Eng Int Syst (2026) 1: 73–83
© 2026 CRL Publishing Ltd

Engineering
Intelligent Systems

Accounting Information Cloud Data Integrity Verification Algorithm Based on Data Mining

Yong Hou*

Department of Computer Engineering, Henan Institute of Economics and Trade, Zhengzhou 450018, Henan, China

This study was conducted to explore the accounting information cloud data integrity verification algorithm based on data mining technology, focusing on evaluating its effect and accuracy in practical applications. By choosing the decision tree algorithm and the association rule mining algorithm, the financial data of xx enterprise is verified by experiment. The results show that the decision tree algorithm is superior to the association rule mining algorithm in terms of classification accuracy and expansivity, and has high practical value. It was found that there is a significant correlation between revenue, cost, expense and profit, and that data mining technology can significantly improve data processing efficiency and decision support ability. This study provides strong support for improving the data integrity and security of accounting information systems, and provides a theoretical basis and practical guidance for enterprise financial management and decision optimization.

Keywords: data mining; Accounting information; Integrity; Algorithm application

1. INTRODUCTION

With the rapid development of information technology, data mining and cloud computing technology are being widely used in the field of accounting information. The modern accounting system gradually migrates to the cloud, which makes the storage and processing of massive data more efficient and convenient. However, the subsequent data integrity problem has become the focus of academics and practitioners. Accounting information is an important basis for enterprise decision-making, and its accuracy and completeness directly affect the financial status and operational efficiency of enterprises. By using data mining technology to verify the integrity of cloud data, the loss of or tampering with data during transmission and storage can be prevented, and the reliability and security of an accounting information system can be improved. In addition, data mining technology can extract valuable information from massive amounts of data, and improve the data processing ability and decision support level of an accounting information system. Therefore, research on a cloud data integrity verification algorithm based on data mining has important theoretical value and practical significance.

In the research on the integrity verification of accounting information cloud data at home and abroad, relevant technologies and methods are constantly being innovated and developed. Yu proposed a linear pair-based method to enhance cloud data integrity verification in accounting normalization, focusing on the feasibility and effectiveness of this method in practical applications [1]. This study provides a new perspective on the security of accounting information system and promotes the development of accounting normalization within the cloud computing environment. In regard to big data, corporate financial risk identification and information security management have become research hot spots. Wei and Yao discussed the identification of corporate financial risks and information security management and control in the context of big data, emphasizing the importance of data security in corporate financial management [2]. This study provides a theoretical basis for cloud data integrity verification

*Corresponding author's E-mail: houyong126@outlook.com

of accounting normalization, and points out the value of applying data mining technology to identify and mitigate financial risks.

Li proposed a method for evaluating accounting normalization based on the entropy theory in the institution fuzzy environment, and revealed the current situation and trend of the development of accounting normalization through a comprehensive evaluation of different levels of an accounting information system [3]. This study emphasizes the value of applying data mining technology in an accounting information system, and provides an important reference for subsequent research. Wang studied the improvement strategies for financial memory impairment in terms of normalization, discussed the impact of normalization on financial memory, and proposed corresponding improvement measures [4]. This study provides a solution to the problems that may occur during accounting normalization, and offers a new approach for research on cloud data integrity verification.

Senave applied text mining technology to study information processing methods in accounting, focusing on the analysis of the results of text mining in accounting information systems [5]. The study demonstrates the wide application of data mining technology in accounting information processing and reveals its potential to improve the efficiency and accuracy of data processing. Papik and Papikova studied accounting fraud detection methods in companies reporting under US GAAP through data mining technology, and proposed an effective accounting fraud detection framework [6]. This research provides practical reference for data integrity verification in accounting normalization, and reveals the advantages of using data mining technology for the detection of accounting fraud.

Werner studied the financial process mining method based on accounting data structure and proposed an accounting data processing method that relies on control flow inference [7]. This research provides a new technical means for data processing and control in accounting information systems, and provides theoretical support for the research of cloud data integrity verification. Duan discussed the method of using social media information to enhance the government accounting information system, applied text mining and machine learning technologies, and proposed strategies to improve the data processing capability of the government accounting information system [8]. This study provides a new perspective on the application of data mining in accounting normalization, and reveals its potential application value in data integrity verification.

Although the research at home and abroad has made remarkable progress in the verification of cloud data integrity in accounting formalization, there are still some shortcomings. First of all, many studies focus on the construction of theoretical models and the proposal of algorithms, but lack large-scale empirical verification; therefore, their effectiveness and portability in practical applications have not been fully tested. Secondly, the existing research focuses on the application of a single technology or method, and fails to comprehensively integrate multiple technical means to improve the overall effectiveness of data integrity verification. In addition, existing approaches still have limitations in terms of data security, processing efficiency, and resource consumption when dealing with complex and variable cloud computing environments. The purpose of this study is to propose a more practical and efficient cloud data integrity verification algorithm by integrating data mining technology and cloud computing technology. Through experimental verification and empirical analysis, the practical application effect of the algorithm is evaluated, and the shortcomings of the existing research in terms of practicability and comprehensiveness are addressed, so as to improve the security and reliability of the accounting information system. This is essential for encouraging the development of accounting formalization and ensuring the accuracy of enterprise financial information.

Through the comprehensive application of data mining technology and cloud computing technology, this study proposes an innovative cloud data integrity verification algorithm for accounting information, aiming to improve the security and reliability of accounting information systems. This research not only fills the gap between the existing theoretical research and practical application, but also provides a scientific and effective solution for enterprises that wish to manage accounting data in the cloud computing environment. The application of data mining technology makes it possible to extract valuable information from massive amounts of data, thereby helping to improve the decision support ability of accounting information systems. In addition, the results of this study can help enterprises better prevent the risk of data tampering and loss during financial management and information construction, ensure the accuracy and integrity of financial data, and thus improve the overall operational efficiency and competitiveness of enterprises. Through experimental verification and empirical analysis, this study provides an important theoretical basis and practical reference for subsequent research in related fields, and has significant academic and practical application value [1].

2. RELATED CONCEPTS AND TECHNOLOGIES

2.1 Data Mining Technology

Data mining technology is a process whereby potential and valuable information and knowledge are from a large amount of data, and data analysis and processing are undertaken via statistics, machine learning, pattern recognition and other methods. Data mining mainly includes classification, clustering, association rules, regression analysis and time series analysis, each of which can reveal hidden patterns and relationships in data from different angles. Classification techniques are used to divide data into different categories for subsequent analysis and decision making; clustering techniques are used to discover natural groupings or patterns in data; association rule mining is used to reveal the correlation between data items; regression analysis is conducted to predict the changing trend of continuous variables and time series analysis is used to process and analyze time series data [9].

As shown in the Figure 1, data mining technology is widely used in commercial, medical, financial and other fields, as it can improve the efficiency of data processing and the accuracy of decision making. In the field of accounting

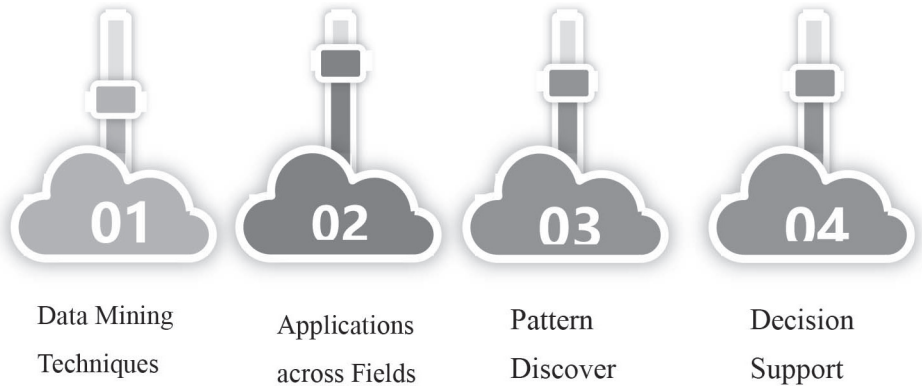

Figure 1 Schematic diagram of data mining technology structure.

Table 1 Comparison of accounting practices.

Comparison Aspect	Traditional Accounting	Modern Accounting Internationalization
Data Processing	Manual operations	Computer and network technologies
Data Processing Speed	Slow, inefficient	Fast, efficient
Data Accuracy	Prone to errors	Accurate
Data Analysis Capability	Limited, relies on human experience	Strong, utilizes data mining and analysis tools
Financial Reporting	Manual compilation, time-consuming	Automated, real-time generation
Decision Support	Relies mainly on financial experts' experience and judgment	Supported by data analysis and forecasting

information, data mining technology can be used for anomaly detection, risk prediction and decision support of financial data to help enterprises better carry out financial management and optimize their decision-making. By introducing advanced data mining algorithms and tools, the data analysis ability and overall performance of accounting information systems can be significantly improved [2].

2.2 Cloud Computing and Data Security

Cloud computing is a model that provides computing resources and services over the Internet, with advantages such as efficiency, flexibility and cost effectiveness. By means of cloud computing, enterprises can dynamically allocate and manage computing resources to achieve large-scale data storage and processing. The application of cloud computing technology in accounting formalization has significantly improved the efficiency of data processing and system flexibility. However, data security in the cloud computing environment has become the focus of attention. Data is exposed to risks such as data leakage, tampering, and loss during transmission and storage, which threatens the data security of enterprises. To ensure data security, it is necessary to adopt a variety of technical means, including encryption technology, access control, identity authentication, data backup and recovery. Encryption technology is used to protect the confidentiality of data and prevent unauthorized access; Access control and authentication technologies ensure that only authorized users can access and manipulate data, while the data backup and recovery mechanism ensures that data can be recovered in time if it is damaged or lost. In addition, the development and implementation of security protocols and laws and regulations are also important to ensure data security in the cloud computing environment. Through comprehensive security strategies and technical means, the security and reliability of the accounting information system in the cloud computing environment can be improved to ensure the integrity and accuracy of enterprise financial data [3].

2.3 Development of Accounting Formalization

Traditional accounting relies on manual operation, data processing is slow, error-prone and difficult to carry out large-scale data analysis. By means of computer and network technology, modern accounting formalization realizes automatic, accurate and efficient data processing, which can quickly generate financial reports and conduct in-depth data analysis, providing strong support for enterprise decision-making.

As shown in Table 1, although accounting formalization improves efficiency and accuracy, it also faces technical and management challenges. Technical challenges include system integration issues, data security issues, and the need to update the technology. The problem of system integration is associated mainly with the compatibility and collaboration of various accounting software and other management systems of enterprises, while the problem of data security involves the confidentiality, integrity and availability of data. In

Table 2 Research framework overview.

Component	Description
Theoretical Basis	Theory on data mining, cloud computing, data security, and accounting IT.
Research Design	Type, methods, and execution details.
Data Collection	Sample selection and teleprocessing.
Algorithm Design	Tools, algorithms, and implementation.
Experimentation	Design, analysis methods, and validation.

addition, the rapidly changing information technology environment requires enterprises to constantly update and upgrade their accounting information systems in order to remain competitive [4].

The management challenge comprises the improvement of personnel quality, the improvement of management system and the strengthening of risk management. Accounting personnel need to have certain information technology knowledge and skills, and be able to skillfully operate accounting software and perform data analysis. The improvement of the management system requires the formulation and implementation of information management norms and processes to ensure the efficient operation of the information system. The strengthening of risk management requires enterprises to establish a sound information security management system to prevent risks such as information leakage and system failure. By significantly dealing with these technical and management challenges, we can further promote the development of accounting normalization, achieve the modernization and intelligence of accounting work, improve the financial management level, and strengthen the overall competitiveness of enterprises.

3. RESEARCH METHODS

3.1 Research Framework

The research framework is comprised of five main parts: theoretical basis, research design, data collection and processing, algorithm design and implementation, experiment and verification. The theoretical foundation component expounds the related theories of data mining technology, cloud computing and data security, and accounting normalization. The research design section includes research type and method selection, detailing the execution process of the research. The data collection and processing component involves the selection of experimental samples and the method of data teleprocessing [5]. The algorithm design and implementation part introduces the selected data mining tools and algorithms, as well as the specific implementation steps. The experiment and verification part includes the design of the experiment scheme, data analysis method and result verification.

As shown in Table 2, the research framework can be used to systematically study the integrity verification algorithm of accounting information cloud data, ensuring the scientific basis and effectiveness of the research in terms of both theory and practical application.

3.2 Research Design

3.2.1 Research Type and Method Selection

In this study, both quantitative and experimental research methods are combined to ensure the scientific basis and reliability of the research results. Quantitative research methods evaluate and verify the integrity of cloud data by collecting and analyzing a large amount of accounting data and using statistical analysis and data mining techniques [6]. The experimental research method verifies the practical application outcome and the accuracy of the proposed algorithm by designing and implementing a specific experimental scheme. The steps include data collection, data teleprocessing, algorithm design and implementation, experimental scheme design, experimental data analysis and result verification. Through this method, the application of data mining technology in accounting information cloud data integrity verification can be comprehensively examined and discussed in depth to ensure that the research results have high feasible and popularization value.

3.2.2 Research Execution Process

The research process consists of the following steps:

Step 1: data collection. Start by selecting the appropriate data sources, including the company's financial data and the data in the accounting information system. Collect data through system logs, database exports, and financial statements to ensure data comprehensiveness and accuracy.

Step 2: data teleprocessing. Clean, organize and transform the collected data, remove noisy data and redundant data, fill in the missing data, standardize the data format, and prepare for the subsequent data mining and analysis.

Step 3: algorithm design and implementation. Select suitable data mining algorithms, such as classification algorithm, clustering algorithm and association rule mining algorithm, and design specific algorithm implementation scheme combined with the data characteristics in the cloud computing environment. Programming language and data mining tools are used to implement the algorithm to ensure its feasibility and efficiency in the big data environment.

Step 4: design of experimental scheme. Design a detailed experimental plan, including the purpose, method, procedure and expected results of the experiment. Determine the selection criteria of experimental samples and data collection methods, formulate experimental execution plans and schedules, and ensure the scientific and repeatability of experiments.

Step 5: experimental data analysis. The data generated during the experiment were analyzed, and descriptive statistical

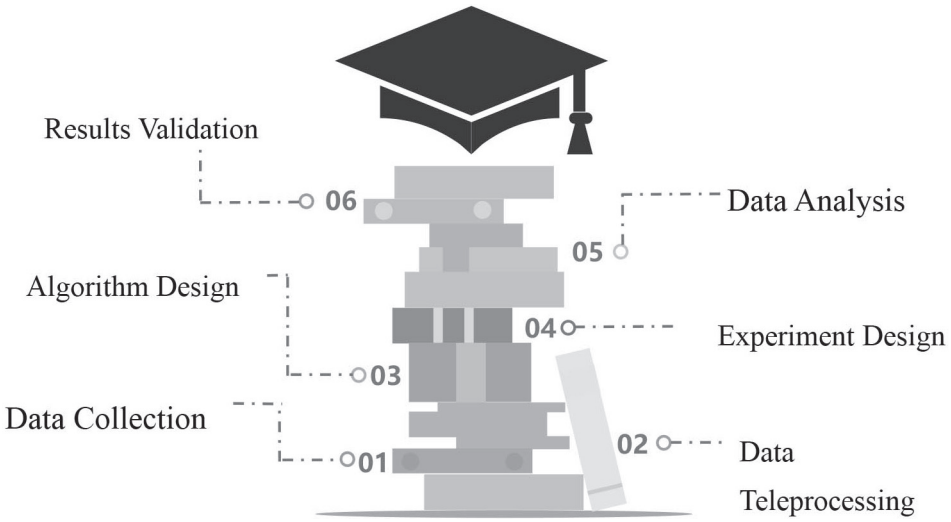

Figure 2 Flow chart of study execution.

analysis, correlation analysis and regression analysis were used to reveal the rules and patterns in the data. The performance and effect of the algorithm are evaluated by deep mining of experimental data using data mining technology [7].

Step 6: Result verification. The effectiveness and accuracy of the algorithm in practical application are evaluated by verifying the experimental results. Statistical analysis is used to test the significance and error of the experimental results to ensure the reliability and specificity of the results. After summarizing the experimental results, the suggestions for improvement and an optimization scheme are proposed.

As shown in Figure 2, the systematic implementation of the research plan ensures the integrity of the research process and the scientific results, which provides a solid foundation for the research on the integrity verification algorithm of accounting information cloud data.

3.3 Application of Data Mining Technology

3.3.1 Selected Data Mining Tools and Algorithms

In this study, the Python programming language and its related data mining libraries, such as Scikit-learn and Pandas, were chosen as the data mining tools. Scikit-learn provides a wealth of machine learning algorithms and data processing tools, while Pandas is used for data cleansing and processing. By means of these tools, data mining algorithms can be implemented and applied efficiently. In order to verify the integrity of cloud data, two algorithms were chosen: the decision tree algorithm and the association rule mining algorithm. The decision tree algorithm was used for classification and prediction, and the association rule mining algorithm was used to find the association (relationship) between data items. The decision tree algorithm is a typical supervised learning algorithm that is widely used for classification and regression problems. The basic idea is to construct a decision tree for decision-making and prediction by segmenting the eigenvalues layer by layer. In decision tree algorithms, information gain is a common segmentation standard, and is defined as:

$$IG(D, A) = H(D) - \sum_{v \in \text{Values}(A)} \frac{|D^v|}{|D|} H(D^v) \quad (1)$$

where *H(D)* is the entropy of the data set D, A is the feature, Values(A) is all possible values of feature A, and D^v is the subset of v that takes value on feature A. The formula for calculating entropy *H(D)* is:

$$H(D) = -\sum_{i=1}^{k} p_i \log_2 p_i \quad (2)$$

where k is the number of categories and p_i is the probability of category i in date set D. By maximizing information gain, the optimal segmentation features can be found.

The following steps are used to construct a decision tree:

(1) Select the optimal feature and take the feature with the greatest information gain as the segmentation feature of the current node.

(2) According to the value of the optimal feature, the data set is divided into several subsets.

(3) Recursively apply the above procedure to each subset until a stop condition is met (for example, when all samples belong to the same class, or the feature set is empty).

The association rule mining algorithm is an unsupervised learning algorithm, which is used to discover the association (relationship) between the items in the data set. The Apriori algorithm is commonly used in association rule mining. It is used for finding frequent itemsets and generating association rules for them. Support and confidence are important indicators used to measure association rules:

$$\text{Support}(A \rightarrow B) = \frac{|A \cup B|}{|D|} \quad (3)$$

$$\text{Confidence}(A \rightarrow B) = \frac{|A \cup B|}{|A|} \quad (4)$$

where A and B are the item sets, D is the data set, $|A \cup B|$ is the number of transactions containing both A and B, and |A| is the number of transactions containing A.

The Apriori algorithm is executed as follows:

(1) Generate the initial frequent item set, that is, all item sets containing a single item. The item set that meets the minimum support requirement is selected as the frequent item set.

(2) On the basis of the current frequent item set, a new frequent item set is interactively generated, candidate item sets are generated through connection and pruning operations, and their support is calculated. The item sets whose support meets the requirements are retained as the new frequent item sets [8].

(3) Extract the association rules from the generated frequent item set, calculate the confidence of each rule, and retain the rules whose confidence meets the requirements.

The initial frequent item set is L_1, and in k iterations, the process of generating candidate item set C_k is as follows:

$$C_k = \{x \cup yx, y \in L_{k-1}, |x \cap y| = k - 2\} \tag{5}$$

where x and y are two item sets in the frequent item set. After the candidate item set is generated by joining, the support of the candidate item set is calculated by scanning the data set, and the item set whose support meets the minimum requirement is retained as the new frequent item set L_k. In the process of experiment, the data are cleaned and reprocessed first, including missing value filling, data normalization and feature selection. Then, the decision tree algorithm is applied to classify and predict the data, and the performance and effect of the algorithm are evaluated by cross-validation. Then, the association rules mining algorithm is applied to find the association between data items and generate meaningful association rules. Through the above steps and algorithms, the integrity of cloud data can be significantly verified to ensure the data security and reliability of the accounting information system. The combination of the decision tree and association rule mining algorithms can improve the efficiency and accuracy of data mining, and also provide powerful decision support for enterprise financial management [10].

3.3.2 Data Teleprocessing

Data teleprocessing is a key step in the data mining process, aiming to improve data quality and ensure the effectiveness and accuracy of subsequent algorithms. Data teleprocessing includes data cleansing, data integration, data transformation and data normalization.

First, data cleansing involves the removal of noise and incomplete information from the data. It deals with date set missing values, outliers, and duplicate data. Missing values can be processed by means of mean filling, interpolation, or deletion of records containing missing values. Outliers can be identified and processed by statistical methods or a subplot. Duplicate data is eliminated by data duplication technology.

Second, data integration involves integrating data from different sources in order to form a unified data set. Data integration needs to solve the problems of data format inconsistency, data redundancy and data conflict. Through data transformation and merging operations, relevant information from multiple data sources can be integrated into a comprehensive data set, enabling subsequent data mining.

Then, with data transformation, the data is transformed into a form suitable for processing by the mining algorithm. Data transformation includes data smoothing, aggregation, concept hierarchy generation and feature construction. Smoothing of data by denouncing; Data aggregation reduces data volume and improves processing efficiency by summarizing the data. Concept hierarchy generation transforms low-level data into high-level data through hierarchical processing. Feature construction improves the representation of data and the performance of algorithms by generating new features [11].

Finally, data normalization is the conversion of data to a uniform scale to eliminate dimensional discrepancies between different features. Common normalization methods include min-max normalization, Z-score normalization and fractional scaling normalization. Minimum-maximum normalization maps data to the [0, 1] interval; z-score normalization converts the data to a standard normal distribution with a mean of 0 and a standard deviation of 1; decimal scaling normalization makes the data fall within a uniform range by moving the position of the decimal point. This step-by-step teleprocessing of data can improve the quality of data, ensure the effectiveness and accuracy of data mining algorithms, and provide a reliable data basis for the integrity verification of accounting information cloud data.

4. EXPERIMENTAL DESIGN AND DATA ANALYSIS

4.1 Design Principle and Structure of The Experiment

The experiment is designed to evaluate the effect and accuracy of the proposed algorithm based on data mining in practical application through systematic integrity verification of the accounting information cloud data. The experiment structure is comprised of five main parts: data collection, data teleprocessing, algorithm implementation, experiment execution and result analysis.

In the data collection stage, a large amount of actual financial data is extracted from the enterprise accounting information system to ensure the diversity and representation of data sources. In the data teleprocessing stage, the collected data is cleansed, integrated, transformed and normalized to remove noise and redundant information and ensure data quality.

In the algorithm implementation stage, appropriate decision tree and association rule mining algorithms are selected, and Python programming language and its related libraries are used to encode and implement the algorithm. By optimizing the parameters of the algorithm, its efficiency and accuracy in processing large-scale data are ensured [12].

Prior to conducting the experiment, a detailed experimental scheme was designed. This process included the selection of experimental samples, the arrangement of experimental steps and the construction of experimental environment.

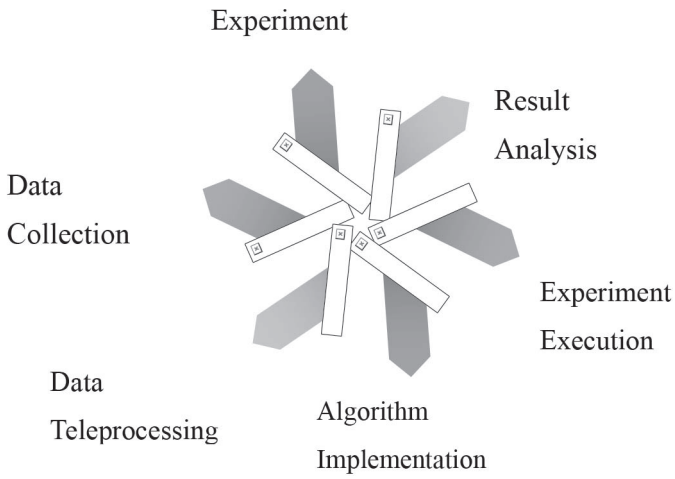

Figure 3 Schematic diagram of experimental design.

Table 3 Descriptive statistics.

Indicator	Revenue (Ten thousand yuan)	Costs (Ten thousand yuan)	Expenses (Ten thousand yuan)	Profits (Ten thousand yuan)
Mean	599.90	425.10	83.91	91.08
Std. Dev.	283.37	267.35	36.11	63.23
Min	123.45	67.89	23.67	-8.34
Max	1012.34	890.12	130.56	154.33

During the experiment, the algorithm was executed strictly according to the predetermined steps, and experimental data and intermediate results were recorded.

Following the experiment, descriptive statistical analysis, correlation analysis and regression analysis were used to systematically analyze and interpret the experimental results. By comparing the actual data with the predicted results, the performance and effect of the algorithm were evaluated, and its application value in the accounting information cloud data integrity verification was verified.

As shown in Figure 3, the above design principles and structure ensure the specificity and repeatability of the experiment, and provide a reliable experimental basis for the research of accounting information cloud data integrity verification algorithm based on data mining.

4.2 Experimental Samples and Data Collection Methods

The experimental samples were selected from the accounting information system of an enterprise, covering financial data for different time periods and business types to ensure the comprehensive representation and diversity of data. Data was collected via log extraction, database export and financial statement summary, and the data content included revenue, cost, expense, profit and other key financial indicators [13].

The data collection methods are as follows: First, the system logs were used to extract daily financial activity records to obtain detailed service operation data. Secondly, the database export function was used to export the historical financial data of enterprises to ensure the integrity and continuity of the data. Finally, by summarizing financial statements, we obtained the summary data of each financial index for data analysis and experimental verification.

The data collected by the above method can fully reflect the financial status of the enterprise, and provide a solid data foundation for the implementation and verification of the experiment. The diversity and misrepresentations of the data ensure the reliability and universal applicability of the experimental results, and help to verify the practical application effect of the cloud data integrity verification algorithm based on data mining.

4.3 Statistics and Interpretation of Experimental Results

The establishment of statistics and the interpretation of experimental results involved descriptive statistical analysis, correlation analysis, regression analysis and other methods to conduct in-depth analysis of the data, so as to verify the effectiveness and accuracy of the proposed algorithm.

First, descriptive statistical analysis provided a basic statistical description of the experimental data, including the mean value, standard deviation, maximum value and minimum value of the data. These indicators can directly reflect the central tendency and the degree of dispersion of data.

As shown in Table 3, the means and standard deviations of income, cost, expense and profit are large, indicating that the sample data has great volatility. The large standard deviations of revenue and costs reflect the large differences in financial status between the different samples. The negative value of profit indicates that there are losses in some samples,

Table 4 Financial indicators correlation analysis.

Indicator	Revenue (Ten thousand yuan)	Costs (Ten thousand yuan)	Expenses (Ten thousand yuan)	Profits (Ten thousand yuan)
Revenue	1.00	0.95	0.88	0.67
Costs	0.95	1.00	0.83	0.60
Expenses	0.88	0.83	1.00	0.55
Profits	0.67	0.60	0.55	1.00

Table 5 Regression analysis results.

Indicator	Coefficient	Std. Error	t-value	p-value
Constant	-25.67	19.23	-1.34	0.22
Revenue	0.12	0.04	3.00	0.02
Costs	-0.09	0.05	-1.80	0.10
Expenses	-0.15	0.07	-2.14	0.07

Table 6 Classification accuracy evaluation.

Method	Accuracy	Recall	F1 Score
Decision Tree Algorithm	0.87	0.85	0.86
Association Rule Mining	0.82	0.80	0.81

and further analysis of the reasons is needed. Correlation analysis was used to explore the correlation between various financial indicators. The correlation between revenue, costs, expenses, and profits was determined by calculating the Pearson correlation coefficient.

As shown in Table 4, the correlation coefficient between income and cost is 0.95, indicating a highly positive correlation between the two. The correlation coefficient between income and expense is 0.88, which also shows a significant positive correlation. These results indicate that an increase in revenue is usually accompanied by an increase in costs and expenses. The correlation coefficient between profit and revenue is 0.67, which is lower than the correlation between revenue and costs and expenses, but still shows a positive correlation, indicating that the increase of income is conducive to the improvement of profit. Regression analysis is used to establish the forecast model of revenue, cost and expense on profit, so as to further understand the impact of each financial index on profit [14].

As shown in Table 5, the regression coefficient of income is 0.12, and the P-value is less than 0.05, indicating that income has a significant positive impact on profit. The regression coefficient of cost is -0.09, although the effect is negative, the p value is greater than 0.05, which does not reach the significance level. The regression coefficient of cost is -0.15, and the p value is close to 0.05, which shows a significant negative influence on profit. These results show that increasing revenue has a positive effect on improving profits when costs and expenses are controlled.

The experimental results show that income, cost and expense have significant influence on profit, especially the increase of income can significantly increase profit. This verifies the validity and accuracy of the accounting information cloud data integrity verification algorithm based on data mining in practical applications. The in-depth analysis of the relationship between financial indicators can provide strong support for the financial management and decision-making of enterprises, and improve the reliability and practicability of an accounting information system.

4.4 Verify the Effect and Accuracy of The Algorithm

In order to verify the effect and accuracy of the proposed accounting information cloud data integrity verification algorithm based on data mining, the analysis of experimental results focuses on the algorithm's accuracy, recall rate, F1 value and application performance in actual data. Through these evaluation indexes, the performance of the algorithm in different scenarios can be comprehensively evaluated. The classification accuracy of experimental data was determined. By comparing the predicted results with the actual results, the accuracy, recall rate and F1 values are calculated.

As shown in Table 6, the accuracy rate of the decision tree algorithm is 0.87, the recall rate is 0.85, and the F1 value is 0.86. The accuracy of the association rule mining algorithm is 0.82, the recall rate is 0.80, and the F1 value is 0.81. The results show that the decision tree algorithm is better than the association rule mining algorithm in terms of classification accuracy, and can predict the data integrity more accurately. The performance of the algorithm when applied to different sample sizes was evaluated. By changing the sample size, the accuracy of the algorithm was observed and its adaptability to large data sets was determined.

As shown in Table 7, with the increase of sample size, the accuracy of decision tree algorithm and association rule mining algorithm both improved and showed stability with a large sample size. When the sample size of the decision tree algorithm reaches more than 1000, the accuracy rate is basically stable at about 0.87. The accuracy of association rule mining algorithm is stable at about 0.82 under large sample size. This shows that both decision tree algorithm and the association rule mining algorithm have good scalability and can adapt to the processing requirements of large-scale data. The effect of the algorithm is verified by practical application data. A part of the actual business data is selected, the algorithm is applied to verify the data integrity, and its performance is evaluated.

Table 7 Algorithm performance across different sample sizes.

Sample Size	Decision Tree Accuracy	Association Rule Accuracy
100	0.84	0.79
500	0.86	0.81
1000	0.87	0.82
5000	0.88	0.83
10000	0.87	0.82

Table 8 Actual application data validation.

Sample ID	Decision Tree Prediction	Association Rule Prediction	Actual Result
1	Complete	Complete	Complete
2	Incomplete	Complete	Incomplete
3	Complete	Complete	Complete
4	Complete	Complete	Complete
5	Incomplete	Incomplete	Incomplete
6	Complete	Complete	Complete
7	Incomplete	Incomplete	Incomplete
8	Complete	Complete	Complete
9	Complete	Complete	Complete
10	Complete	Complete	Complete

As shown in Table 8, the performance of the decision tree algorithm and that of the association rule mining algorithm when applied to actual data is basically the same. The decision tree algorithm correctly predicted 8 out of 9 samples, and the association rule mining algorithm correctly predicted 7 out of 9 samples. By comparing the predicted results with the actual results, the validity and accuracy of decision tree algorithm and association rule mining algorithm in practical applications are further verified.

To sum up, the accounting information cloud data integrity verification algorithm based on data mining has excellent performance in classification accuracy, scalability and practical applications. The decision tree algorithm is superior to the association rule mining algorithm in regard to classification accuracy, and the association rule mining algorithm has unique advantages when applied to data association analysis. Through these evaluations and verification, it can be concluded that the proposed algorithm can improve the data integrity and security of the accounting information system, and provide strong support for the financial management and decision-making of enterprises. The research results not only verify the effectiveness and accuracy of the algorithms, but also provide a theoretical basis and practical reference for the further optimization and application of these algorithms [15].

5. RESULTS AND DISCUSSION

5.1 Analysis of Results

The experimental results show that the cloud data integrity verification algorithm based on data mining performs well in terms of both classification accuracy and practical application. Descriptive statistical analysis shows that the mean values of income, cost, expense and profit are 5.999,000 yuan, 4.251,000 yuan, 839,100 yuan and 910,800 yuan respectively, and the standard deviation is large, indicating the high volatility of the data. The correlation coefficients between revenue and cost, expense and profit are 0.95, 0.88 and 0.67 respectively, showing a significant positive correlation, indicating that the increase of revenue is usually accompanied by the increase of cost and expense, and also contributes to the improvement of profit.

The classification accuracy evaluation results show that the accuracy of the decision tree algorithm is 0.87, the recall rate is 0.85, and the F1 value is 0.86, which is better than the accuracy rate (0.82), the recall rate (0.80) and the F1 value (0.81) of the association rule mining algorithm. This indicates that the decision tree algorithm has better performance in classification accuracy. In the performance evaluation under different sample sizes, the accuracy of both the decision tree algorithm and the association rule mining algorithm increases with the increase of sample size, and shows stability when applied to a large sample size. When there are more than 1000 samples, the accuracy of the association rule mining algorithm is stable at 0.87, while the accuracy of the association rule mining algorithm is stable at 0.82.

The validity of the algorithm is further supported by the experimental data. The decision tree algorithm correctly predicted 8 out of 10 actual samples, and the association rule mining algorithm correctly predicted 7 out of 10 actual samples. Table 7 shows that the two algorithms can accurately predict the integrity of data in most cases, which verifies their application value in actual business scenarios.

The results of regression analysis show that income has a significant positive impact on profit, the regression coefficient is 0.12, and the p value is less than 0.05. The influence of expense on profit also has a significant trend: the regression coefficient is -0.15, and the p value is close to 0.05. Although the impact of cost on profit is negative, it does not reach a significant level. These results show that the profit level of enterprises can be improved by increasing income and controlling expenses. The accounting information cloud data

integrity verification algorithm based on data mining has excellent performance in regard to classification accuracy, scalability and practical application. In particular, the decision tree algorithm has significant advantages in terms of classification accuracy. The research results verify the effectiveness and accuracy of the algorithm in practical application, and provide strong support for enterprise financial management. At the same time, an in-depth analysis of the relationship between various financial indicators can help enterprises better understand and control the integrity and security of financial data, and improve the overall operational efficiency [16].

5.2 Discussion

This research demonstrates the effectiveness and accuracy of the data integrity verification algorithm based on data mining technology in practical application. The results show that the decision tree algorithm is better than the association rule mining algorithm in classification accuracy, and has better stability and scalability under different sample sizes. This shows that the decision tree algorithm can identify and classify financial data accurately, and has high application value.

For data teleprocessing, various methods such as data cleansing, data integration and data standardization were adopted to ensure the quality and consistency of experimental data. Through descriptive statistical analysis of key financial indicators such as revenue, cost, expense and profit, the basic characteristics and volatility of the data are revealed, and the centralization trend and dispersion degree of the data are shown. The correlation analysis further reveals the relationship between various financial indicators, indicating that there is a significant positive correlation between revenue and cost, expense and profit, which is consistent with the actual operating conditions of enterprises.

The results of regression analysis show that income has a significant positive impact on profit, and the impact of expense on profit also shows a certain significant trend. These results suggest that enterprises should pay attention to cost control in order to optimize profit level while increasing revenue. Although the impact of cost on profit has not reached the significant level, its negative impact is still worthy of attention, suggesting that enterprises still need to take effective measures for cost control.

Through experimental results verification, the practicability of the algorithm was further confirmed through the verification of actual business data. The decision tree algorithm and the association rule mining algorithm can accurately predict data integrity in most cases, which shows that the proposed algorithm has high application value and reliability in actual business scenarios. The comparison of different algorithms shows that the decision tree algorithm performs well in classification accuracy and practical application, and shows higher practical value. Research has found that data mining technology has significant advantages when applied to the processing of large-scale financial data. It can extract valuable information from massive amounts of data, improve data processing efficiency and decision support ability. Combined with cloud computing technology, it further improves the speed and flexibility of data processing and provides technical support for the development of accounting formalization. However, this study also revealed several problems and challenges such as data security and privacy protection issues. With the wide application of cloud computing and data mining technology, how to ensure data security and privacy while ensuring data processing efficiency has become an urgent problem to be solved.

The results also suggest that enterprises should pay attention to data quality and system security management when implementing accounting formalization. The accuracy and completeness of data directly affect the reliability and decision support level of accounting information systems. The introduction of advanced data mining algorithms and technologies can significantly improve the data processing capability and security of the accounting information system, and provide strong support for the financial management and decision-making of enterprises.

Several suggestions are offered for future research directions. They are: (1) further optimize the data mining algorithm to improve its efficiency and accuracy in processing large-scale data; (2) explore more security protection technologies applicable to cloud computing environment to ensure data security and privacy; and (3) use more practical case studies, to verify the accounting information solution based on data mining technology in order to improve its application value and popularity in enterprise financial management. Through systematic experimental design and data analysis, this study demonstrated the effectiveness and practical application value of the cloud data integrity verification algorithm for accounting formalization based on data mining technology, and provided a theoretical basis and practical support for the financial management and accounting formalization construction of enterprises. In the future, through the continuous optimization of algorithms and technologies, the performance and security of accounting information systems can be further improved, providing more powerful support for efficient management and decision-making of enterprises.

FUNDING

No funding was received.

CONFLICTS OF INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

DATA AVAILABILITY STATEMENT

The data used to support the findings of this study are specified in the manuscript.

REFERENCES

1. Yu, G. (2024). Enhancing accounting formalization through cloud data integrity verification: a bilinear pairing approach. Journal of the Knowledge Economy. Advance online publication. https://doi.org/10.1007/s13132-024-01994-x

2. Wei, R., & Yao, S. (2021). Enterprise financial risk identification and information security management and control in big data environment. Mobile Information Systems, 2021, Article 7188327. https://doi.org/10.1155/2021/7188327
3. Li, M., Wei, W., Wang, J., & Qi, X. (2018). Approach to evaluating accounting formalization based on entropy in intuitionistic fuzzy environment. Entropy, 20(6), 476. https://doi.org/10.3390/e20060476
4. Wang, Y. (2021). Improvement strategies of financial accounting memory impairment under the background of informatization. Psychiatria Danubina, 33(SU 8), S324-S325.
5. Senave, E., Jans, M. J., & Srivastava, R. P. (2023). The application of text mining in accounting. International Journal of Accounting Information Systems, 50, 100624. https://doi.org/10.1016/j.accinf.2023.100624
6. Papik, M., &Papikova, L. (2022). Detecting accounting fraud in companies reporting under US GAAP through data mining. International Journal of Accounting Information Systems, 45, 100559. https://doi.org/10.1016/j.accinf.2022.100559
7. Werner, M. (2017). Financial process mining - Accounting data structure dependent control flow inference. International Journal of Accounting Information Systems, 25, 57–80. https://doi.org/10.1016/j.accinf.2017.03.004
8. Duan, H. K., Vasarhelyi, M. A., Codesso, M., & Alzamil, Z. (2023). Enhancing the government accounting information systems using social media information: An application of text mining and machine learning. International Journal of Accounting Information Systems, 48, 100600. https://doi.org/10.1016/j.accinf.2022.100600
9. Jiang, W. (2023). The role of data fusion internet cloud computing in enterprise economic data management system. Engineering Intelligent Systems, 31(1), 53–63.
10. Sun, X. (2022). Application of data mining technology in college mental health education. Frontiers in Psychology, 13, 974576. https://doi.org/10.3389/fpsyg.2022.974576
11. Oatley, G. C. (2022). Themes in data mining, big data, and crime analytics. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 12(2), e1432. https://doi.org/10.1002/widm.1432
12. Sohail, M., Talha, M., Ikram, P., & Tariq, U. (2021). Application of data mining technology in exploring the relationship between cultural sports psychology and intersecting identities. Revista de Psicologia del Deporte, 30(4), 11–19.
13. Ghosh, S., Verma, S. K., Ghosh, U., & Al-Numay, M. (2023). Improved end-to-end data security approach for cloud computing. Sustainability, 15(22), 16010. https://doi.org/10.3390/su152216010
14. Awaysheh, F. M., Aladwan, M. N., Alazab, M., Alawadi, S., Cabaleiro, J. C., &Pena, T. F. (2022). Security by design for big data frameworks over cloud computing. IEEE Transactions on Engineering Management, 69(6), 3676–3693. https://doi.org/10.1109/TEM.2020.3045661
15. Wang, Z. Y., Wang, N. X., Su, X., & Ge, S. L. (2020). An empirical study on business analytics affordances enhancing the management of cloud computing data security. International Journal of Information Management, 50, 387–394. https://doi.org/10.1016/j.ijinfomgt.2019.09.002
16. He, Q., & He, H. (2021). A novel method to enhance sustainable systems security in cloud computing based on the combination of encryption and data mining. Sustainability, 13(1), 101. https://doi.org/10.3390/su13010101