Research on the Application of Big Data Technology in Financial Audit Data Analysis

Xi Zhu

Wuhan Technology and Business University; Hubei, 430065, China

mmxzmm123@163.com

Abstract. Against the backdrop of the rapid expansion and increasing complexity of data in the field of financial auditing, big data technology has brought new methods and tools to auditing work. The study analyzed the diverse applications of big data in the field of financial auditing, including data integration, risk detection, technological improvement, and compliance review from multiple dimensions. The concept of big data was explained, and the difficulties encountered in financial auditing, such as scattered data sources, privacy and security risks, and technological bottlenecks, were discussed. Proposed response strategies such as building a unified data integration platform, using intelligent algorithms for risk warning, adopting cutting-edge technology tools to enhance technical strength, and developing audit processes that comply with regulatory standards, aiming to provide solid technical support for innovative development and industry upgrading in the audit field.

Keywords: big data technology; Financial auditing; Data analysis; Risk identification; Compliance Audit

1 Introduction

In the financial sector, the audit process plays a crucial role in ensuring the stable operation of enterprises and the effective implementation of regulatory policies. But with the dramatic increase in information volume and the increasing complexity of business forms, traditional auditing methods are no longer sufficient to meet the challenges of modern financial markets. In this context, big data technology, with its outstanding ability to process large datasets and quickly filter key information, has gradually become an important technical means in the financial auditing industry. Outlined the connotation of big data technology and explored the challenges encountered in the application of financial auditing. Corresponding strategies have been proposed to address these issues, aiming to provide reference for improving the efficiency of financial auditing and promoting the digital transformation of the industry through in-depth analysis and resolution of these problems.

2   Definition of Big Data Technology

The series of technological systems that utilize cutting-edge information technology and efficient computing devices to collect, store, process, interpret, and use massive, heterogeneous, and real-time updated structured and unstructured data is called big data technology. The core of technology is to use powerful data processing capabilities to uncover the hidden information value of data, assist in scientific decision-making, and predict future trends. Technology values the magnitude of data scale, with a greater emphasis on the diversity, speed, and important role of data in value creation.
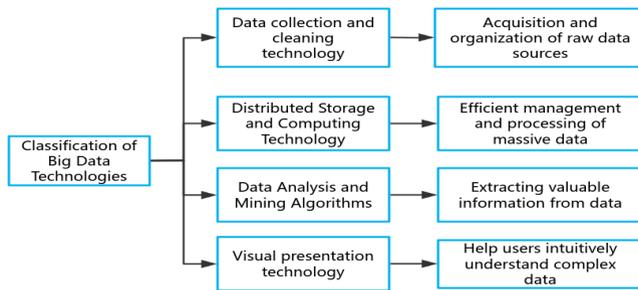

Fig. 1. Big data technology

The core framework of big data technology includes data collection and cleaning technology, distributed storage and computing technology, data analysis and mining algorithms, and visualization presentation technology, as shown in Fig 1. The collection and organization of raw materials is the task of data collection and cleaning. With the help of distributed systems, storage and computing technologies can effectively manage and process large-scale data. By applying statistical principles and machine learning techniques, data analysis and mining techniques can extract valuable information from massive amounts of data. Visualization technology helps users grasp complex data information more vividly. In the financial audit industry, the integration and application of big data technology injects new vitality into traditional audit processes, helps enterprises efficiently process audit information, uncover hidden risk factors, and promote the development of audit work towards more accurate and dynamic directions.

3   Challenges Faced By Big Data Technology in Financial Auditing

3.1   Data Sources Are Scattered, With Diverse Formats and Uneven Quality

The core issue in the financial auditing industry lies in the complexity of data sources and the dispersion of their distribution[1]. The various types of information that must be processed during the audit process come from a wide range of sources, including

bank transaction details, financial reports, electronic bills, and consumer behavior information. These information are scattered across different systems and platforms, without forming a unified standard and interface, making data collection and integration difficult. The format of data also varies, covering structured information, semi-structured information (such as JSON, XML documents), and unstructured information (such as documents, images, etc.). Before conducting in-depth data analysis, tedious data preprocessing is necessary, which increases the difficulty of audit work and prolongs the required time. The inconsistent quality of data is also a major challenge in the audit process. Some data may be incomplete, missing, or even incorrect, which may be caused by limitations of recording devices or human errors. However, the organization of historical data lacks systematic cleaning and updating, resulting in a large amount of redundant information, occupying storage space, and affecting the accuracy of data analysis results.

3.2   Data Privacy and Security Risks

The data processed by financial institutions includes sensitive information and trade secrets of customers, such as bank account information, fund transaction history, personal identity authentication, and other key information[2]. Once this information is leaked, it will cause property damage to customers, have a negative impact on the reputation of financial institutions, and may trigger legal proceedings. Financial auditing utilizes big data technology, with its core being the ability to store and process large-scale data. However, this centralized data management approach also increases the risk of being hacked. The financial industry is frequently subjected to cyber attacks and data breaches. Given the increasingly advanced nature of cyber attack methods, traditional security strategies are insufficient to provide sufficient protection. During the data transmission phase, if sensitive information is not encrypted, it may be illegally intercepted or tampered with, endangering the integrity and security of the data. The power of big data analysis lies in its ability to uncover deeper information through the correlations between data, but this also means an increase in privacy risks.

3.3   There Are Certain Technological Bottlenecks

Although big data technology has opened up new avenues in the field of financial auditing, it still faces many limiting factors in practical application, which affect its widespread application and in-depth exploration. The issue of computing power is particularly prominent, and data processing in the field of financial auditing requires rapid response and extremely high accuracy[3]. Processing massive amounts of data relies on complex computing methods and efficient computing resources, which poses a serious challenge to current hardware facilities and software platforms. Especially when dealing with unstructured data, traditional methods may not be efficient and cannot meet the timeliness requirements of auditing. Financial auditing involves many industry-specific norms and complex data association patterns, and current big data tools and algorithms often struggle to fully adapt to these unique requirements. For example, using machine learning models in auditing requires a large amount of anno-

tated data for model training, but in the financial industry, such high-quality annotated data is not easy to obtain[4].

4 Application Strategies of Big Data Technology in Financial Auditing

4.1 Building a Data Integration and Unified Platform

Creating a comprehensive data integration and unified platform can improve the efficiency of data processing and achieve standardized workflow[5]. This platform utilizes various methods such as API calls, database connections, and file transfers to gather data information from internal enterprise platforms (such as financial platforms, sales platforms, customer relationship management platforms) and external resources (such as open databases, API interfaces from other service providers). After data aggregation, the platform uses ETL tools to purify, transform, and store the data, ensuring consistency in data format (such as unified time format and amount expression) and removing duplicate and inaccurate data items. The platform introduces a metadata management system to ensure data consistency and reduce communication errors between different departments through unified data description and specifications. In addition, adopting distributed storage technologies such as HDFS, HBase, etc., improves the speed of data reading and writing. The standardized conversion after data cleaning can be modeled using the following formula:

$$D_{cleaned} = Normalize(D_{raw}, f_1, f_2, ..., f_n) \quad where \quad f_i \in \{Date, Amount, ...\} \tag{1}$$

In formula (1), D_{raw} represents the raw data, f_1, f_2,..., f_n are conversion standards for different data fields (such as date, amount, etc.), while $D_{cleaned}$ is the standardized and cleaned dataset. The platform has implemented real-time monitoring of data quality, ensuring the integrity, accuracy, and uniformity of information, while using permission allocation and encryption methods to maintain the legitimate permissions and privacy security of data. Sensitive information is encrypted and stored, and data leakage is avoided through the setting of job permissions. The system log records every data operation in detail to meet compliance requirements. The system integrates advanced data modeling tools and analysis modules (such as Apache Spark, TensorFlow) to enhance data processing capabilities, utilizing big data for in-depth mining, revealing hidden development trends and abnormal situations. The analysis results are presented using visualization software such as Tableau and Power BI, which facilitates decision-makers to quickly grasp key information. In addition, the system can automatically generate standardized reports to assist in decision-making and subsequent audit work.

4.2   Using Machine Learning to Identify Risks

In traditional financial institution audits, risk identification relies on subjective judgments and established rule sets, which is inefficient and may overlook some implicit risk factors[6]. Thanks to the development of big data technology and continuous breakthroughs in machine learning technology, the financial risk assessment industry is gradually introducing intelligent algorithms to enhance its ability to identify risks. Machine learning extracts potential patterns and abnormal actions from a large amount of historical data through deep mining, providing more accurate technical support for risk management. The application of machine learning technology is mainly divided into two categories: supervised learning and unsupervised learning. In the context of supervised learning, algorithms are trained by analyzing a labeled historical case library to master the ability to identify specific types of risks. In fraud transaction detection, a logistic regression model is used to predict the fraud risk of a certain transaction, and its core formula is:

$$P(y = \square X) = \frac{1}{1 + e^{-\left(\beta_0 + \sum_{i=1}^{n} \beta_i X_i\right)}} \tag{2}$$

In formula (2), $P(y=1|X)$ represents the probability of the transaction being fraudulent, β_0 is the bias term, β_i is the feature weight, and x_i is the transaction feature (such as amount, location, time, etc.). The audit platform evaluates the suspicious level of transactions and prioritizes the review of transactions with high potential risks, enhancing the accuracy and operational efficiency of financial risk control.

In the field of unsupervised learning, algorithms do not rely on labeled datasets, but mainly rely on their own mining of the internal structure of the data to explore hidden risk patterns. Through clustering techniques, it is possible to identify customer groups or transaction behaviors that do not conform to conventional practices, revealing risk factors that are difficult to detect through conventional means. During the training phase of the model, machine learning algorithms analyze numerous historical cases, extract risk related features, and construct a feature library. With the continuous expansion of information volume and improvement of quality, the feature library will also be updated accordingly to enhance the accuracy of model prediction. In this way, machine learning models have the ability to automatically detect and predict risks, issue real-time warning signals, and improve the efficiency and accuracy of financial audits. By adopting this intelligent risk screening method, the financial audit process breaks away from the traditional manual audit mode and relies on big data support and automated analysis processes to achieve precise identification and early warning of potential risks, supporting more efficient risk decision-making and management.

4.3   Introducing Big Data Analysis Tools to Enhance Technical Capabilities

In the field of financial auditing, big data analysis technology has become a key means to improve audit efficiency and accuracy, and its core lies in the implementation of the following key processes. Integrating data resources from different channels, such as

transaction records, financial reports, and market dynamics of financial institutions, big data systems purify and standardize these complex data for further in-depth analysis. With the help of distributed computing frameworks such as Hadoop or Spark, the system can quickly read, write, and process data, ensuring the comprehensiveness and uniformity of audit data[7]. Then, advanced data mining techniques, including but not limited to clustering, correlation analysis, and anomaly detection algorithms, are applied to conduct in-depth analysis of massive financial datasets. The system can intelligently identify potential risks and abnormal transaction patterns, helping auditors discover potential risk points from numerous transactions. Combining real-time monitoring systems and machine learning algorithms, dynamically analyze financial transaction behavior and trends, and automatically output risk warning information. In the anomaly detection process, the system usually uses Z-Score or standard deviation methods to locate outliers in the data. A common formula is the Z-Score formula, used to measure the degree of deviation between a data point and the mean of the dataset:

$$Z = \frac{X - \mu}{\sigma}$$

(3)

In formula (3), X is the observed value (i.e. a specific financial transaction data point), μ is the mean of the dataset, and σ is the standard deviation of the dataset. Using the Z-Score algorithm, the audit system can automatically detect transactions or activities that differ from the regular pattern, triggering risk alerts or anomaly detection mechanisms. Once an anomaly is detected, the system will immediately notify, allowing auditors to quickly intervene and reduce the risk of human error. By using data visualization software such as Power BI or Tableau, audit results can be presented intuitively to management in the form of charts and dashboards, accelerating the interpretation of complex data and promoting efficient decision-making. This visual data presentation also helps auditors identify potential hazards in their analysis work and improve the accuracy of decision support. According to the analysis, an automated report generator can quickly complete the preparation of audit reports, saving time and cost of manual writing, and improving the accuracy and timeliness of reports.

4.4   Develop Audit Processes Based on Regulatory Requirements to Reduce the Risk of Violations

Financial audit work must strictly comply with laws, regulations, and compliance standards, which permeate every detail of the audit process. Building audit procedures that comply with regulatory requirements is a core strategy to reduce the likelihood of violations. Faced with the increasingly complex global financial regulatory mechanisms, audit firms must adhere to relevant regulations such as the General Data Protection Regulation (GDPR) and the Anti Money Laundering Law to ensure the legality and compliance of audit activities. In the information collection process, the audit work must be integrated into the intelligent compliance audit unit to ensure that the collection, application, and data storage strictly comply with relevant regulations. When analyzing data, privacy information masking and encryption functions should be em-

bedded to ensure that customer secrets are not leaked and maintain data security. During the report generation phase, automated audit verification should be implemented to prevent overlooking the legally mandated review points. In anti money laundering audits, a multidimensional risk assessment model can be designed according to regulatory requirements to identify high-risk nodes by calculating the risk distribution in the transaction network. The audit model defines the comprehensive risk score Ri of customer i as:

$$Ri = \alpha \cdot \frac{T_i}{\bar{T}} + \beta \cdot N_i + \gamma \cdot \frac{C_i}{\sum_{j=1}^{n} C_j} \tag{4}$$

In formula (4), Ti is the total transaction amount of customer i, and $\bar{T}$ is the average transaction amount of all customers,Ni is the high-frequency trading frequency of customer i, Ci is the number of connection nodes of customer i (such as trading frequency with other accounts), and α, β, and γ are weight coefficients. The audit system sorts risk scores and can quickly identify customer groups with higher potential risks, ensuring accurate risk assessment is completed in accordance with regulatory standards.

5   Conclusion

In the financial auditing industry, the deep application of big data technology has gradually become a trend. With the integration of data processing and the construction of a unified management platform, the integration of intelligent algorithms, and the utilization of advanced tools, the efficiency of audit work has been improved, the sensitivity of risk prevention and control has been enhanced, and potential risks in legal and security aspects have been effectively reduced. Despite this, the improvement of technology and the deep integration of the industry still require time, especially in areas such as data quality control, personal privacy and security, and technological innovation, which still face continuous challenges and room for improvement. In the future, with the continuous advancement and widespread application of big data technology, the technical framework and operational methods in the field of financial auditing are expected to undergo deeper innovative changes.

References

1. Xie Tian."Artificial intelligence and automatic recognition application in B2C e-commerce platform consumer behavior recognition."Soft Computing 27.11(2023):7627-7637.
2. Min Ye,and Tan Chai Ching."Research and Application of Compulsive Buying Behaviors of Consumers in E-Commerce Livestreaming on Big Data."Human-Centric Intelligent Systems 2.3-4(2022):113-123.
3. YananSong."Application of Big Data Analysis Technology in Cross-Border E-Commerce."Journal of Electronic Research and Application 5.4(2021):1-3.

4. Johnson Kolawole Olowookere, et al."AUDIT ATTRIBUTES AND FINANCIAL REPORTING QUALITY AMONG LISTED NON-FINANCIAL FIRMS IN NIGERIA."Zbornik radova Ekonomskog fakulteta Sveučilišta u Mostaru 29(2023):60-85.
5. Mary Tiezel G. Rufin,and Rose A. Arceño."Evaluation of Cooperative Risks in Eastern Visayas:Basis for the Development of Risk Management Manual."Journal of Economics, Management and Trade 29.12(2023):28-53.
6. El Khoury Rim,Harb Etienne,and Nasrallah Nohade."Triple vectorial analysis of financial development impact on economic growth: evidence from the Middle East and Central Asia."International Journal of Emerging Markets 18.10(2023):4301-4319.
7. Kuzey Cemil, et al."Do shareholders appreciate the audit committee and auditor moderation? Evidence from sustainability reporting."International Journal of Accounting & Information Management 31.5(2023):808-837.