International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18 (2026) pp. 1-17

https://doi.org/10.70917/ijcisim-2026-1873
Article

Research on refined cost management in enterprises based on the integration of business and financial systems

Yaxuan Shi [1,*]

[1] Chengdu University of Technology, Chengdu, Sichuan, 610059, China
* Correspondence author: shiyaxuan06@163.com

Abstract: In the context of economic globalization, enterprises should focus on the integration of business and finance and leverage refined cost management methods to continuously improve their cost control capabilities. By effectively managing costs, enterprises can achieve higher profits. This paper first proposes a method for refined enterprise cost management based on the business-finance integration model. It designs an algorithm that integrates the classic ID3 decision tree with principal component analysis, thereby circumventing the limitations of pre-pruning. This approach enables the selection of more representative decision attributes, reduces the time required for decision tree modeling, and enhances the efficiency of the modeling process. This method is applied to a real-world cost management case study at E-commerce Company A. Analysis of indirect cost allocation across various production processes reveals that, within the production operations center, the three processes—cup body and base stamping, bottom heating, and rim curling—take the longest to complete. These processes account for as much as 36.61% of total production indirect costs, and the scrap rate for defective products is significantly higher than the industry average. This demonstrates that the method proposed in this paper can resolve the issue of unreasonable cost allocation at Company A, enabling the tracing of resources to specific operations and allowing the company to implement effective control over cost allocation.

Keywords: Decision tree algorithm; Principal component analysis; Business-finance integration model; Fine-grained enterprise cost management

1. Introduction

Cost management is a core component of corporate operations and a key factor in achieving efficient operations and sustainable development [1–2]. As the market environment evolves and competitive pressures intensify, traditional cost management models can no longer meet the demands of modern enterprises for refined and data-driven management [3–4]. The business-finance integration model represents a novel management philosophy that embeds financial management into every stage of business operations, achieving seamless connectivity between business and finance and replacing the traditional model where corporate management and financial management were isolated from one another [5-8].

As a significant innovation in modern enterprise management, the essence of business-finance integration lies in breaking down the information barriers between traditional business and finance functions and establishing a two-way, mutually reinforcing collaborative mechanism [9-10]. Through the deep integration of business processes and financial data, this management model achieves a dynamic balance in the optimal allocation of resources [11-12]. In the field of refined corporate cost management, the value of the business-finance integration model is primarily reflected in three dimensions: First, it establishes a closed-loop data system from the business side to the financial side, making cost data collection more precise; second, through real-time data exchange, it significantly enhances the timeliness of cost analysis; finally, a decision-making mechanism based on business-finance synergy makes cost control more forward-looking and scientific.

This paper first proposes a method for enterprise refined cost management from the perspective of business-finance integration, followed by an exposition of the relevant theories of decision tree classification algorithms and principal component analysis (PCA). It then proposes an intelligent identification method (PCA-ID3) that integrates the traditional decision tree ID3 algorithm with principal component analysis (PCA). By introducing attribute thresholds and information gain, the traditional ID3 algorithm is optimized. Principal component analysis is used to select more representative decision attributes, thereby improving the modeling efficiency and accuracy of the decision tree. Taking E-commerce Company A as the specific research subject, the PCA-ID3 algorithm is applied to the company's cost control to achieve modeling and prediction of cost control. Meanwhile, by integrating the unique characteristics of E-commerce Company A, the method proposed in this paper is used to implement comprehensive and refined management of the company's costs before, during, and after operations.

2. Optimized Design of Decision Tree Algorithms

2.1. Methods for Refined Cost Management in Enterprises from a Business-Finance Integration Perspective

1) Establish a Cost Management Indicator System

(1) Scientifically set cost management objectives. When implementing refined cost management, enterprises must establish cost management objectives based on their strategic goals, striving to achieve higher profits at the lowest possible cost.

(2) Analyze the value chains of the industry and competitors. Enterprises need to conduct thorough analyses of the value chains within their industry and among competitors. Based on this analysis, they should estimate competitors' costs to gain the greatest possible cost advantage.

(3) Reduce Tasks with Unclear Objectives. After establishing clear cost management objectives, the company should streamline its tasks, eliminating those with unclear objectives to ensure that all activities contribute to achieving the cost management goals.

(4) Define Cost Responsibilities for Each Department. When constructing the cost management indicator system, the company should ensure that all employees participate in cost management while strengthening the link between indicator performance and individual performance evaluations, thereby creating a more comprehensive cost management indicator system.

(5) Implement cost accounting and refined cost management. After completing the allocation of cost management metrics and tasks, the company must comprehensively implement cost accounting and refined cost management. This involves quantifying all expenses and ensuring the smooth progress of cost control efforts through the aggregation and analysis of monthly and quarterly expense data.

2) Strengthen Communication and Collaboration Between Finance and Business Departments

From the perspective of business-finance integration, when implementing refined cost management, enterprises should prioritize strengthening communication and collaboration between finance and business departments to achieve deep integration between business operations and finance. First, when implementing refined cost management, enterprises must adopt a holistic perspective to conduct research on industry conditions and market trends, ensuring the efficient execution of cost management. Second, management should take the lead in fostering active communication between business and finance departments, requiring staff across all departments to promptly share relevant information in their daily work to ensure that finance personnel can produce more objective and reliable cost analysis reports.

3) Optimizing Cost Management Through Information Technology

When implementing refined cost management, enterprises should leverage information technology to continuously optimize their cost management processes and enhance the level of precision in cost management. In practice, enterprises should take the following steps: First, enterprises need to increase investment in the development of information systems for cost management. Second, they should establish an integrated business and financial management system to gain a detailed understanding of market developments through technological means. Third, enterprises need to apply big data technology to scientifically set cost warning thresholds; when the system detects abnormal cost expenditures, it should promptly issue alerts to finance and business personnel.

4) Implementing Full-Process Cost Management

First, strengthen pre-event cost management. Enterprises should reinforce pre-event cost control, break down overall objectives based on their own development direction, and distribute work plans to all departments.

Second, emphasize in-process cost management. Companies need to establish distinct cost standards based on their development status and business characteristics, then compare these with actual costs to identify variances between actual and standard costs. They should conduct a detailed analysis of the causes of these variances and implement targeted improvements.

Third, strengthen post-event cost management. Companies need to continuously broaden their perspectives, draw on the cost management experiences of other enterprises, comprehensively analyze key cost management indicators from other companies, and, based on their own circumstances, summarize and analyze issues. Through business-finance collaboration, they should formulate strategies to resolve these issues.

2.2. An Improved ID3 Decision Tree Algorithm Based on PCA

2.2.1. Decision Tree Classification Algorithm

1) Overview of Decision Tree Classification Algorithms

Decision trees are used to classify data. In this process, they calculate the probability that the expected net present value is non-negative, thereby estimating a project's risk and determining its feasibility. They represent a graphical method for intuitively applying probability analysis. A decision tree consists of three types of nodes: decision nodes, state nodes, and leaf nodes. Decision trees are typically constructed using a top-down recursive approach. After a decision tree is constructed, it must be pruned. There are two types of pruning algorithms: pre-pruning and post-pruning. Generally, the post-pruning algorithm is used. There are many algorithms for building decision trees; the most commonly used ones include the ID3 algorithm, the C4.5 algorithm, and the CART algorithm.

(1) Attribute Selection Metrics

The formula for attribute selection metrics calculates the splitting rules for attributes. It is a heuristic classification strategy that partitions D, a sample set of training data for known class attributes, into individual classes.

(2) Pre-pruning and Post-pruning

The choice of attribute selection metrics and decision tree pruning strategies are the primary factors distinguishing different decision tree algorithms. There are two basic pruning methods for classification decision trees: pre-pruning and post-pruning.

Pre-pruning: The construction of the decision tree is halted in advance, and then the classification decision tree is pruned. At the point of termination, the current node becomes a leaf node.

Characteristics of pre-pruning: The criterion for stopping the construction of the decision tree is a predefined threshold. Selecting an appropriate threshold is very difficult.

Post-pruning: Pruning is performed after the decision tree has finished "fully growing." The deleted node is replaced with a leaf node, and the corresponding subtree is pruned.

Characteristics of post-pruning: The most frequently occurring class in the pruned subtree is used to replace that subtree as a leaf node.

Pruning methods for decision tree classification algorithms generally fall within the scope of the two strategies described above. A method known as pessimistic pruning, which is used in the C4.5 algorithm, is also a post-pruning strategy.

Additionally, there is a pruning strategy that combines pre-pruning with post-pruning. This pruning method combines the advantages of both strategies, resulting in a final decision tree that is more compact, simpler, and more accurate.

2) ID3 Algorithm

In decision tree learning, ID3 is an algorithm used to generate decision trees from a dataset. ID3 is commonly used in the fields of machine learning and natural language processing. Decision tree technology involves the process of constructing a tree model for classification. Once the decision tree is constructed, it is used to classify each sample in the prediction database and display the classification results.

The ID3 algorithm is a classification algorithm based on information entropy. Its fundamental principle is that all examples are mapped to different classes based on the differences in the median values of the attributes. Its core lies in identifying the best classification attribute from a set of candidate attributes. The attribute selection criterion for splitting in the ID3 algorithm is information gain; generally, the candidate attribute with the highest information gain is selected as the splitting attribute for the current node, in order to minimize the information entropy required by the split subsets. Based on different attribute values, the branch structure of the decision tree can be constructed. The creation of branch structure nodes and subsequent branches is performed recursively by repeating the above steps until all samples are on the same branch, i.e., belong to the same class.

(1) Information Entropy and Information Gain

In physics, information entropy describes the uncertainty of a source of information, while in mathematics, it represents the relationship between information redundancy and probability. Information gain is the difference between two information entropies. The information entropy of the samples in a dataset regarding a particular class is the value of one of these entropies. The other information entropy is the entropy of the samples in that dataset regarding the same class after the value of a certain attribute is known. The specific definitions are as follows:

Let P_i be the probability that a sample takes on the attribute value C_i, and let $\sum p_i = 1$ be defined; then Equation (1) defines the concept of entropy.

$$H\left(p_1, p_2, \cdots, p_s\right) = \sum_{i=1}^{s} -\left(p_i \log p_i\right) \tag{1}$$

Entropy determines the ranking of an attribute within its dataset. $H = 0$ indicates the optimal classification setting for the attribute. In other words, the higher the entropy value, the greater the potential for improvement in the classification process.

Information gain is defined as the difference between the original information requirement and the new information requirement after classification. It is determined by calculating the sum of the original entropy and the weighted entropy for each segmented dataset, as expressed by Equation (2).

$$G(D,S) = H(D) - \sum P\left(D_i\right) H\left(D_i\right) \tag{2}$$

D : Represents a dataset.

D_i : Indicates a sample belonging to class i.

S : Indicates the number of class attributes in the data.

(2) The Concept Behind the ID3 Algorithm

In practical applications, the information gain formula from the ID3 algorithm is used to calculate each feature in the dataset, yielding a series of values. The largest value is selected, and the feature corresponding to that value is chosen as the test feature for the current dataset. If we use a decision tree to represent the dataset, the node corresponds to the test attribute selected through this calculation. Using this attribute as a label, each of its values represents a branch, and the samples in the dataset are partitioned accordingly. The core idea of the algorithm is as follows:

For a dataset T, let T be a training set of samples with known class labels. Let:

$C_i (i = 1, 2, \ldots\ldots, n)$: The class label attribute has n distinct values.

TC_i : The set of samples in the training set T that belong to class C_i.

$|T|$: The number of samples in the training set TC_i.

$\left|TC_i\right|$: The number of samples in the training set TTT.

Equation (3) represents the expected information required for sample classification.

$$Info(T) = -\sum_{i=1}^{n} p_i \log_2\left(p_i\right) \tag{3}$$

p_i : The non-zero probability that any sample in the T dataset belongs to class C_i, estimated using the $\frac{\left|TC_i\right|}{|T|}$ method.

$Info(T)$: The average information entropy required to identify the class labels of samples in the T dataset.

This formula uses only the proportion of tuples in each class relative to the total number of tuples in the dataset.

Now, let's partition the samples in the training set according to the attribute A, which has m distinct values, namely $\left\{a_1, a_2, \cdots, a_m\right\}$. Assuming A is a discrete attribute, the values correspond one-to-one with the m outputs tested on attribute A. The attribute A divides the training set T into m subsets $\left\{T_1, T_2, \cdots, T_m\right\}$, where T_j contains the tuples from T whose values for attribute A are a_j.

Equation (4) represents the information entropy of classifying tuples into T based on the A classification.

$$Info_A(T)=\sum_{j=1}^{m}\frac{|T_j|}{|T|}Info\left(T_j\right) \tag{4}$$

$\frac{|T_j|}{|T|}$: Represents the proportion of the number of samples where attribute A takes the j th value out of the total number of samples.

$Info_A(T)$: Represents the expected information required to classify the samples in dataset T based on attribute A . The smaller this value, the higher the purity of the class to which each attribute value belongs.

Based on the definition of information gain, the specific calculation formula is shown in Equation (5).

$$Gain(A)=Info(T)-Info_A(T) \tag{5}$$

Equation (5) represents the reduction in the expected information gain given the value of attribute A. The attribute with the highest information gain is selected as the splitting attribute for the current node.

(3) ID3 Algorithm Flowchart

A decision tree consists of three types of nodes: decision points, internal nodes, and leaf nodes. Decision trees are typically constructed using a top-down recursive strategy. After the decision tree is constructed, it must be pruned; there are two types of pruning algorithms: pre-pruning and post-pruning.

The algorithmic flow for the ID3 algorithm to create a decision tree is shown in Figure 1.

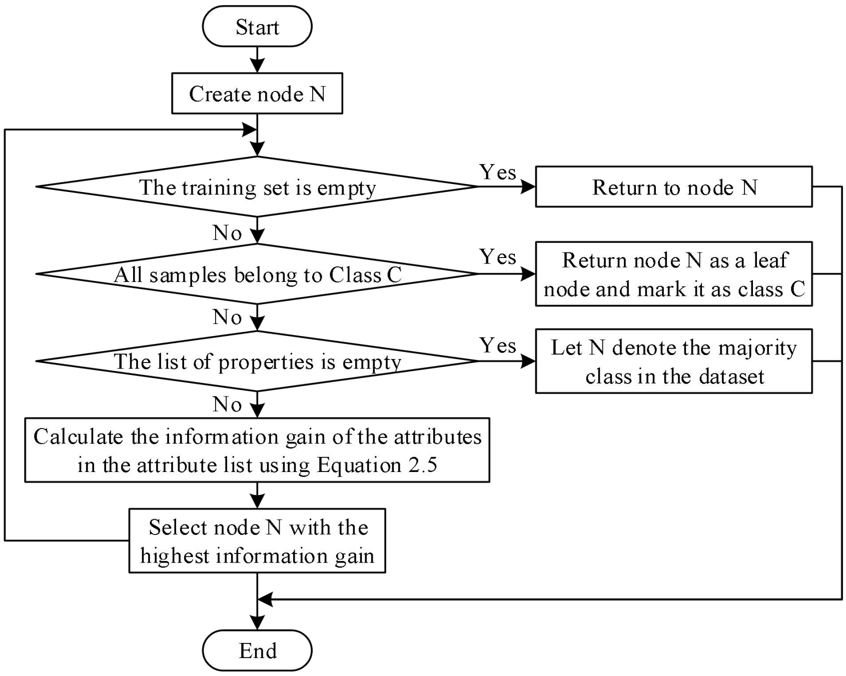

Figure 1. The ID3 algorithm creates a decision tree process

2.2.2. Principal Component Analysis

Principal Component Analysis (PCA) is a statistical method used to simplify datasets; it is commonly employed for dimensionality reduction, which involves transforming multiple variables into a small number of principal components. These principal components reflect the vast majority of the information contained in the original variables and are typically represented as linear combinations of the original variables.

The primary objective of Principal Component Analysis is to transform multidimensional variables into lower-dimensional ones and eliminate redundant information to reduce data complexity. This method primarily involves performing an eigenvalue decomposition on the covariance matrix to derive the principal components of the data and their respective weights. The specific calculation steps are as follows:

1) Suppose the dataset X contains n samples, denoted as $X = [x_1, x_2, \cdots, x_n]$, where each sample has m dimensions.

2) Compute the covariance matrix R, where $i = 1, 2, \cdots, n$.

$$R = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \frac{\sum_{i=1}^{n} x_i}{n}\right)\left(x_i - \frac{\sum_{i=1}^{n} x_i}{n}\right)^T = \frac{1}{\sqrt{n}}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(x_i - \frac{\sum_{i=1}^{n} x_i}{n}\right)\left(x_i - \frac{\sum_{i=1}^{n} x_i}{n}\right)^T = UU^T \tag{6}$$

3) Compute the eigenvalues and eigenvectors.

Compute the eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_n$ and corresponding eigenvectors $U = [u_1, u_2, \cdots, u_n]$ of the covariance matrix R, and reorder the eigenvalues so that λ_1开λ_2开$\cdots$开λ_n.

$$\lambda_i u_i = R u_i, i = 1, 2, \cdots, n \tag{7}$$

4) Calculate the cumulative information contribution rate η_p of the first $p(p \bullet n)$ principal components in the eigenvalues.

The formula for calculating the information contribution rate y_i of the eigenvalue λ_i is as follows:

$$y_i = \frac{\lambda_i}{\sum_{i=1}^{n} \lambda_i} \tag{8}$$

The formula for η_p is as follows:

$$\eta_p = \frac{\sum_{i=1}^{p} \lambda_i}{\sum_{i=1}^{n} \lambda_i} \tag{9}$$

Generally, if the cumulative information contribution rate η_p of the first $p(p \bullet n)$ principal components among the eigenvalues reaches 85% or higher, it indicates that these first p principal components can represent the vast majority of the information in the entire dataset.

5) Calculate the dimension reduction results.

The transformation matrix $T_p = (t_1, t_2, \cdots, t_p)$ is composed of the eigenvectors corresponding to the first p eigenvalues, and the result of the dimension reduction is $p_i = T_p X_1$.

Principal component analysis not only reduces the dimensionality of a dataset but also retains the features that contribute most to the variance in the dataset. This is achieved by retaining the lower-order principal components and ignoring the higher-order ones, as the lower-order components typically capture the most important aspects of the data. When data has multiple dimensions, some dimensions contribute significantly to the data, while others contribute less. By using PCA to retain important dimensions and eliminate less important ones, data complexity can be reduced, computational demands for data processing can be minimized, and classification efficiency can be improved. It is evident that using PCA for dimensionality reduction has a minor impact on classification accuracy but a significant impact on classification efficiency.

2.2.3. Based on the Integration of the Improved ID3 Decision Tree Algorithm and Principal Component Analysis

1) Improvements to the ID3 Decision Tree

In the traditional classical ID3 decision tree algorithm, feature attribute values vary across different categories. The algorithm partitions subsets based on the method of distinct values, and the number of subsets used to construct the decision tree corresponds one-to-one with the number of individual attributes. The branch endpoints form leaf nodes, which constitute the decision tree rules. In practice, the number of feature types in attribute values affects the leaf nodes, decision paths, and decision rule results of the decision tree, increasing the computational load and leading to issues such as multi-valued tendencies and low accuracy. Therefore, this paper proposes an improved and optimized ID3 algorithm, with the specific steps as follows:

(1) Select a training attribute to determine the classification.

(2) Create a node.

(3) If the current data belongs to the same class, set the node as a leaf node of the decision tree and label it with the class.

(4) If the current data is an empty set, set the node as an empty leaf node and label it with the class based on the majority rule.

(5) Calculate the expected value and select the optimal attribute as the test attribute for the node based on the expected values of the attributes.

2) Fusion Description Based on the Integration of the Improved Decision Tree and PCA Algorithms

The node attributes and classes are fused by combining the improved ID3 decision tree algorithm with the principal component analysis (PCA) algorithm. The specific steps are as follows:

(1) Select dataset A . First, compress the dataset using the PCA algorithm to select important features.

Step 1: Initialize the matrix containing the required data volume to create data matrix $X_{m\times n}$, where m represents the number of data rows and n represents the spatial dimension of the data records in the matrix.

Step 2: Data standardization involves normalizing the data so that the mean is 0 and the standard deviation is 1, thereby aligning the data within the same matrix. The formula used for standardization is:

$$Z(X)=\frac{x-\overline{x}}{s(x)}=\frac{x-\overline{x}}{\sqrt{\frac{\left(x-x^{2}\right)}{n}}} \tag{10}$$

Step 3: Covariance describes the relationship between two measures of a physical quantity. Its formula is:

$$\operatorname{cov}(x,y)=\frac{\sum_{i=1}^{n}\left(X_{i}-\overline{X}\right)\left(Y_{i}-\overline{Y}\right)}{n-1} \tag{11}$$

In an n -dimensional data matrix, the covariance between any two data points can be calculated. Therefore, from an n -dimensional matrix of data, an $n\times n$ -dimensional covariance matrix can be obtained.

$$\operatorname{cov}\left(X_{n*n}\right)=\left(C_{ij},C_{ij}=\operatorname{cov}\left(\operatorname{dim}_{i},\operatorname{dim}_{j}\right)\right) \tag{12}$$

If X_i represents the i th element of the matrix, it can be expressed as:

$$coeff=\begin{bmatrix}\operatorname{cov}\left(x_{1},x_{i}\right) & \cdots & \operatorname{cov}\left(x_{1},x_{n}\right)\\ \vdots & \ddots & \vdots\\ \operatorname{cov}\left(x_{n},x_{1}\right) & \cdots & \operatorname{cov}\left(x_{n},x_{n}\right)\end{bmatrix} \tag{13}$$

By solving the linear equations of the original variables, we can obtain the principal components of the composite indicators. We then analyze the structural characteristics of the resulting matrix and calculate its corresponding eigenvalues.

(2) Determine the classification attributes.

(3) Create a node N .

(4) If all data points belong to the same class, N is a leaf node; label the class to which it belongs on the leaf.

(5) Apply the PCA algorithm to the selected class to further compress the data and identify important features.

(6) If no attributes remain in the residual dataset, N is also a leaf node of the decision tree branch; label the category based on the attribute with the highest selected weight influence.

(7) Otherwise, select the optimal attribute based on the expected average information gain as the test attribute for node N .

The process of integrating the principles of Principal Component Analysis (PCA) with the ID3 decision tree algorithm is shown in Figure 2.

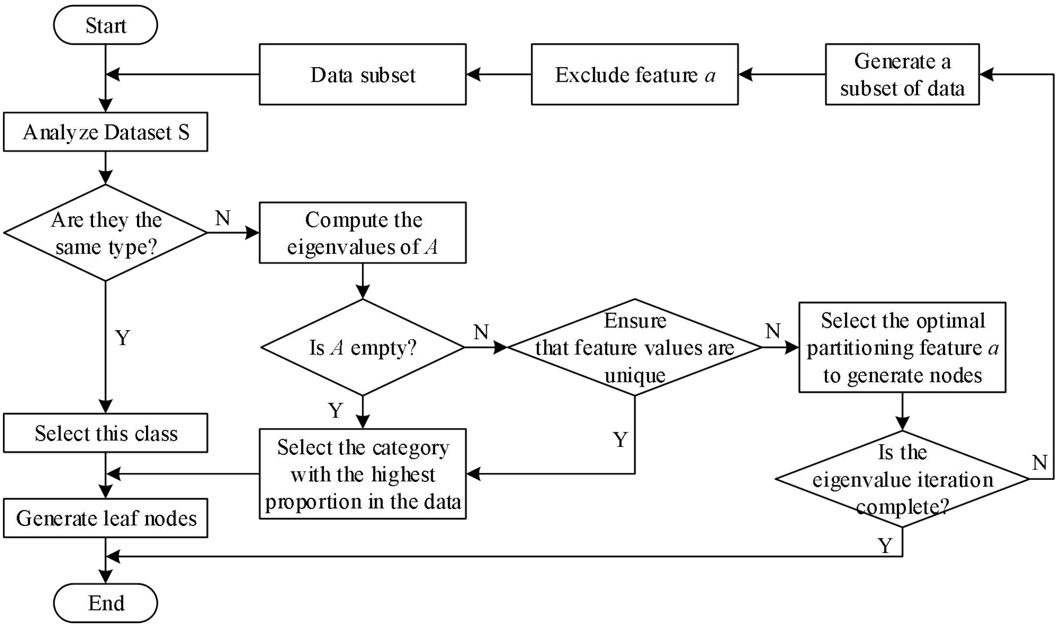

Figure 2. Improve the fusion process of decision tree and PCA algorithm

3. Application and Analysis of an Improved Decision Tree Algorithm in Enterprise-Level Fine-Grained Cost Management

To validate the PCA-ID3 algorithm, the authors selected data from the UCI Machine Learning Repository for testing. The first dataset chosen was the wine dataset, which contains 200 samples and 13 features. Figure 3 shows the comparison results for the wine dataset: Figure a shows the results obtained using the original ID3 algorithm, while Figure b shows the results of the optimized algorithm. It is clear from these two figures that there are six non-overlapping points under the original ID3 algorithm, whereas there is only one non-overlapping point after optimization. This demonstrates that the PCA-ID3 algorithm achieves higher accuracy than the original algorithm.

3.1. The Application of PCA-ID3 in Corporate Cost Control

3.1.1. The Application of PCA-ID3 in Corporate Cost Control

To validate the PCA-ID3 algorithm, the author selected data from the UCI Machine Learning Repository for testing. The first dataset chosen was the wine dataset, which contains 200 samples and 13 attributes. A comparison of the results for the wine dataset is shown in Figure 3: Figure a displays the results obtained using the original ID3 algorithm, while Figure b shows the results from the optimized algorithm. From the two figures, we can clearly see that there are a total of 6 non-overlapping points under the original ID3 algorithm, whereas after optimization, there is only 1 non-overlapping point, proving that the accuracy of the PCA-ID3 algorithm is higher than that of the original algorithm.

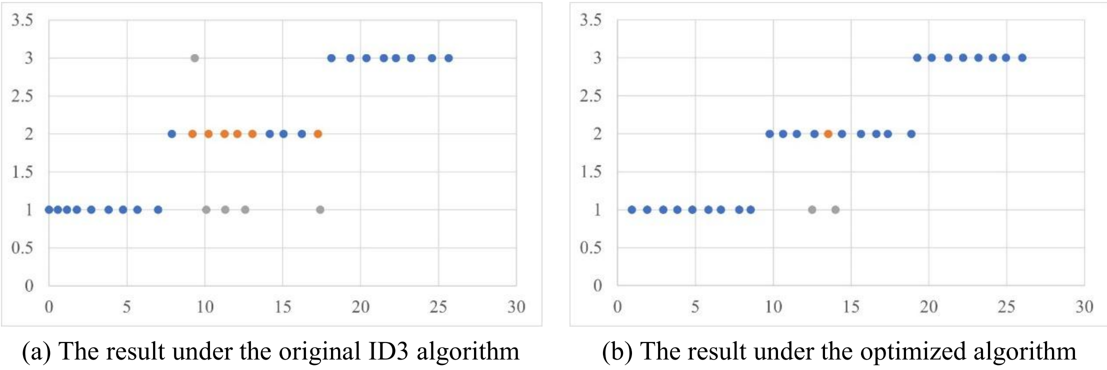

Figure 3. Comparison of results from the wine dataset

In addition, I tested two other datasets, namely the "adult" and "car" datasets, and compared them with the standard ID3 algorithm combined with PCA. The comparison of accuracy rates is shown in Table 1. As can be seen from the table, across all compared datasets, the algorithm proposed in this paper outperforms both the original ID3 algorithm and the standard PCA-combined algorithm in terms of accuracy, providing reason to believe that this algorithm offers superior optimization results.

After undergoing double compression via PCA, the accuracy improved to a certain extent. In the Wine dataset, the optimized algorithm achieved an accuracy 1.3% higher than the traditional ID3 algorithm and 0.2% higher than the standard PCA-fusion algorithm. In addition, two data mining classification algorithms—KNN and Naive Bayes—were selected for comparison on the Adult dataset. The KNN algorithm achieved an accuracy of 80.1%, while the Naive Bayes algorithm achieved 91.18%, both of which were lower than the accuracy of the optimized algorithm in this paper, demonstrating the practical significance of our algorithm.

Table 1. Comparison of algorithm accuracy

	Wine dataset/%	**Adult dataset/%**	**Car dataset/%**
The original ID3 algorithm	92.8	93.4	93.7
Common PCA combined algorithm	93.9	95	93.1
This algorithm	94.1	95.8	94.3

3.1.2. Comparison of Algorithms Before and After Optimization

1) Decision Tree Modeling of the Optimization Algorithm

To further validate the advantages of the optimization algorithm, this paper uses the dataset from E-commerce Company A as a basis. The algorithm was rewritten in Python under the Microsoft Windows operating system, and the rewritten algorithm was then simulated. The decision tree for E-commerce Company A generated by the optimization algorithm is shown in Figure 4.

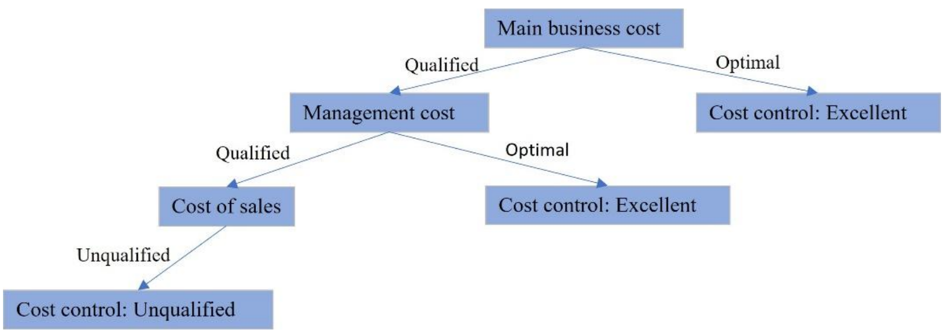

Figure 4. The enterprise decision tree obtained by optimization algorithm

In addition, based on the new decision tree, we can derive certain rules related to corporate cost control decisions. These rules differ somewhat from those obtained using the ID3 algorithm. The specific rules are as follows:

(1) IF Core Business Cost Control = Excellent, THEN Cost Control = Excellent

(2) IF Core Business Cost Control = Acceptable AND Administrative Cost Control = Excellent,

THEN Cost Control = Excellent

(3) IF Core Business Cost Control = Acceptable AND Administrative Cost Control = Acceptable AND Sales Cost Control = Unacceptable, THEN Cost Control = Unacceptable

From the above experimental results, we can observe that when conducting experiments using the improved algorithm described in this paper, the PCA algorithm retained only three cost items for analysis—specifically, production costs—resulting in a more streamlined overall decision tree. Furthermore, based on the evaluation results after decision tree modeling, the accuracy of the decision tree modeled using the optimization algorithm was higher, reaching 96.3%, while the modeling time was shorter, at 8.5 seconds. In addition, I applied the data to a standard PCA fusion algorithm, which achieved an accuracy of 93.1%, but took less time than the optimized algorithm described in this paper, at 6.2 seconds.

2) Comparative Analysis

A summary of the comparison between the pre- and post-optimization algorithms is shown in Table 2. The optimized algorithm is faster in terms of decision tree modeling speed, reducing the modeling time by 5.9 seconds compared to the traditional ID3 algorithm. In terms of modeling time, total number of nodes, number of leaf nodes, number of decision rules, and average accuracy, the algorithm in this paper demonstrates significant advantages, making it more practical for constructing decision trees for enterprise cost control.

Table 2. A comparison and summary of the algorithms before and after optimization

Comparison item and	**ID3 algorithm**	**This paper optimizes the algorithm**	**Comparison results**
Modeling time	13.5s	7.6s	The optimization has been improved by 5.9 seconds before and after
Total number of nodes	10	5	5 were reduced before and after optimization
The number of leaf nodes	5	3	3 were reduced before and after optimization
The number of decision rules	5	2	3 items were reduced before and after optimization
Average accuracy rate	92%	96.7%	It has increased by 4.7% before and after optimization

3.1.3. Research Expansion

In addition, we can use other attributes from the company's cost accounts as decision attributes for the experiment. After reducing the company's cost data, we retained four attributes: posting date, voucher type, account name, and debit amount. We can also use voucher types from the company's cost accounts as classification attributes for testing. In the company's cost data, voucher types include AF depreciation entries, WA (Shipment), and SA (General Ledger Voucher). Therefore, this paper conducts exploratory experiments using the voucher type from the enterprise's cost accounts as a decision attribute. First, the experimental data is aggregated based on voucher type. The statistical table of cost data with voucher type as the decision attribute is shown in Table 3 (Unit: 10,000 yuan).

Table 3. Quick Strategy Attribute statistics

Month	**Administrative expenses**					**Production cost**				**Cost of sales**				**Main business**	**Assembly book**	**marking**
	AF	**WA**	**ZX**	**ZY**	**ZZ**	**SA**	**WA**	**ZZ**	**AF**	**SA**	**WA**	**ZY**	**ZZ**	**WL**		
1	11.69	0.44	0.89	15.71	25.39	115.7	3882.79	86.75	0.07	820.75	5.41	0.1	7.886	2535.8	7509.376	0
2	12.48	0.21	1.87	3.78	34.21	90.09	2732.16	47.28	0.07	605.73	8.45	0.41	0.69	3247.7	6785.13	0
3	11.65	1.02	2.03	4.6	28.15	301.44	4733.71	76.69	0.07	698.74	12.41	11.28	15.34	4490.9	10388.03	1
4	12.21	1.46	1.64	4.09	20.956	173.26	3068.25	58.84	0.07	654.25	15.05	1.88	9.25	3944.2	7965.406	0
5	11.58	2.32	1.69	3.42	19.36	444.35	4552.27	86.87	0.07	660.05	9.98	0.57	8.24	5093.9	10894.67	1
6	11.87	0.51	2.43	5.21	21.31	461.68	5654.08	95.01	0.07	666.66	12.19	0.57	9.43	5495.2	12436.22	1
7	12.4	1.54	1.85	33	25.38	354.19	5727.34	117.64	0.07	917.68	11.81	7.43	8.71	6063.1	13282.14	1
8	--	--	--	--	--	--	--	--		--	--	--	--	--	0	0
9	12.02	0.12	1.98	4.61	18.67	446.71	627.17	106.18	0.08	910.48	16.24	0.18	11.3	7031.6	9187.34	0
10	11.76	1.45	1.23	2.48	29.1	490.92	5861.55	115.47	0.08	892.37	10.34	3.85	11.07	--	7431.67	0
11	11.85	0.14	2.62	11.5	27.87	333.32	5926.59	98.62	0.08	870.44	8.28	1.92	6.36	5475	12774.59	1
12	102.55	2.83	2.79	1.77	35.81	202.73	4978.28	83.08	0.08	1488.72	38.89	2.24	13.43	5336.6	12289.8	1
Average value	34.545	1.135	2.155	5.09	27.8625	368.42	4348.3975	100.8375	0.08	1040.5025	18.4375	2.0475	10.54	5947.733333	10420.85	

Data Notes: 1. The data categorizes monthly cost figures into administrative expenses, production costs, sales costs, and cost of goods sold. Since the raw data is missing August figures, those figures have been excluded here. 2. The total cost control process is as follows: values exceeding the average cost are marked as 1, while those below the average cost are marked as 0. This also serves as the criterion defined in this paper for evaluating the effectiveness of cost control. A mark of 1 indicates that the cost value is high, while a mark of 0 indicates that cost control is satisfactory. 3. A decision tree is used for evaluation. The cost data is normalized using voucher type as the decision attribute, with the results shown in Table 4.

Table 4. The cost data is normalized as the decision attribute based on the type

Month	Administrative expenses					Production cost			Cost of sales					Main business	Assembly book
	AF	WA	ZX	ZY	ZZ	SA	WA	ZZ	AF	SA	WA	ZY	ZZ	WL	
1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	1	0	1	0	0	0	0	1	1	0	0
4	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	1	0	0	0	1	1	0	0	0	0	0	0	0	1
6	0	0	1	1	0	1	1	0	0	0	0	0	0	0	1
7	0	1	0	1	0	0	1	1	0	0	0	1	0	1	1
9	0	0	0	0	0	1	0	1	1	0	0	0	1	1	0
10	0	1	0	0	1	1	1	1	1	0	0	1	1	--	0
11	0	1	1	1	1	1	1	0	1	0	0	0	0	0	1
12	1	1	1	0	1	0	1	0	1	1	1	1	1	0	1

3.2. A Study on the Fine-Grained Cost Management of E-Commerce Enterprises Based on the Methodology Presented in This Paper

3.2.1. The Process of Detailed Cost Management in A E-commerce Company

1) Preliminary Preparations for Detailed Cost Management at E-commerce Company A

(1) Based on the actual circumstances of E-commerce Company A, this paper defines the scope of accounting as the production, logistics, warehousing, and marketing departments' entire business activities for March 2025.

(2) Identification of Corporate Resources

Based on business operations, the resources of E-commerce Company A have been classified as follows, as shown in Table 5. In practice, most companies tend to focus solely on reducing direct costs—such as raw materials and direct labor—while neglecting indirect costs. However, from the perspective of enhancing corporate competitiveness, controlling indirect costs is particularly important, as it directly impacts the company's profitability and competitive advantage. Therefore, the following cost analysis will center on indirect costs.

Table 5. Classification of e-commerce enterprise resources

Resource type	Specific content	
Direct cost	Personnel fee	Employee pay
	Materials cost	Raw material cost
	Site charge	Site rent, etc
Indirect costs	Daily consumption	Low value consumption, auxiliary materials, etc
	Transport cost	Transportation fees, printing fees, etc
	Equipment cost	Equipment depreciation, purchase, maintenance, maintenance, etc
	Power cost	Water and electricity charges, etc.
	Other expenses	Advertising fees, entertainment fees, etc

(3) Identifying Activity Centers and Integrating Activities

Identifying activity centers and integrating specific activities based on the defined product and service processes is a critical step in cost accounting using the Activity-Based Costing (ABC) method. In an e-commerce enterprise, activity centers can be divided by department, including production, logistics, warehousing, and marketing activity centers.

2) Real-Time Cost Management System for an E-Commerce Enterprise

Based on the above data, the following three sets of data can be derived:

(1) Actual Total Costs for Each Activity Center

The summary of costs for each activity center is shown in Table 6 (in RMB). As indicated by the cost summary in the table, the amount of indirect costs incurred varies across different activity centers, resulting in differing proportions of total actual costs. Some activity centers have a lower proportion, while others have a higher proportion. Clearly, the accurate allocation of indirect costs is particularly important for understanding the consumption of resources across each activity center.

Table 6. Cost summary of each operation center(yuan)

Name of the operation center	Name of the operation center	Indirect costs	Actual total cost
Production center	1608900	716309	2325209
Logistics operation center	148723	285480	434203
Warehouse operation center	155613	64889	220502
Marketing operation center	70922	123887	194809
Total	1984158	1190565	3174723

(2) Theoretical Working Hours for Each Operational Center

At E-commerce Company A, with the exception of functional positions that follow a "9-to-5" work schedule, the standard working hours for other departments are 8:00 a.m. to 5:00 p.m. Additionally, the production and warehousing departments are required to work an extra 4 hours every Wednesday evening and work as usual on Saturdays. In summary, there were 22 working days in March 2025, while the production and warehousing departments worked for 27 days. According to statistics provided by relevant managers, no employees at E-commerce Company A took leave in March. The theoretical working hours for each operational center are shown in Table 7.

Table 7. Theoretical working hours of each operation center

Name of the Operation Center	Average number of employees	Theoretical working hours (hours)	Theoretical working hours (minutes)
Production center	57	13448	807330
Logistics operation center	8	1581	95032
Warehouse operation center	12	2801	167022
Marketing operation center	21	3181	190069

(3) Unit operation time and number of operations for each task

Time equations are required when collecting unit operation times. In fact, the need for time equations varies by work center, but production departments tend to use them more frequently due to the greater complexity of their operations. Here, we will use a production work center as an example. The processing times for each sub-operation in the production work center are shown in Table 8, illustrating the data collection process for unit processing times. When an operation occurs, the number of executions for each sub-operation should be recorded in the database to facilitate the direct application of time equations during cost accounting.

Table 8. The working hours of each sub-operation in the production operation center

Homework content	Homework time (minutes)	Number of assignments	Unit operation time (minutes)
Original paper coating	2	3751	7502
Tangential	3	16496	49488
Typesetting	2	8413	16826
Print	5	7430	37150
Membrane Cutting	6	9556	57336
Disinfection	7	3587	25109
The body and bottom of the cup are mixed together	3	13493	40479
Bottom heating and ironing	7	12840	89880
The rim of the cup is curled	11	11986	131846
Forming	6	11278	67668
Auxiliary material addition	2	3558	7116
Basic inspection	3	2880	8640
Scrap of defective products	15	2038	30570
Disinfection packaging	5	3872	19360

Packing	8	1494	11952
Total	85	112672	600922

Based on the collected baseline data, the following cost accounting process is initiated:

(1) Calculate the unit-time production cost for each center. Assuming a rough effective working hours ratio of 83% for E-commerce Company A, the aggregate unit-time production costs for each operational center are estimated based on actual conditions. The aggregated unit production costs for each operational center are shown in Table 9.

Table 9. Summary of the production capacity of each operation center unit

Homework Center	Average number of employees	Indirect cost (yuan)	Effective working hours (minutes)	Unit time capacity cost (yuan/min)
Production center	57	716309	670016.8	1.069
Logistics operation center	8	285480	79069.5	3.610
Warehouse operation center	12	64889	138700.5	0.468
Marketing operation center	21	123887	157781	0.785

(2) Allocate indirect costs based on the standard operating time of each activity center

The allocation of indirect costs and cost ratios for each activity center and activity at E-commerce Company A in March 2025 are shown in Tables 10 through 13.

Table 10. Indirect allocation table of production operation centers

Homework content	Unit operation time (minutes)	Indirect cost allocation (yuan)	Current period cost proportion
Original paper coating	7502	8019.638	1.05%
Tangential	49488	52902.672	6.91%
Typesetting	16826	17986.994	2.35%
Print	37150	39713.35	5.19%
Membrane cutting	57336	61292.184	8.00%
Disinfection	25109	26841.521	3.51%
The body and bottom of the cup are combined	40479	43272.051	5.65%
Bottom heating and ironing	89880	96081.72	12.55%
The rim of the cup is curled	131846	140943.374	18.41%
Forming	67668	72337.092	9.45%
Auxiliary material addition	7116	7607.004	0.99%
Basic inspection	8640	9236.16	1.21%
Scrap of defective products	30570	32679.33	4.27%
Disinfection packaging	19360	20695.84	2.70%
Packing	11952	12776.688	1.67%
Total	600922	642385.618	83.89%
Remaining production capacity	69094.8	73923.382	16.11%

Table 11. Indirect allocation table of Logistics operation center

Homework content	Unit operation time (minutes)	Indirect cost allocation (yuan)	Current period cost proportion
Assignment of goods	17425	62904.25	22.03%
Distribution inspection	13434	48496.74	16.99%
Hauling	31296	112978.56	39.57%
Refueling	3279	11837.19	4.15%
Maintenance	302	1090.22	0.38%
Total	65736	237306.96	83.13%
Remaining production capacity	13333.5	48173.04	16.87%

Table 12. Indirect allocation table of the warehousing operation center

Homework content	Unit operation time (minutes)	Indirect cost allocation (yuan)	Current period cost proportion
Purchasing warehousing	50069	23432.292	36.11%
Acceptance	12464	5833.152	8.99%
Split sorting	15283	7152.444	11.02%
Feed out	43587	20398.716	31.44%
Total	121403	56816.604	87.56%
Remaining production capacity	17297.5	8072.396	12.44%

Table 13. Indirect allocation table of the Marketing Operations Center

Homework content	Unit operation time (minutes)	Indirect cost allocation (yuan)	Current period cost proportion
Merchandise editor	17695	13890.575	11.21%
Brand management	4491	3525.435	2.85%
Activity planning	13745	10789.825	8.71%
Promote channel traffic	15063	11824.455	9.54%
Customer relationship maintenance	6220	4882.7	3.94%
Pre-sale consultation	18589	14592.365	11.78%
After-sales maintenance	5619	4410.915	3.56%
Product design	13594	10671.29	8.61%
Public relations	5887	4621.295	3.73%
Customer service training	2917	2289.845	1.85%
Introduction to goods	18839	14788.615	11.94%
total	122659	96287.315	77.72%
Remaining production capacity	35122	27599.685	22.28%

This concludes the allocation of indirect costs across the four key operational centers—production, logistics, warehousing, and marketing—at E-commerce Company A. The entire process of real-time, detailed cost management has been completed, and relevant finance personnel have compiled statistics and categorized the data, thereby providing robust data support for the subsequent detailed cost management process.

(3) Detailed Post-Event Cost Management at E-commerce Company A

The variances between budgeted and actual amounts are shown in Table 14 (in RMB).

Table 14. The discrepancy between the budget and the actual amount incurred(yuan)

Cost classification	Budget amount	Actual expenditure	Difference	Rate of difference
Production cost	2125537	2319879	194342	9.14%
Logistics cost	378551	433949	55398	14.63%
Warehousing cost	209152	222975	13823	6.61%
Marketing cost	173833	192895	19062	10.97%
Total	2887073	3169698	282625	9.79%

Based on the data comparison in the table and the allocation of indirect costs across each process, the following analysis can be conducted:

1) In the production center, the operations of punching and joining the cup body and bottom, bottom heating, and rim curling are performed manually. These three processes are time-consuming and account for as much as 36.61% of total production indirect costs. Additionally, the scrap rate on the entire production line is significantly higher than the industry average.

2) Regarding the company's logistics operations, goods distribution and inspection account for 16.99% of total indirect costs. These activities are non-value-added tasks that only generate indirect value for the company.

3) Receiving and issuing goods are indispensable components of the warehousing operations center; however, the combined indirect costs for these two operations account for as much as 67.55%. Reducing indirect costs associated with receiving and issuing goods can effectively lower the total indirect costs of the warehousing operations center.

4) To meet marketing requirements, the company spends excessively on website maintenance and hiring relevant professionals. However, this operations center exhibits the highest proportion of unused capacity among all centers, indicating low employee productivity and consequently low effective working hours. The company should appropriately streamline its workforce by reducing the number of low-performing employees and hiring more talented individuals to drive the development of the company's marketing capabilities.

3.2.2. Comparative Analysis of Methods

The indirect costs allocated using the method described in this paper are shown in Table 15. The indirect costs allocated using traditional cost accounting methods are shown in Table 16. After calculating and comparing the results, we can see that the method described in this paper yields unique information—unused time and excess capacity. This is due to the discrepancy between available resources and actual resource consumption in real-world operations. Based on practical experience, no single operation can utilize existing resources 100%. Data on idle production capacity serves as crucial information for managers in formulating corporate governance strategies, enabling them to address inefficiencies at their source and further enhance corporate profitability. In summary, the implementation of the method proposed in this paper is feasible in the specific context of E-commerce Company A.

Table 15. The indirect costs allocated by the method proposed in this paper

Cost accounting items		Total time (minute)	Indirect cost allocation (yuan)	Cost composition
Logistics operations	Assignment of goods	17425	62904.25	26.51%
	Distribution inspection	13434	48496.74	20.44%
	Hauling	31296	112978.56	47.61%
	Refueling	3279	11837.19	4.99%
	Maintenance	302	1090.22	0.46%
Effective utilization		65736	237306.96	-
Practical usability		17425	62904.25	-
Remaining production capacity		13434	48496.74	-

Table 16. Indirect costs are allocated by using traditional cost accounting methods

Cost accounting items		Business volume	Indirect cost allocation (yuan)	Cost composition
Logistics operations	Assignment of goods	1880	25.44%	25.44%
	Distribution inspection	1418	20.15%	20.15%
	Hauling	3605	49.76%	49.76%
	Refueling	323	4.36%	4.36%
	Maintenance	12	0.30%	0.30%
Total		7238	282443.04	100.00%

4. Conclusion

Through cost control and resource allocation, enterprises can achieve greater economic benefits in their operations and development. This paper explores methods for refined cost management from the perspective of business-finance integration and investigates the application of the PCA-ID3 algorithm in cost control at an e-commerce enterprise. The following conclusions were drawn:

1) Validation of the improved algorithm using the UCI database revealed that the PCA-ID3 algorithm achieved a 1.3% higher accuracy rate than the traditional ID3 algorithm and a 0.2% higher accuracy rate than standard PCA fusion algorithms. This demonstrates that PCA-based decision tree algorithms can enhance result accuracy to a certain extent, holding practical significance.

2) In the post-event refined cost analysis of E-commerce Company A, the combined proportion of indirect costs associated with the two operations—purchasing and warehousing (inbound) and material issuance and warehousing (outbound)—reached as high as 67.55%. It is necessary to reduce the indirect costs of inbound and outbound operations to lower the total indirect costs of the warehousing operations center. This also indicates that the application of the method proposed in this paper is feasible in e-commerce enterprises, as it can more objectively and accurately reflect the company's

costs while also adapting to complex business processes.

References

1. Inna, B., Tetiana, Y., Anatolii, B., Oleksandr, K., Liudmyla, Y., & Svitlana, B. (2024). Features of Enterprise Cost Management. In Opportunities and Risks in AI for Business Development: Volume 1 (pp. 699-709). Cham: Springer Nature Switzerland.
2. Korzhengulova, A. A., Serkebayeva, R. K., Aikupesheva, D., Babakhanob, N. E., & Ziyatdinov, A. M. (2017). Management of enterprise costs based on relevant information. Espacios, 38(49), 2.
3. Fan, W. (2025). Research on Cost Management Issues of Small and Medium-Sized Enterprises: A Case Study of Company H. Asia Pacific Economic and Management Review, 2(5).
4. Radionova, N., Skrypnyk, M., & Voronkova, T. (2019). Dual nature of industrial enterprise cost management system. Baltic Journal of Economic Studies, 5(2), 184-190.
5. Liu, H., & Sun, G. (2021). Research on the application of big data technology in the integration of enterprise business and finance. Journal on Big Data, 3(4), 175.
6. Chen, N. (2019, March). Exploration and Application of Industry Financial Integration in Enterprise Financial Management and the Enlightenment--Taking Mengniu Group as an Example. In International Academic Conference on Frontiers in Social Sciences and Management Innovation (IAFSM 2018) (pp. 266-271). Atlantis Press.
7. Fostolovych, V. (2020). Integrated enterprise management model in the post-industrial development system. Norwegian Journal of development of the International Science, (52-3), 3-13.
8. Fagbore, O. O., Ogeawuchi, J. C., Ilori, O., Isibor, N. J., Odetunde, A., & Adekunle, B. I. (2024). Building cross-functional collaboration models between compliance, risk, and business units in finance. International Journal of Scientific Research in Science and Technology, 11(4), 488-524.
9. Guo, M. (2019, April). Research on Financial Business Integration and Financial Management Function Transformation Based on Thermal Power Plants. In IOP Conference Series: Earth and Environmental Science (Vol. 252, No. 3, p. 032050). IOP Publishing.
10. Andreeva, T., Popova, L., Yashina, M., Babynina, L., Yaksanova, N., & Natsypaeva, E. (2019). Integration of the quality management and strategic management systems into unified management system of enterprises. Calitatea, 20(171), 3-8.
11. Hamza, O., Collins, A., Eweje, A., & Babatunde, G. O. (2023). A unified framework for business system analysis and data governance: Integrating Salesforce CRM and Oracle BI for cross-industry applications. International Journal of Multidisciplinary Research and Growth Evaluation, 4(1), 653-667.
12. Vitale, G., Cupertino, S., Rinaldi, L., & Riccaboni, A. (2019). Integrated management approach towards sustainability: an Egyptian business case study. Sustainability, 11(5), 1244.

About the Author

Yaxuan Shi was born in Changsha, Hunan Province, P.R. China, in 2006. She is currently pursuing her undergraduate degree in Business Administration under the Sino-British Joint Program at Chengdu University of Technology. Her primary research interests mainly include market analysis, marketing strategies, and corporate accounting cost research.