Implementing Big Data Auditing: Pathways and Institutional Safeguards for Effective Adoption

Wen Qi *

School of Economics and Management, NingXia University, Yinchuan 750021, China

* Corresponding Author Email: qiw7988@gmail.com

Abstract. In the digital era, big data auditing has gradually emerged as a crucial approach in audit work, boasting its distinct advantages. This paper delves into the implementation pathways and safeguard mechanisms that facilitate the advancement of big data auditing. Initially, it elaborates on the theoretical essence of big data auditing and its alignment with the theory of comprehensive audit coverage. Subsequently, through empirical investigation, it demonstrates the application effects of big data analysis technology in auditing, encompassing trend analysis, cluster analysis, anomaly detection, correlation analysis, and machine learning prediction technology. The research findings indicate that big data auditing can effectively enhance the ability to identify problems in auditing and achieve comprehensive audit coverage. In terms of implementation pathways, suggestions are put forward, such as strengthening data governance, improving technical capabilities, optimizing audit procedures, and establishing coordination mechanisms. Regarding safeguard mechanisms, emphasis is placed on the significance of organizational guarantees, system guarantees, talent guarantees, and security guarantees. Despite the fact that big data auditing still encounters numerous challenges in practical applications, it plays a vital role in promoting the healthy development of the economy and society and safeguarding national economic security. In the future, efforts should be made to further improve the implementation pathways and safeguard mechanisms to promote the sustainable development of big data auditing.

Keywords: Big Data Auditing; Implementation Path; Safeguard Mechanism; Data Analysis Technology.

1. Introduction

In the digital era, the rapid development of information technology has profoundly changed the operation modes of various industries, and the auditing field is no exception. As an advanced direction of traditional auditing informatization, big data auditing is gradually becoming an important means of auditing work with its unique advantages [1]. Big data auditing uses huge, scattered, and diverse data of economic and social operation to carry out in-depth data mining and analysis across levels, regions, systems, departments, and businesses, effectively improving the ability of auditing to find problems, evaluate and judge, and conduct macro analysis [2].

Compared with the traditional auditing mode [3], big data auditing has realized the transformation from sampling auditing to full-sample auditing, avoiding sampling risks and making auditing results more comprehensive and accurate; it has expanded from on-site auditing to off-site auditing, breaking through time and space constraints and greatly improving auditing efficiency; it has shifted from focusing on causal relationships to focusing on correlation relationships, breaking through the limitations of traditional auditing perspectives, providing a new perspective for auditing work, and significantly improving auditing efficiency. These transformative advantages of big data auditing make it play an increasingly important role in promoting strengthened management, improving performance, and maintaining national economic security [4].

At present, various departments and fields are actively promoting informatization construction, and the characteristics of business dataization and management informatization of auditing objects are becoming more and more significant. This provides a broad space for the development of big data auditing, but also brings many challenges. How to effectively promote the implementation of big data auditing and build a sound implementation path and safeguard mechanism has become a key issue to be solved urgently in the auditing field [2,4]. In-depth research on this topic is of great practical

significance for giving full play to the effectiveness of big data auditing, improving the quality of auditing work, and promoting the modernization of the national governance system and governance capacity.

2. Theoretical Foundations and Research Methods

2.1 Overview of Big Data Auditing Theory

Big data auditing is an independent supervision activity carried out by auditing institutions using the technologies, methods, and tools of auditing science and big data science to check the reliability of major financial matters, operation and management activities, and related materials of the audited units. Its core lies in using big data technology to process and analyze massive amounts of data to obtain valuable auditing information. Big data auditing theory integrates multidisciplinary knowledge such as auditing, computer science, and statistics, providing a solid theoretical support for the innovative development of auditing work.

From the perspective of auditing, big data auditing follows the basic principles of auditing, such as independence, objectivity, and fairness. Based on risk-oriented auditing, it determines the key points and scope of auditing through risk assessment of the audited units. At the same time, big data auditing expands the scope and acquisition methods of audit evidence, including not only traditional paper vouchers and electronic data but also new types of data such as social media data and sensor data, making audit evidence more abundant and comprehensive.

From the perspective of computer science, big data auditing relies on technologies such as distributed computing, cloud computing, and data warehouses to realize the storage, processing, and analysis of massive data. Distributed computing technology can distribute large-scale data processing tasks to multiple computing nodes for parallel processing, improving data processing efficiency; cloud computing technology provides elastic computing and storage resources for big data auditing, reducing the hardware investment and maintenance costs of auditing institutions; data warehouse technology is used to integrate and manage data from different sources and in different formats, providing a unified data platform for data analysis.

2.2 Application Principles of Big Data Analysis Technology in Auditing

Combined with the technical methods involved in the analysis of the research results, the following is a detailed discussion on the application principles of relevant big data analysis technologies in auditing [5,6].

(1) Trend Analysis Technology

Trend analysis technology is a data analysis method that extracts and analyzes the characteristics of data sequences changing over time to reveal the development laws and changing trends of data. Its core principle is based on time series analysis theory, arranging data in chronological order, and capturing the characteristics of long-term trends, seasonal fluctuations, and random fluctuations of data by calculating indicators such as moving averages, growth rates, and month-on-month ratios of data, or constructing time series models (such as ARIMA models).

In auditing, trend analysis technology can be used to analyze the financial data (such as operating income, cost and expense, profit, etc.) and business data (such as sales volume, purchase volume, etc.) of the audited units. By comparing the changing trends of data in different periods, it can be judged whether there are abnormal fluctuations in the data. For example, if the operating income of an enterprise has maintained steady growth in the past few years but suddenly drops significantly in a certain period without a reasonable explanation, it may imply that the enterprise has problems such as financial fraud or poor management. Trend analysis technology can help auditors quickly find abnormal trends in data and provide clues for further auditing.

(2) Cluster Analysis and Anomaly Detection Technology

Cluster analysis technology is an unsupervised learning method that groups samples in a data set according to similarity. Its principle is to define a similarity measure between data samples (such as

Euclidean distance, Manhattan distance, etc.), classify samples with high similarity into the same category, and samples with low similarity into different categories, so that samples in the same category have high similarity and samples in different categories have low similarity.

Anomaly detection technology is a technology that identifies abnormal data points in a data set that are inconsistent with the behavior or characteristics of most data. Its principle is to establish a model of normal data based on the distribution characteristics of normal data. When new data is input, it is compared with the normal model to judge whether the data is abnormal. Anomaly detection technology can be divided into statistical-based methods, distance-based methods, density-based methods, etc.

In auditing, cluster analysis and anomaly detection technology are usually used in combination. Firstly, cluster analysis technology is used to cluster the transaction data of the audited units according to characteristics such as transaction amount, transaction frequency, and transaction object, dividing them into different transaction mode categories. Then, anomaly detection technology is used to identify abnormal transaction behaviors that deviate from their category or do not conform to the normal transaction mode. For example, when analyzing the purchase transaction data of an enterprise, cluster analysis can divide purchase transactions into different categories such as large-amount low-frequency transactions and small-amount high-frequency transactions, and then anomaly detection technology can find transactions whose amounts are much higher or lower than the average level of their category. These abnormal transactions may involve problems such as false purchases and interest transmission.

(3) Correlation Analysis Technology

Correlation analysis technology is a data analysis method used to discover the correlation between data. Its principle is based on association rules, determining the strength of association between data items by calculating the support and confidence between data items. Support refers to the proportion of transactions that contain two data items at the same time in all transactions; confidence refers to the proportion of transactions that contain one data item and also contain another data item.

In auditing, correlation analysis technology can be used to discover the correlation between different business data, such as the correlation between sales data and inventory data, and the correlation between purchase data and payment data. By analyzing these correlations, it can be judged whether the business process is reasonable and whether there are abnormal situations. For example, under normal circumstances, there should be a negative correlation between sales data and inventory data, that is, an increase in sales volume leads to a decrease in inventory. If correlation analysis finds that sales data increases but inventory data does not decrease accordingly, it may imply problems such as poor inventory management and false sales data.

(4) Machine Learning Prediction Technology

Machine learning prediction technology is a technology that uses machine learning algorithms to train historical data, build a prediction model, and then use the model to predict future data or unknown data. Its principle is to select appropriate feature variables, input historical data into machine learning algorithms (such as logistic regression, decision trees, neural networks, etc.) for training, adjust model parameters so that the model can accurately fit historical data, and then use the trained model to predict new data.

In auditing, machine learning prediction technology can be used to build an audit doubt probability evaluation model. By training the known problem data and normal data in historical audit data, the model can learn the characteristics of problem data, and then predict the probability of each audit doubt having problems according to the characteristics of the current audit data. Auditors can reasonably arrange audit resources according to the level of doubt probability, giving priority to verifying high-probability doubts, and improving audit efficiency. For example, using logistic regression algorithm to build an audit doubt probability evaluation model, taking transaction amount, transaction frequency, credit rating of transaction objects, etc. as feature variables, through training on historical data, the model can output the probability value of each audit doubt having problems.

2.3 The Fit between the Theory of Full Coverage of Auditing and Big Data Auditing

Full coverage of auditing refers to the full coverage of auditing of public funds, state-owned assets, state-owned resources, and the performance of economic responsibilities by leading cadres. Traditional auditing modes are limited by factors such as manpower and time, making it difficult to truly achieve full coverage of auditing. However, big data auditing, with its characteristics of high efficiency, real-time performance, and wide coverage, can quickly process and analyze massive amounts of data, thus providing a strong guarantee for full coverage of auditing.

The theory of full coverage of auditing requires that auditing work can comprehensively, objectively, and timely reflect the economic activities of the audited objects, find existing problems and risks, and put forward effective audit suggestions. The fit points between big data auditing and the theory of full coverage of auditing are mainly reflected in the following aspects:

(1) Fit in coverage scope. Big data auditing can integrate massive amounts of data from different departments and fields, realize all-round and multi-angle supervision of the audited objects, break through the limitations of traditional auditing in terms of region, industry, and department, and meet the requirements of full coverage of auditing for coverage scope.

(2) Fit in timeliness. Traditional auditing modes are often post-audit, and the timeliness of audit results is poor. Big data auditing can monitor and analyze the data of the audited objects in real-time or quasi-real-time, find problems and risks in a timely manner, provide timely information support for audit decisions, and meet the requirements of full coverage of auditing for timeliness.

(3) Fit in depth. Full coverage of auditing not only requires that auditing work covers all audited objects but also requires in-depth analysis and evaluation of the economic activities of the audited objects. Through in-depth mining and analysis of massive data, big data auditing can find deep-seated problems and risks that are difficult to find in traditional auditing, improve the depth and quality of auditing work, and meet the requirements of full coverage of auditing for depth.

By building an audit big data warehouse, integrating data resources from various fields, and using big data analysis technology to conduct overall analysis of all data, we can comprehensively grasp the economic activities of the audited objects, find potential problem doubts, and then conduct targeted decentralized verification, realizing all-round and whole-process supervision of the audited objects, effectively making up for the shortcomings of traditional auditing, and enabling the full coverage of auditing to be truly implemented. For example, in the audit of financial budget execution, big data auditing can be used to centrally analyze the financial data of all budget units in the city, quickly screen out units and matters that may have problems, and then arrange auditors to conduct on-site verification, which greatly improves the efficiency and coverage of auditing.

3. Results

3.1 Data Analysis Results

This study collected rich data through various channels, including financial data, business data, and related external data of the audited units. In the process of data collection, full consideration was given to the completeness, accuracy, and timeliness of the data. To ensure data quality, data cleaning technology was used to preprocess the collected data, removing noise data, correcting wrong data, and reasonably filling in missing values. After data cleaning, the availability of data was effectively improved, laying a solid foundation for subsequent data analysis.

The trend chart of the financial data of the audited unit in the past five years was drawn using MATLAB (see Figure 1). It can be clearly seen from the figure that the operating income of the unit shows an increasing trend year by year, but the growth rate has slowed down in the past two years; the cost and expense have been increasing year by year, and the increase is more obvious. Through further analysis of the composition of cost and expense, it is found that the increase in raw material purchase cost and labor cost is the main reason for the increase in total cost. This result suggests that

auditors need to focus on possible problems in the cost control of the unit, such as whether the procurement process is standardized and whether the labor cost accounting is accurate.

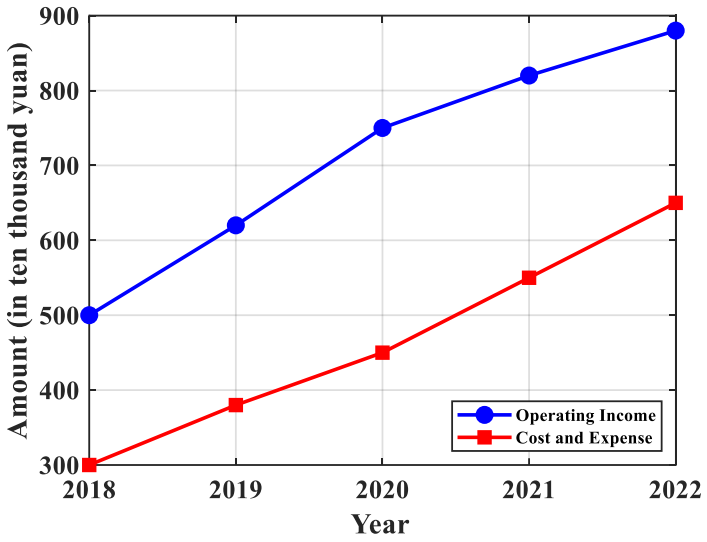

Figure 1. Trend Chart of Financial Data

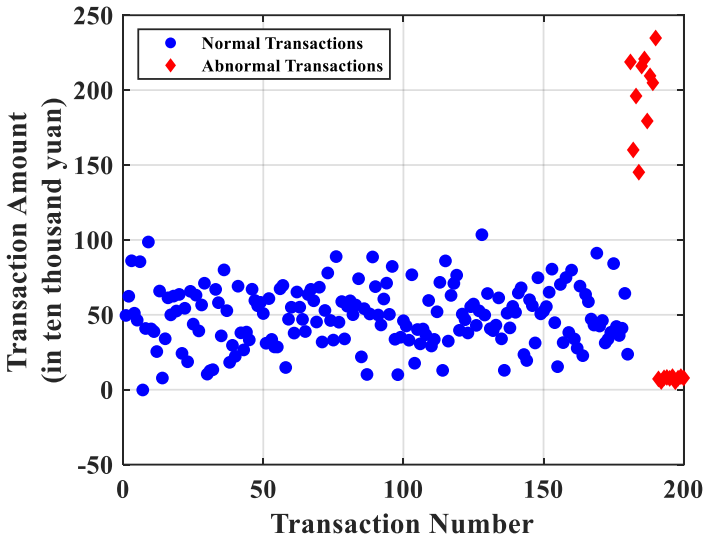

Figure 2. Scatter Plot of Transaction Amount Clustering

Using cluster analysis and anomaly detection algorithms, the transaction data of the audited unit was analyzed, and a number of abnormal transaction behaviors were successfully identified (see Figure 2). In the scatter plot of transaction amount distribution, it was found that some transaction points deviated from the clustering range of normal transaction amounts, and the amounts of these abnormal transactions were significantly higher or lower than the normal level. Further investigation found that these abnormal transactions involved frequent large-amount transactions with specific suppliers and some businesses with very small amounts but frequent transactions, which had potential risks of interest transmission and illegal operations, and auditors could carry out in-depth investigations based on this.

To sum up, an audit doubt probability evaluation model was built. By learning and training historical audit data and known problem data, the model can predict the probability of each audit doubt having problems according to the characteristics of the current audit data. The analysis results are presented in the form of probability values (see Figure 3). Auditors can reasonably arrange audit resources according to the level of doubt probability, giving priority to verifying high-probability

doubts, and improving audit efficiency. The model evaluation found that the probability of some doubts involving large-amount fund expenditures is as high as more than 0.9. A

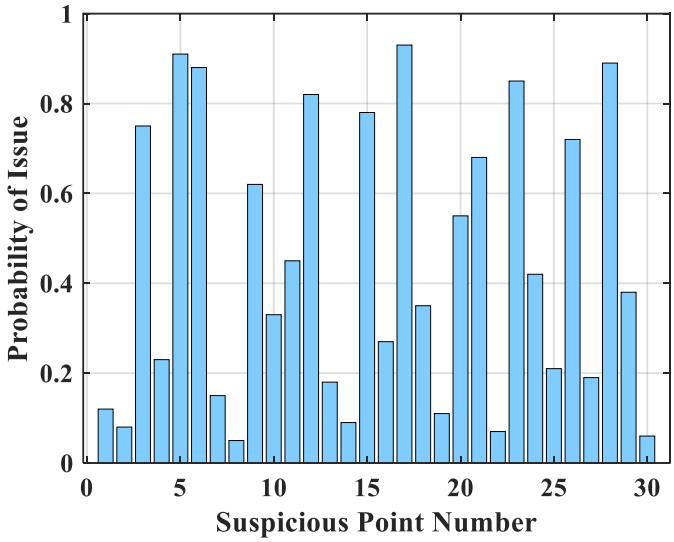

Figure 3. Scatter Plot of Transaction Amount Clustering

fter on-site verification by auditors, it was finally confirmed that these doubts do have violations.

4. Discussion

A leading group for big data auditing work was established, with the main leaders of auditing institutions as the team leaders, responsible for overall coordination of the planning, deployment, and promotion of big data auditing work. The responsibilities of each department in big data auditing work were clarified, and a sound work responsibility system was established to ensure that all tasks are implemented. Strengthen the organizational leadership and supervision and inspection of big data auditing work, timely solve problems arising in the work, and ensure the smooth progress of big data auditing work. Formulate and improve relevant systems and norms for big data auditing, including data management methods, audit operating procedures, quality control systems, security and confidentiality systems, etc. Through system construction, the processes, standards, and requirements of big data auditing work are clarified, the behaviors of auditors are standardized, and ensure that big data auditing work is carried out in accordance with laws and regulations. Strengthen the supervision and inspection of the implementation of the system, and regularly evaluate and revise the system to ensure its effectiveness and adaptability.

Increase efforts in training big data auditing talents and formulate scientific and reasonable talent training plans. Improve the professional ability and comprehensive quality of auditors in big data auditing through various methods such as internal training, external further study, and training through audits. Establish a talent incentive mechanism, commend and reward personnel who have performed outstandingly in big data auditing work, and stimulate the enthusiasm and creativity of auditors. At the same time, actively introduce outstanding talents with professional backgrounds in big data analysis, computer science, etc., to enrich the big data auditing talent team. Attach great importance to information security issues in big data auditing and establish a sound information security protection system. Strengthen the security management of data, take measures such as encrypted storage, access control, and data backup to ensure the security and confidentiality of data. Strengthen the security protection of the audit information system, conduct regular security vulnerability scanning and repair, and prevent the risks of network attacks and data leakage. Strengthen the education on information security awareness for auditors to improve their security awareness and emergency response capabilities.

5. Conclusions

This study deeply explores the implementation paths and safeguard mechanisms for promoting the implementation of big data auditing, and draws the following conclusions through theoretical analysis and empirical research: As an innovative development direction in the auditing field, big data auditing has significant advantages, which can effectively improve the quality and efficiency of auditing work and provide strong support for the realization of full coverage of auditing. In terms of implementation paths, efforts should be made to strengthen data governance, improve technical capabilities, optimize auditing processes, and establish coordination mechanisms to promote the in-depth integration of big data auditing and auditing business. In terms of building safeguard mechanisms, it is necessary to establish organizational guarantee, system guarantee, talent guarantee, and security guarantee systems to provide all-round guarantee for the smooth implementation of big data auditing.

However, big data auditing still faces many challenges in practical applications, such as uneven data quality, the need to improve the level of technical application, and talent shortage. In the future, the auditing field should further strengthen the research and practical exploration of big data auditing, continuously improve the implementation paths and safeguard mechanisms, continuously improve the application level of big data auditing, and give full play to its important role in promoting the healthy development of the economy and society and maintaining national economic security. At the same time, with the continuous development of information technology, big data auditing will also continue to evolve and innovate. Auditors need to keep up with the times and actively learn and master new technologies to adapt to the new requirements of the development of auditing work.

References

[1] Gepp A, Linnenluecke M K, O'neill T J, et al. Big data techniques in auditing research and practice: Current trends and future opportunities[J]. Journal of Accounting Literature, 2018, 40(1): 102-115.

[2] Salijeni G, Samsonova-Taddei A, Turley S. Big Data and changes in audit technology: contemplating a research agenda[J]. Accounting and business research, 2019, 49(1): 95-119.

[3] Sun Y, Li J, Lu M, et al. Study of the impact of the big data era on accounting and auditing[J]. arXiv preprint arXiv:2403.07180, 2024.

[4] Hezam Y A A, Anthonysamy L, Suppiah S D K. Big data analytics and auditing: A review and synthesis of literature[J]. Emerging Science Journal, 2023, 7(2): 629-642.

[5] Mehta A M. 08.04. The Role of Technology in Auditing[J]. Internal Auditing: A Practical Approach, 2024.

[6] De Santis F, D'Onza G. Big data and data analytics in auditing: in search of legitimacy[J]. Meditari Accountancy Research, 2021, 29(5): 1088-1112.