. (2020). Voluntary disclosures on control system over financial reporting and corporate governance mechanisms: evidence from poland. Journal for East European Management Studies, 25(4), 698-729.

[15] Mubako, G. (2019). Internal audit outsourcing: a literature synthesis and future directions. Australian Accounting Review, 29.

[16] Bummgin Park. (2019). Audit committees and managerial influence on audit quality: 'voluntary' versus 'mandatory' approach to corporate governance. Australian Accounting Review, 29.

[17] Omer, T. C., shelley@unl.edu, & Tice, F. M. (2020). Do director networks matter for financial reporting quality? evidence from audit committee connectedness and restatements. Management Science, 66.

[18] Giudice, A. D., & Rigamonti, S. (2020). Does audit improve the quality of esg scores? evidence from corporate misconduct. Sustainability, 12.

[19] Marnet, O., Fairchild, R. J., & Gwilliam, D. (2019). Audit within the corporate governance paradigm: a cornerstone built on shifting sand? published in the british journal of management in january 2019. British Journal of Management.

[20] Basu, S., & Lee, E. I. (2021). Antecedents of and outcomes after finance committee use. Journal of Business Finance & Accounting.

[21] Wu, J. Y., Ronán Feehily, & Lord, B. R. (2021). The corporate governance role of audit committees: through the lenses of new zealand institutional investors. Australian Accounting Review(9).

[22] Christopher, J., Leung, P., & Leong, S. (2018). Can employees be used to overcome independent audit limitations?. Australian Accounting Review.

[23] Wei, Z., & Yang, W. (2017). Research on building of cloud accounting financial sharing management model based on analysis in the era of big data. Revista de la Facultad de Ingenieria, 32(14), 34-40.

[24] Gao, B., & Balyan, V. (2022). Construction of a financial default risk prediction model based on the lightgbm algorithm. Journal of Intelligent Systems, 31(1), 767-779.

[25] Liao, S., & Liu, Z. (2022). Enterprise financial influencing factors and early warning based on decision tree model. Scientific Programming.

[26] Yang, Y., Cao, Y., & Yang, C. (2020). Innovation method for centralised management and control mode of enterprise financial audit under the background of industry 4.0. International Journal of Technology Management, 84(3/4), 197.

[27] Chen, L., Dejana, R., & Nassar, T. (2021). Sharing enterprise cloud securely at ibm. IT Professional, 23(1), 67-71.

[28] Valipour, H., & Vahed, M. S. (2017). Reviewing conservative relationship between accounting and corporate governance mechanisms. International Journal of Biometrics, 12, 172.

[29] Loseva, O. V., Tazikhina, T. V., & Fedotova, M. A. (2020). Transformation of corporate governance and property relations in a digital society. Management Science.

[30] Gilver Rosero Chasoy,Omar Vladimir Pabon,Leidy Johana Valencia Hernández & Liliana Serna Cock.(2024).Kinetic analysis of lactic acid fermentation by using an extended logistic model.Process Biochemistry1-5.

[31] Jeongseok Bang & Doojin Ryu.(2024).ESG factors and the cross-section of expected stock returns: A LASSO-based approach.Finance Research Letters105482-.

[32] Xiangbo Liu,Jinwei Long,Xinwei Du,Jiasi Cai & Yanhong Wei.(2024).Systematical factor analysis of dendrite morphology and solute distribution in entire welding molten pool of Al[sbnd]Cu alloy: An MPF-LBM study.Journal of Manufacturing Processes74-91.

About the Author

Hongli Wu (1985.12—), female, Meizhou Guangdong, doctoral student of José Rizal University, Associate professor at Guangdong Innovative Technical College, main research direction: financial accounting.