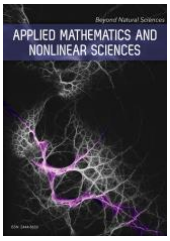
sciendo Applied Mathematics and Nonlinear Sciences, 9(1) (2024) 1-19



Applied Mathematics and Nonlinear Sciences

https://www.sciendo.com

# Optimising the design of financial data processing models in accounting information systems based on artificial intelligence techniques

Yanhua Song[1,†]

1. Tangshan Polytechnic University, Tangshan, Hebei, 063299, China.

**Submission Info**

Communicated by Z. Sabir
Received June 13, 2024
Accepted October 9, 2024
Available online November 27, 2024

## Abstract

Financial assessment and early warning analysis can help enterprises find potential financial problems earlier, make timely plans and take necessary measures to avoid risks. This paper uses a Bagging algorithm to integrate Random Forest, Support Vector Machine, and Plain Bayesian method to achieve the processing and classification of enterprise financial imbalance data. The entropy weight method is used to select and empower financial indicators to construct an accounting and financial data assessment model based on artificial intelligence technology. The model is applied to a consumer electronics enterprise, Company W, to analyze its financial situation and operating level. It is found that the composite score from 2019 to 2022 is 60.29, 70.80, 73.11, and 76.52, and the operating condition gradually improves from 2019. Debt service capacity, profitability, operating capacity, and growth capacity also show a positive trend. This is consistent with the actual development of Company W. Accordingly. It is recommended that Company W while maintaining its R&D advantages, focus more on the long-term operating ability of the enterprise, compress the operating cycle, reduce the risk of repayment and inventory pressure, and continue to enhance the competitiveness of the enterprise. This paper presents new ideas and methods for the innovation of enterprise management and the intelligence of accounting information systems.

**Keywords:** Unbalanced data; Random forest; Support vector machine; Plain bayes; Financial data processing.
**AMS 2010 codes:** 68T05

†Corresponding author.
Email address: hbts2495@126.com

ISSN 2444-8656
https://doi.org/10.2478/amns-2024-3603

## 1 Introduction

With the reform of China's economic system and the continuous development of the market economy, enterprises need to make decisions on their affairs, which is bound to analyse the enterprise's accounting information to find out the regularity factors to provide a basis for its prediction and decision-making, and to guide, supervise and control the various departments of the enterprise management [1-2]. Therefore, the accounting information system in the centre of the enterprise management information system only has the accounting function is far from enough, but also has to participate in the management and decision-making functions. It should support business management, including financial planning, analysis and control management functions, support for business decision-making, financing decisions, investment decisions, capital use and management decisions, etc. [3-5]. Therefore, to achieve the informationisation of accounting, accounting management and accounting decision-making in order to meet the requirements for the comprehensive realisation of accounting informationisation. Meanwhile, in the enterprise, accounting is the information core of enterprise operation and management, which has a close connection with other parts of the enterprise. The accounting accounting information system is almost a closed system for financial departments only, and other information systems are independent of each other, forming an information island. However, with the continuous development of network technology, it is possible to achieve seamless linkage of information from various departments of the enterprise, which is the requirement of enterprise business management informatisation and the development trend of accounting informatisation [6-9].

In summary, accounting needs to solve many complex operational, management and decision-making problems, but due to the complexity of the financial accounting function, part of which can be solved based on the historical and realistic data of accounting and mathematical model calculations, while more problems are dependent on the financial experts and accounting staff's rich practical experience, valuable wisdom and thinking in order to be able to solve them satisfactorily [10-11]. The development of artificial intelligence technology provides strong support for financial data processing in accounting information systems. Therefore, in order to establish a full-featured and effective managerial accounting and decision-making accounting information system, to adapt to the needs of the enterprise with a large increase in the amount of data and predictive decision-making, it is also necessary to study and learn from international data warehousing, online analysis and processing, data control and other technologies, to analyse and observe the data from multiple sources, and it is very necessary to introduce advanced technologies and methods such as artificial intelligence technology and network technology into the system [12-15].

Artificial intelligence is an important field of current information technology development, which helps to improve the productivity of enterprises, and in the field of accounting, the in-depth application of artificial intelligence is completely necessary and very urgent. Literature [16] stresses the importance of data and points out that the full use of big data leads to the development of traditional accounting to accounting information technology, and at the same time, analyses the challenges faced by financial enterprises' computerized accounting in investment and financing decision-making, tax management, performance evaluation, etc., and puts forward relevant solution strategies. Literature [17] describes the definition and wide application of big data technology, artificial intelligence and informatisation theory, then examines the influencing factors of the development of accounting informatisation in the form of a questionnaire, and designs an accounting informatisation construction plan combining big data and artificial intelligence, and compares it with the traditional accounting informatisation construction plan in order to verify its applicability. Literature [18] highlights the important position of accounting informatisation in enterprise financial management, and through research and analysis, it is found that accounting informatisation can not only improve the efficiency and refinement level of enterprise financial management, but also

promote the overall development of enterprise financial management, in addition, in order to improve the overall level of financial management, it also proposes the integration strategy of the accounting information system with the supply chain management, customer relationship management and other systems. Literature [19], after literature review and empirical research, compared and assessed the advantages of traditional methods and cloud computing-based models in implementing accounting informatisation construction, and further analysed the factors and decision-making criteria affecting SMEs' choice of accounting informatisation construction models.

With the advent of the information technology era, financial data processing plays a crucial role in corporate decision-making and risk management. Literature [20] proposed an adaptive real-time architecture for financial data integration based on hybrid financial ontology, elastic distributed dataset and real-time discrete streaming, which can solve the problems of data processing delays, functional misinterpretation, and metadata heterogeneity in real-time financial data integration, and this architecture helps to improve the quality and usability of financial reports in a short period of time. Literature [21] states the trend of traditional financial accounting to management accounting, based on the existing information system, based on artificial intelligence. It proposes the financial management optimisation design method and accounts receivable management optimisation framework, constructs the financial distress early warning model based on the neural network technology, and verifies the validity of the above methods, frameworks and models through experiments, which can create greater value for the enterprise. Literature [22] proposed a financial early warning model for listed companies based on optimised BP neural network, took the financial data of listed companies from 2017 to 2020 as an example, verified the feasibility and practicability of the proposed financial early warning model, which can effectively carry out the financial data processing, discover the financial problems in a timely manner and take relevant and effective measures to cope with the crisis.

In this paper, in order to construct an accounting and financial data evaluation model based on artificial intelligence technology, an unbalanced data processing algorithm integrating Random Forest, Support Vector Machine, and Plain Bayesian method is designed to automate the processing of corporate financial data. A hybrid sampling method is used to oversample and then undersample the dataset, and then the Bagging algorithm is used to achieve integrated machine learning. Compare and validate the predictive effect of the models in this paper using seven machine learning models, namely logistic regression, K-nearest neighbor, decision tree, neural network, random forest, CatBoost, LightGBM, and XGBoost, as baselines. The models are used to evaluate and analyze publicly available financial data of Company W from 2019 to 2022.

## 2 Artificial Intelligence Based Accounting Financial Data Evaluation Model

### 2.1 Financial data and potential operational crisis

The emergence of an enterprise financial crisis is not sudden. Often, the enterprise financial data appears as a risk first. Risk in the case of not getting good management can develop into a financial crisis. In the business activities of enterprises involved in different aspects of large and small, and they are interlinked, once a link of the financial data problems and not get a good repair adjustment, it will eventually lead to other items of financial data also gradually bad, so it can be said that the emergence of financial crisis is the potential problems of financial data from the quantitative change from the accumulation of qualitative changes occurring in the process.

For an enterprise, from the beginning of its founding to the beginning of the scale, in the process of continuous growth and development, to experience countless challenges and bumps, the emergence

of the financial crisis undoubtedly has a fatal blow to the enterprise. For different enterprises, the emergence of a financial crisis may put them at risk of bankruptcy. Although the emergence of the financial crisis is not without a solution, the price to pay for an enterprise may be unable to afford.

Enterprise financial data from good to crisis, according to the severity of the problem is a stage, enterprises in different stages with different performance characteristics, which also means that the financial crisis in the generation before there is enough time and signs to be found, then also can be combined with the stage of its targeted policy adjustments to avoid unnecessary losses. Generally speaking, the financial crisis of enterprises can be divided into four stages, namely the latent stage, the development stage, the deterioration stage, and the final stage.

In this paper, we will construct an accounting and financial data evaluation model based on artificial intelligence technology to monitor and evaluate the financial data in the enterprise operation so as to facilitate the enterprise to instantly discover the financial data problems and adjust the operation activities in time. Accordingly, it can reduce the risk of business management, avoid crises, and help enterprises achieve smooth development.

## 2.2 Algorithm for unbalanced data classification

Data imbalance refers to the fact that the number of samples in each class of the dataset varies considerably, whereas in real-world problems, a small number of class samples is usually important for classification, and too few samples in a small number of classes may make the classifier less effective.

### 2.2.1 Evaluation indicators for disaggregated imbalance data

In order to adapt to the characteristics of unbalanced data, it is necessary to select suitable evaluation metrics that contain more information in order to assess unbalanced data appropriately [23]. There are two evaluation metrics selected in this paper, which are the F1 score and AUC.

The Positive category represents the minority category, while the Negative category represents the majority category. The true values are classified as True and False, and the predicted values are classified as Positive and Negative. According to the different combinations of true values and predicted values, there are four types: TP, FP, FN and TN. If the true values are consistent with the predicted values, it means that the classification is correct, and the number of samples is TP+TN; if the true values are not consistent with the predicted values, it means that the classification is incorrect, and the number of the samples is FP+FN. The total number of samples is the correct prediction plus the incorrect prediction. The total number of samples is the correctly predicted plus the incorrectly predicted, totalling TP+TN+FP+FN.

The formula for calculating the accuracy rate is:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{1}$$

Denotes the proportion of samples that are correctly predicted among all samples. Since the accuracy rate is not suitable as a classification evaluation index for unbalanced data, the concepts of precision rate and recall (Recall) are introduced.

The formula for calculating the precision rate is as follows:

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

Indicates the proportion of true positive samples out of all positively classified samples.

Recall represents the proportion of positive samples that are accurately classified out of all positive samples:

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

The F1 score is a reconciled average of the two obtained based on Precision and Recall, the larger the value of the F1 score, the more effective the classifier is on a small number of classes:

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \tag{4}$$

AUC is the area of the zigzag shape enclosed by the ROC curve and the horizontal axis. And the ROC curve is based on the prediction results of the learner to rank the samples, in this order one by one, the positive class samples are predicted, and each time, the true class rate and the false positive class rate are calculated, the formula is as follows:

$$FPR = \frac{FP}{FP + TN} \tag{5}$$

$$TPR = \frac{TP}{TP + FN} \tag{6}$$

Where the horizontal and vertical axes of the ROC curve are FPR and TPR, respectively, if the position of the curve is skewed towards the upper left corner, it indicates a higher accuracy of training.

### 2.2.2 Resampling methods

The resampling method is an indispensable tool in modern statistics, which mainly involves the oversampling of a few samples and undersampling of the majority of samples or a combination of the two so as to make the unbalanced samples balanced.

1) Oversampling method

   The oversampling method is also called upsampling method, the main idea is to generate new samples from a few samples in order to achieve the majority of samples and minority samples balanced results. Currently, widely used oversampling methods mainly include SMOTE and its improved algorithms.

   The SMOTE algorithm is an improvement of random oversampling, and the basic idea is to artificially create a new minority class sample between two neighbouring minority class samples, so as to achieve a relative balance in the number of samples of the two classes. The specific process is to first analyse and calculate the K-nearest neighbours of each minority sample $X_i$, and then use the linear interpolation formula (7) to randomly select sample $X_j$

from the K-nearest neighbours, and the sample is located between sample $X_i$ and sample $X_j$:

$$X_{new} = X_i + rand(0,1) * | X_i - X_j |, i = 1,2,...,n \tag{7}$$

Where $rand(0,1)$ represents a random number between 0 and 1.

2) Under-sampling method

Under-sampling methods, also called under-sampling methods, the main idea is to remove samples belonging to the majority class from the dataset in order to better balance the class distribution. Random undersampling is usually relatively simple, i.e., randomly selecting samples from the majority class for deletion. However, its limitation is that the randomly deleted samples may be located at decision boundaries, which are usually important.

Tomek Links denotes the pair of samples that are the closest distance between different categories, i.e. these two samples are nearest neighbours to each other and belong to different categories. The main idea is to assume that $s_i$ and $s_j$ belong to different categories, and $d(s_i, s_j)$ represents the distance between the two sample points. If there is no third sample point $s_k$ such that $d(s_k, s_i) < d(s_i, s_j)$ or $d(s_k, s_j) < d(s_i, s_j)$ holds.

Then $(s_i, s_j)$ is said to be a Tomek Link pair and $s_i$ and $s_j$ are close neighbours of each other.

Thus, if two samples form a Tomek Link pair, either one of them is noise or both samples are near the boundary. By eliminating most of the class samples in a Tomek Link pair, the overlapping samples can be cleaned, so that the samples that are nearest neighbours to each other all belong to the same class, which leads to a better classification.

3) Hybrid sampling algorithm

The hybrid sampling algorithm is designed to address the shortcomings of oversampling and undersampling methods. The disadvantage of the oversampling algorithm is that the minority class samples generated are easily overlapped with the surrounding majority class samples, which makes it difficult to classify. However, undersampling can remove the overlapping samples. So it is possible to combine the two by oversampling the dataset and then undersampling it.

### 2.2.3 Classification methods

Classification is an important task in machine learning, i.e., constructing a function to determine the class to which the input belongs, either as a binary classification problem or as a multiclassification problem.

Common classification methods are mainly Random Forest (RF), Support Vector Machine (SVM) and Plain Bayes (NB).

1) Random Forest

The weak classifier used in Random Forest is the CART tree. The CART decision tree, also known as the categorical regression tree, is a binary decision tree that uses the Gini coefficient to select features [24]. The Gini coefficient chosen is determined by the purity of the child nodes. The greater the purity, the better. When all the samples located at the child node have the same category, it means that the child node has the highest purity, and the Gini coefficient is also the smallest.

The formula calculates the Gini coefficient:

$$Gini(p) = 2p(1-p) \tag{8}$$

Random forests use the idea of Bagging, so-called Bagging: each base learner is trained by taking only a portion of the initial training samples and then using a simple voting method for the classification task.

2) Support Vector Machine

Support Vector Machine (SVM) is a linear classifier that classifies binary samples according to a supervised learning approach. Its decision function is to find the hyperplane that maximizes the classification interval. Support Vector Machines are classified into three categories based on the degree of linear separability of the training data, namely Linearly Separable Support Vector Machines, Linear Support Vector Machines, and Non-Linear Support Vector Machines [25]. SVMs can classify samples in a non-linear manner using kernel methods. The basic idea of the kernel function is to compute the inner product between two asked quantities after letting the vectors that are in the low dimensional space be transformed to the high dimensional rarefaction interval through linear mapping. Since this paper focuses on binary classification problems, and the training data are rarely linearly differentiable in reality, the following section focuses on the nonlinear support vector machine that introduces the kernel function.

The original optimisation problem is:

$$\begin{aligned} &\min_{w,b,\xi} \frac{1}{2}\| w\|^2 + C\sum_{i=1}^{N}\xi_i \\ &s.t. \quad y_i(w\cdot x_i + b) \geq 1-\xi_i \\ &\xi_i \geq 0, i = 1,2,...,N \end{aligned} \tag{9}$$

Here, $c > 0$ is known as the penalty parameter, where a larger value of C results in a larger penalty for classification errors. $\xi_i$ is called the slack variable and represents the degree to which the sample is unsatisfied with the constraints. To solve this convex quadratic programming problem, the original problem is generally considered to be transformed into a pairwise problem of convex optimisation. Its dual problem is:

$$\begin{aligned} &\min_{\alpha} \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j (x_i \cdot x_j) - \sum_{i=1}^{N}\alpha_i \\ &s.t. \quad \sum_{i=1}^{N}\alpha_i y_i \\ &0 \leq \alpha_i \leq C, i = 1,2,...,N \end{aligned} \tag{10}$$

Where $\alpha_i$ is the Lagrange multiplier.

Solving the problem yields the classification decision function:

$$f(x) = sign(\sum_{i=1}^{N} \alpha_i^* y_i (x_i \cdot x_j) + b^*) \tag{11}$$

In order to make the nonlinear classification problem solvable, the kernel trick is needed: i.e., a new sample space is initially obtained by mapping the original data, and the model is later trained in a higher dimensional space. The kernel function is defined as follows:

Let $\Omega$ be the input space (a subset of the Euclidean space $R^n$) and $H$ be the feature space, if there exists a mapping from $\Omega$ to $H$

$$\phi(x) : \Omega \to \mathrm{H} \tag{12}$$

Such that for all $x, z \in \Omega$, the function $K(x, z)$ satisfies the condition

$$K(x, z) = \phi(x) \cdot \phi(z) \tag{13}$$

Then $K(x, z)$ is said to be the kernel function and $\phi(x)$ is the mapping function.

Therefore, $(x_i \cdot x_j)$ in Eq. (11) can be replaced by the kernel function $K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$. At this point, the objective function of the dyadic problem becomes:

$$W(\alpha) = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j K(x_i \cdot x_j) - \sum_{i=1}^{N} \alpha_i \tag{14}$$

The classification decision function is:

$$f(x) = sign(\sum_{i=1}^{N_i} \alpha_i^* y_i \phi(x_i) \phi(x) + b^*) = sign(\sum_{i=1}^{N_i} \alpha_i^* y_i K(x_i \cdot x) + b^*) \tag{15}$$

Commonly used kernel functions include linear kernel, polynomial kernel function and Gaussian kernel function, which are denoted as follows:

$$K(x, x_i) = x \cdot x_i \tag{16}$$

$$K(x, x_i) = ((x \cdot x_i) + 1)^d \tag{17}$$

$$K(x, y) = \exp(-\gamma \| x - y \|^2) \tag{18}$$

3) Plain Bayes

The basic principle of the plain Bayesian method is Bayes' theorem, based on which the method is also combined with the conditional independence assumption [26]. The joint probability distribution $P(X, Y)$ is first calculated and then the posterior probability

distribution $P(Y \mid X)$ is obtained. Specifically, the joint probability distribution is obtained by using the training data to learn the estimates of $P(Y \mid X)$ and $P(Y)$:

$$P(X,Y) = P(Y)P(X \mid Y) \tag{19}$$

Probability estimation methods are great likelihood estimation or Bayesian estimation. The underlying assumption is conditional independence: that is, it is assumed that the features of a given training set are independent of each other:

$$\begin{aligned} P(X = x \mid Y = c_k) &= P(X^{(1)} = x^{(1)}, \ldots X^{(n)} = x^{(n)} \mid Y = c_k) \\ &= \prod_{j=1}^{n} P(X^{(j)} = x^{(j)} \mid Y = c_k) \end{aligned} \tag{20}$$

### 2.2.4 Integrated learning

Integrated learning algorithms are not a separate class of machine learning algorithms but rather a fusion of multiple machine learning models to get the final result. In this paper, we will integrate the Random Forest, Support Vector Machines, and Plain Bayesian Classification described in the previous section to finally achieve financial data evaluation.

The bagging algorithm, also known as bagging algorithm, is a parallel integration algorithm that is based on the self-help sampling process Boostrap sampling mentioned earlier: for a given dataset, $m$ the sample is randomly selected as the sampling set, and then that sample is put back into the initial dataset so that the sample is still possible to be selected in the next sampling. This continues for $M$ times. Then we can get $M$ sampling sets containing $m$ training samples, and then train the sampling sets on the base learners separately, and finally combine the results of these $M$ base learners as the final result output.

For the financial data processing problem in this paper, the voting method is used to get the category with the most votes as the final model output. The voting method is further divided into the absolute majority voting method and the relative majority voting method:

Absolute majority voting method:

$$H(x) = \begin{cases} c_j, & if \sum_{i=1}^{r} h_i^j(x) > 0.5 \sum_{k=1}^{N} \sum_{i=1}^{r} h_i^k(x) \\ reject, & otherwise \end{cases} \tag{21}$$

That is, if there is a majority of votes for a mark, the prediction is for that mark; otherwise, the prediction is rejected.

Relative majority voting method:

$$H(x) = c_{\arg\max} \sum_{i=1}^{r} h_i^j(x) \tag{22}$$

That is, the prediction is for the marker with the highest number of votes, and if more than one marker receives the highest votes at the same time, one of them is randomly selected. In summary, integrated learning can be achieved.

## 2.3 Indicator system and entropy method of weighting

### 2.3.1 Establishment of an indicator system

This paper selects the following four directions of financial data-related indicators based on the review of information and expert consultation.

1) Solvency. The company's high proportion of debt is likely to cause the occurrence of business risk. The company's solvency risk is generally manifested in the company's inability to pay cash and repay debts. The strength of solvency is related to the sustainable and healthy development of the company. The selected solvency indicators are gearing ratio A1, earned interest multiple A2, and cash current debt ratio A3.

2) Profitability. Insufficient profitability will affect the company's level of operation, not only to focus on the profitability of the company, but also focus on the management of costs, so the cost margin is used to measure the cost of the company to obtain revenue. The indicators selected for profitability are return on net assets B1, return on total assets B2, and cost margin B3.

3) Operating Capacity. If the business undertaken has a long cycle, a large period, and obstacles to payback, there is a huge risk to operating capacity. To assess operating capacity, the indicators chosen are total asset turnover rate C1, accounts receivable turnover rate C2, and inventory turnover rate C3.

4) Growth capacity. Under the fierce market competition environment, the enterprise must improve its innovation and development ability and constantly make adjustments according to the direction of market demand in order to keep pace with the market and realise the continuous improvement of market share and operating income. The growth capacity indicator chosen is operating growth rate D1, total assets growth rate D2, and capital accumulation rate D3.

### 2.3.2 Entropy weight method model

In information management methods, the entropy method is used to measure uncertain information. If the entropy value is higher, the system becomes chaotic, resulting in a lack of valid information. On the contrary, when the entropy value is smaller, it means that the system becomes more orderly and can contain more valid information.

The accuracy of early warning of enterprise business risks mainly depends on the selection and establishment of early warning indicators. This paper selects the entropy value method for data processing and calculates the entropy value and weight of each indicator to select the highest-weight early warning indicators. Because of the selection of more indicators, this paper selects the entropy value method to assign weights to each indicator. The precise calculations are as follows:

Firstly, assuming that there is $n$ evaluation item, representing different years, and $m$ evaluation indicators under each item, then $x_{ij}$ represents the $j$ th indicator in the $i$ th year $(i=1,2,...n;\ j=1,2,...m)$, and the initial data matrix is constructed $X$ based on the financial data of the initial selection of indicators.

$$X = \begin{vmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{vmatrix} \tag{23}$$

Second, standardisation. Due to the difference in the data outline of each indicator, it is not possible to make a simple comparison of the results based on the data, so it is necessary to unify the initial data to obtain a unified outline data. The standardisation method is as follows:

If it is a positive indicator with larger and larger values, it is processed according to formula (24).

$$x_{ij} = \frac{x_{ij} - x_j^{min}}{x_j^{max} - x_j^{min}} \tag{24}$$

In case of inverse indicators with smaller values, they are treated according to equation (25).

$$x_{ij} = \frac{x_j^{max} - x_{ij}}{x_j^{max} - x_j^{min}} \tag{25}$$

If it is a fitness indicator whose value is optimal in a certain interval range, it is processed according to formula (26).

$$x_{ij} = \begin{cases} \dfrac{x_{ij} - x^{\min}}{x_0 - x^{\min}}, x_{ij} < x_0 \\ \dfrac{x^{\max} - x_{ij}}{x^{\max} - x_0}, x_{ij} \geq x_0 \end{cases} \tag{26}$$

Where $X_0$ is the moderation value of the moderation indicator.

Thirdly, the data were non-negativised according to equation (27):

$$Y_{ij} = 1 + x_{ij} \tag{27}$$

Fourth, the data were normalised.

$$y_{ij} = \frac{Y_{ij}}{\sum_{i=1}^{n} Y_{ij}} \tag{28}$$

Fifth, the entropy value of individual indicators is calculated:

$$e_j = -\frac{1}{\ln n} \sum_{i=1}^{n} y_{ij} \ln y_{ij} \tag{29}$$

Sixth, the coefficient of variation for individual indicators is calculated:

$$g_j = 1 - e_j \tag{30}$$

Seventh, the weights of individual indicators were calculated:

$$p_j = \frac{g_j}{\sum_{j=1}^{p} g_j} \tag{31}$$

## 3 Data processing and empirical analyses

### 3.1 Effect of unbalanced data processing

To test the performance of the integrated learning model in this paper when dealing with corporate financial imbalance data, the performance of different predictive classification models is analysed and compared using the dataset.

In order to compare the effect of different machine learning models, seven machine learning models, namely, logistic regression, K-nearest neighbour, decision tree, neural network, random forest, CatBoost, LightGBM, and XGBoost are used as the baseline to compare the prediction effect, and the results obtained are evaluated by four categorical evaluation metrics, namely, precision rate, recall rate, F1 score, and accuracy rate. The results are shown in Table 1. From the evaluation indexes, the accuracy rate, in order from high to low, is based on the models in this paper, LightGBM, Random Forest, Neural Network, K Nearest Neighbor, XGBoost, Logistic Regression, and Decision Tree.

The best classification effect is this paper's model; the main parameters are higher than other classification algorithms, which can be better used for enterprise financial risk and business crisis prediction, and its precision rate, recall rate, and F1 score are 0.95, 0.98, and 0.96, respectively. This paper's integrated learning algorithm has the best performance when applied to unbalanced data.

**Table 1.** Comparison of test results of different models

| Model | P | R | F1 | ACC |
|---|---|---|---|---|
| This model | 0.95 | 0.98 | 0.96 | 0.95 |
| K vicinity | 0.88 | 0.84 | 0.86 | 0.86 |
| LightGBM | 0.82 | 0.89 | 0.84 | 0.89 |
| Logistic regression | 0.82 | 0.83 | 0.89 | 0.84 |
| Neural network | 0.88 | 0.89 | 0.89 | 0.88 |
| Random forest | 0.86 | 0.89 | 0.88 | 0.88 |
| Decision tree | 0.9 | 0.88 | 0.89 | 0.83 |
| XGBoost | 0.87 | 0.84 | 0.87 | 0.84 |

In order to further determine the generalisation ability of the model, as well as to exclude the overfitting phenomenon, the training and testing of the model were re-run after changing the ratio of the test set, and then the evaluation indexes of the comprehensively sampled financial imbalance data under different ratios were obtained. Table 2 displays the results of different proportions.

When changing the proportion of the test set, the accuracy of the integrated learning financial data assessment model in this paper is above 95%, and when the test set is 20%, the model accuracy is the

highest at 98%. From the viewpoint of checking the full rate, the model's early warning ability for companies with financial risks is better, indicating that the model can identify listed companies with financial risks more accurately; from the value of AUC, although the proportion of the test set is changing, the values of AUC are all 98% and above, indicating that the model's generalisation ability is stronger, and there is no overfitting phenomenon.

**Table 2.** Comparison of results of different proportional test results

| Test set ratio | Evaluation index | | | | |
|---|---|---|---|---|---|
| | Accuracy rate | Accuracy ratio | Check rate | F1 value | AUC Value |
| 50% | 0.96 | 0.96 | 0.97 | 0.96 | 0.99 |
| 40% | 0.97 | 0.96 | 0.98 | 0.98 | 0.99 |
| 20% | 0.98 | 0.96 | 0.99 | 0.97 | 0.98 |
| 10% | 0.96 | 0.95 | 0.99 | 0.96 | 0.99 |

## 3.2 Evaluation and analysis of financial data of Company W

The financial data evaluation model designed in this paper is applied to a manufacturing enterprise, Company W, to analyse its financial data and production and operation in recent years, to verify the model's analytical ability in comparison with the actual situation, and to put forward suggestions for the future operation of Company W based on the results of financial data analysis.

### 3.2.1 Empowerment of assessment indicators

According to the algorithm designed above, the entropy value and weight of each indicator are calculated, and the final calculation results are shown in Table 3. The weights of the four first-level indicators of solvency, profitability, operating capacity, and growth capacity are 0.272, 0.236, 0.54, and 0.238, respectively. The A3 cash-current-liability ratio is the top-ranked second-level indicator, and it has a global weight of 0.094.

**Table 3.** The weight of each indicator in the evaluation system

| Primary indicator | Weighting | Secondary indicator | Weighting | Global weight |
|---|---|---|---|---|
| Solvency | 0.272 | Asset liability rate | 0.319 | 0.087 |
| | | Profit ratio | 0.336 | 0.091 |
| | | Cash current liability ratio | 0.345 | 0.094 |
| Profitability | 0.236 | The return on equity | 0.329 | 0.078 |
| | | Total assets return rate | 0.354 | 0.084 |
| | | Profit margin | 0.317 | 0.075 |
| Operational capacity | 0.254 | Total asset turnover | 0.358 | 0.091 |
| | | Turnover rate of accounts receivable | 0.325 | 0.083 |
| | | Inventory turnover | 0.317 | 0.081 |
| Growth ability | 0.238 | Operating rate | 0.324 | 0.077 |
| | | Total asset growth rate | 0.355 | 0.084 |
| | | Capital accumulation rate | 0.321 | 0.076 |

### 3.2.2 Evaluation of financial data of Company W in the last four years

Established in 2009, Company W is an important component supplier in China's consumer electronics industry. The financial data of Company W is calculated and collected from publicly disclosed data, and imported into the model to assess its financial risk, production, and operation.

Table 4 shows the assessment results of the financial data of Company W in 2019. The comprehensive score of the company's financial data in 2019 is 60.29, which indicates some potential financial risks and poor operations. In terms of sub-items, only A3 cash current liabilities ratio, B3 cost and expense margin, D2 total assets growth rate, and D3 capital accumulation rate among the initial scores of the secondary indicators are above 60 points, and the rest of the items are between 54 and 60 points.

In actuality, Company W's products were affected by trade disputes in 2019. Orders were lost to varying degrees both domestically and abroad, and net profit for the year fell sharply.

**Table 4.** W company's financial data assessment results in 2019

| Primary indicator | Financial evaluation | Secondary indicator | Financial evaluation | Initial evaluation |
|---|---|---|---|---|
| A | 16.94 | A1 | 5.19 | 59.82 |
| | | A2 | 5.42 | 59.28 |
| | | A3 | 6.33 | 67.48 |
| B | 14.03 | B1 | 4.32 | 55.66 |
| | | B2 | 4.69 | 56.16 |
| | | B3 | 5.01 | 67.02 |
| C | 14.75 | C1 | 5.41 | 59.47 |
| | | C2 | 4.84 | 58.59 |
| | | C3 | 4.50 | 55.91 |
| D | 14.58 | D1 | 4.21 | 54.57 |
| | | D2 | 5.48 | 64.82 |
| | | D3 | 4.89 | 64.04 |

Table 5 shows the results of the financial data assessment of Company W in 2020. The overall score of the company's financial data in 2020 is 70.80, and the company's operating performance has improved significantly compared to 2019. The lowest of the secondary indicators is C1 total asset turnover 65.36 points, and the highest is C3 inventory turnover 77.92 points.

The actual situation is that Company W is actively exploring new markets, and its market share in Africa and South America has increased significantly. The inventory accumulation situation has greatly alleviated, but there are still more risks in items such as accounts receivable.

**Table 5.** W company's financial data assessment results in 2020

| Primary indicator | Financial evaluation | Secondary indicator | Financial evaluation | Initial evaluation |
|---|---|---|---|---|
| A | 19.55 | A1 | 5.96 | 68.69 |
| | | A2 | 6.34 | 69.32 |
| | | A3 | 7.25 | 77.3 |
| B | 16.97 | B1 | 5.80 | 74.75 |
| | | B2 | 6.12 | 73.28 |
| | | B3 | 5.05 | 67.46 |
| C | 17.73 | C1 | 5.94 | 65.36 |
| | | C2 | 5.51 | 66.8 |
| | | C3 | 6.27 | 77.92 |
| D | 16.55 | D1 | 5.28 | 68.52 |
| | | D2 | 5.84 | 69.08 |
| | | D3 | 5.43 | 71.05 |

Table 6 shows the 2021 financial data assessment results of Company W. The company's financial data composite score in 2021 is 73.11, which is a slight improvement from 2020. Except for the operating ability, which is still in trouble, the solvency, profitability, and growth ability have rebounded compared to the low point in 2019.

In reality, W's turnover capacity is still at a low level in 2021 due to the long cycle of overseas order projects and obstacles to payback. However, despite benefiting from leading technology and stable supply quality, Company W's business volume has already exceeded the 2019 level.

**Table 6.** W company's financial data assessment results in 2021

| Primary indicator | Financial evaluation | Secondary indicator | Financial evaluation | Initial evaluation |
|---|---|---|---|---|
| A | 19.91 | A1 | 6.33 | 73 |
| | | A2 | 7.14 | 78.13 |
| | | A3 | 6.44 | 68.62 |
| B | 18.72 | B1 | 6.10 | 78.51 |
| | | B2 | 6.69 | 80.02 |
| | | B3 | 5.94 | 79.4 |
| C | 17.50 | C1 | 5.88 | 64.63 |
| | | C2 | 6.46 | 78.24 |
| | | C3 | 5.17 | 64.2 |
| D | 16.97 | D1 | 5.59 | 72.54 |
| | | D2 | 5.96 | 70.59 |
| | | D3 | 5.41 | 70.84 |

Table 7 shows the results of the financial data assessment of Company W in 2022. The overall score of the company's financial data in 2022 is 76.52, which is a slight improvement from 2021.C3 Inventory turnover is still below 70, and the rest of the financial data have returned to a good track.

The actual situation is that the demand in the consumer electronics market weakened in 2022, and Company W once again showed a certain degree of tiredness. However, Company W has persisted in prioritising research and development for a long time and has made more progress in technological upgrading and consolidated its leading position in the industry, and its operating conditions have returned to a benign development track.

**Table 7.** W company's financial data assessment results in 2022

| Primary indicator | Financial evaluation | Secondary indicator | Financial evaluation | Initial evaluation |
|---|---|---|---|---|
| A | 21.17 | A1 | 7.20 | 82.93 |
| | | A2 | 7.31 | 80 |
| | | A3 | 6.66 | 70.96 |
| B | 17.31 | B1 | 5.89 | 75.81 |
| | | B2 | 6.38 | 76.39 |
| | | B3 | 5.05 | 67.44 |
| C | 19.58 | C1 | 7.53 | 82.8 |
| | | C2 | 6.77 | 82.04 |
| | | C3 | 5.27 | 65.5 |
| D | 18.46 | D1 | 5.71 | 74.01 |
| | | D2 | 6.76 | 80.04 |
| | | D3 | 5.99 | 78.43 |

In summary, it can be seen that the financial data assessment model in this paper assesses the financial operation of Company W from 2019 to 2022, and the results are all in line with the actual situation, which proves the model's assessment ability and its significance as a guide to production and operation.

Figure 1 shows the trend of debt service, profitability, operation and growth capacity between 2019 and 2022, from which it can be seen that the four first-level indicators basically all move from the low point in 2019 to the high score.

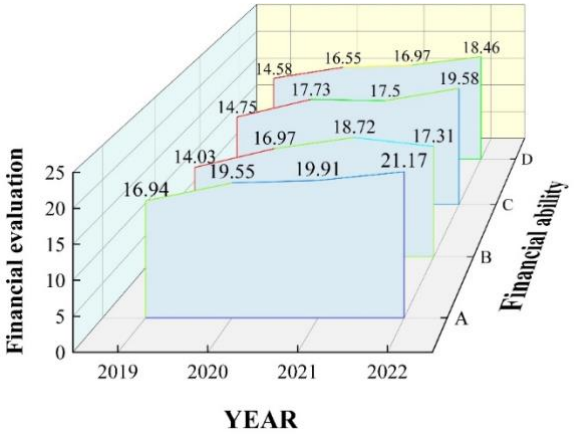



**Figure 1.** Four changes in the ability to change between 2019 and 2022

Based on the results of the analysis, it is recommended that Company W, while maintaining its R&D advantages, focus more on the long-term operating capacity of the enterprise, compress the operating

cycle, reduce the risk of repayment and the pressure on inventories, and continue to enhance the competitiveness of the enterprise.

## 4 Conclusion

In this paper, in order to discover the enterprise financial data problems in time, reduce the risk of enterprise operation and management, and avoid crisis, we construct an accounting and financial data assessment model based on artificial intelligence technology to monitor and assess the various financial data in the enterprise operation. Using the model to analyse the financial data of Company W, it is found that the comprehensive score from 2019 to 2022 is 60.29, 70.80, 73.11 and 76.52, and the financial and operational situation shows a gradual improvement trend. Compared with the actual situation, Company W has experienced events such as a significant drop in net profit due to the loss of product orders in 2019, active development of new markets in 2020, and weakening demand in the consumer electronics market in 2022, resulting in the accumulation of products in storage, etc. However, Company W has persisted in prioritising research and development for a long time and made more progress in technological upgrading to consolidate its industry-leading position, and its operating condition has returned to a benign track of development. The effectiveness of the accounting and financial data assessment model designed in this paper has been verified.

## References

[1] Kureljusic, M., & Karger, E. (2023). Forecasting in financial accounting with artificial intelligence–A systematic literature review and future research agenda. Journal of Applied Accounting Research, (ahead-of-print).

[2] Saukkonen, N., Laine, T., & Suomala, P. (2018). Utilizing management accounting information for decision-making: Limitations stemming from the process structure and the actors involved. Qualitative Research in Accounting & Management, 15(2), 181-205.

[3] VORONKOVA, O. V., KUROCHKINA, A. A., FIROVA, I. P., & BIKEZINA, T. V. (2017). Implementation of an information management system for industrial enterprise resource planning. Revista Espacios, 38(49).

[4] Ernawatiningsih, N. P. L., & Kepramareni, P. (2019). Effectiveness of accounting information systems and the affecting factors. International Journal of Applied Business and International Management (IJABIM), 4(2), 33-40.

[5] Monteiro, A., & Cepêda, C. (2021). Accounting information systems: scientific production and trends in research. Systems, 9(3), 67.

[6] Lv, Y., Li, J., Chen, L., & Li, X. (2021, June). The construction and research of control system and accounting informatization by visualization mode. In Journal of Physics: Conference Series (Vol. 1952, No. 3, p. 032079). IOP Publishing.

[7] Gao, J. (2022). Analysis of enterprise financial accounting information management from the perspective of big data. International Journal of Science and Research (IJSR), 11(5), 1272-1276.

[8] Chyzhevska, L., Voloschuk, L., Shatskova, L., & Sokolenko, L. (2021). Digitalization as a vector of information systems development and accounting system modernization. Studia Universitatis Vasile Goldiș Arad, Seria Științe Economice, 31(4), 18-39.

[9] Deng, J. (2022). The Informatization of Small and Medium-Sized Enterprises Accounting System Based on Sensor Monitoring and Cloud Computing. Mobile Information Systems, 2022(1), 5007837.

[10] Yu, L. (2021, May). Analysis on the application of information processing technology in accounting. In Journal of Physics: Conference Series (Vol. 1915, No. 4, p. 042064). IOP Publishing.

[11] Teng, Y., & Fang, C. (2023). Research on the Transformation from Financial Accounting to Management Accounting Informatization in Universities under the New Government Accounting System. Accounting and Corporate Management, 5(6), 33-42.

[12] Li, M., Wei, W., Wang, J., & Qi, X. (2018). Approach to evaluating accounting informatization based on entropy in intuitionistic fuzzy environment. Entropy, 20(6), 476.

[13] Li, F., & Fang, G. (2022). Process-Aware Accounting Information System Based on Business Process Management. Wireless Communications and Mobile Computing, 2022(1), 7266164.

[14] Junhong, M., & Zehua, W. (2021, February). Research on the intelligentization of accounting in the information technology environment. In 2021 International Conference on Public Management and Intelligent Society (PMIS) (pp. 412-415). IEEE.

[15] Xing, R., & Zhang, J. (2017). Problems and countermeasures of the application of enterprise management accounting informatization. Agricultural Science & Technology, 18(8), 1555-1558.

[16] Ma, Y. M., Huang, J. Y., Danarson, J. H., Huang, Y., & Wei, X. H. (2023, February). Problems and Countermeasures Facing Accounting Informatization in the Era of Big Data. In Proceedings of the 2023 15th International Conference on Machine Learning and Computing (pp. 217-220).

[17] Zhang, M. (2023). Problems and countermeasures of accounting informatization construction in colleges and universities under the background of big data and artificial intelligence. Journal of Computational Methods in Sciences and Engineering, 23(2), 747-757.

[18] Zhou, C. (2024, May). The Influence and Optimization Strategy of Accounting Informatization on Enterprise Financial Management. In 2024 International Conference on Applied Economics, Management Science and Social Development (AEMSS 2024) (pp. 407-414). Atlantis Press.

[19] Luo, R. (2024, May). Research on Accounting Informatization Construction Mode of Small and Medium-Sized Enterprises in Cloud Computing Environment. In International Conference on Artificial Intelligence for Society (pp. 597-605). Cham: Springer Nature Switzerland.

[20] Fikri, N., Rida, M., Abghour, N., Moussaid, K., & El Omri, A. (2019). An adaptive and real-time based architecture for financial data integration. Journal of Big Data, 6, 1-25.

[21] Zeng, Y. (2022). Neural Network Technology-Based Optimization Framework of Financial and Management Accounting Model. Computational Intelligence and Neuroscience, 2022(1), 4991244.

[22] Li, X., Wang, J., & Yang, C. (2023). Risk prediction in financial management of listed companies based on optimized BP neural network under digital economy. Neural Computing and Applications, 35(3), 2045-2058.

[23] Maryam Talebi Moghaddam,Yones Jahani, Zahra Arefzadeh, Azizallah Dehghan, Mohsen Khaleghi, Mehdi Sharafi & Ghasem Nikfar. (2024). Predicting diabetes in adults: identifying important features in unbalanced data over a 5-year cohort study using machine learning algorithm. BMC Medical Research Methodology(1),220-220.

[24] Viet Hoang Ho,Hidenori Morita,Felix Bachofer & Thanh Ha Ho. (2024). Random forest regression kriging modeling for soil organic carbon density estimation using multi-source environmental data in central Vietnamese forests. Modeling Earth Systems and Environment(prepublish),1-22.

[25] Xiaoming Han,Xin Zhao,Yecheng Wu,Zhengwei Qu & Guofeng Li. (2024). A least squares–support vector machine for learning solution to multi-physical transient-state field coupled problems. Engineering Applications of Artificial Intelligence(PA),109321-109321.

[26] Victor Mfon Abia & E. Henry Johnson. (2024). Sentiment Analysis Techniques: A Comparative Study of Logistic Regression, Random Forest, and Naive Bayes on General English andNigerian Texts. Journal of Engineering Research and Reports(9),123-135.

## About the Author

Yanhua Song was born in Hengshui, Hebei, P.R. China, in 1980.She received the Master degree of accounting from Tianjin University of Finance and Economics, P.R. China. Now, she works in School of Management Engineering, Tangshan Polytechnic University, her research interests include financial management and vocational education teaching.