Accounting, Auditing and Finance (2024)
Clausius Scientific Press, Canada

DOI: 10.23977/accaf.2024.050302
ISSN 2616-3136 Vol. 5 Num. 3

Optimization Strategy and Implementation Effect Analysis of Unstructured Data Audit for Internal Audit of Commercial Banks in the Big Data Environment

Yijiao Fan*

JP Morgan, New York, 11101, USA
yijiaofanedit@outlook.com
**Corresponding author*

Keywords: Big Data Audit, Unstructured Data, Internal Audit of Commercial Banks, Audit Informatization, Risk Management Optimization

Abstract: This article deeply analyzes the current situation and challenges of internal auditing in commercial banks, especially unstructured data auditing, in big data technology. With the rapid development of information technology, the financial industry is undergoing profound changes. The widespread application of big data technology has brought unprecedented opportunities for risk management and internal auditing of commercial banks, but it is also accompanied by challenges such as complex data processing and increased difficulty in risk prevention and control. The internal audit of commercial banks has expanded from traditional financial statement audit to comprehensive audit in multiple fields such as internal control, asset quality, credit risk, etc. This article focuses on analyzing the problems in unstructured data auditing. It proposes optimization strategies based on the advantages of big data technology, aiming to improve audit efficiency and quality, form a unified standard, and promote the informatization process of auditing. Taking Bank A as an example, its rapidly developing internal audit is facing new challenges. The research results of this article have important reference value for optimizing internal audit processes and enhancing risk management capabilities. At the same time, it is pointed out that the informatization construction of internal audits in commercial banks in China still needs to be improved, and there are many problems. This article reviews the research on unstructured data audit methods, taking Bank A as an example. Through interviews and data analysis, the current status of unstructured data audit is summarized, optimization methods are proposed, and the necessity of big data technology for audit optimization is emphasized. It is pointed out that Bank A and the entire banking industry should attach importance to unstructured data audits, pay attention to technological innovation, and promote the development of audit informatization. However, research also has limitations, such as the need for practical testing of the effectiveness of optimization methods and continuous research to address technological updates. This article aims to provide innovative ideas and suggestions for the development of audit informationization in unstructured data auditing for commercial banks.

1. Introduction

In the surging wave of big data technology, the financial industry is entering a new era full of changes and opportunities. At present, China's economy is promoting the revolution of information technology and the integration of two technologies at an unprecedented speed. The vigorous development of Internet technology is like a strong easterly wind, prompting traditional industries to step onto the fast lane of digital transformation. As the pillar of the financial industry, commercial banks are no exception. Faced with the profound impact of the big data environment and multiple challenges such as frequent changes in market policies, commercial banks have to accelerate their pace, actively transform their business strategies, and comprehensively enhance their risk management capabilities. In this context, commercial banks have accumulated massive data resources, among which unstructured data such as text documents, videos, audio, images, etc. account for an important proportion. This article takes this as a starting point to deeply analyze the current situation and challenges of unstructured data auditing in commercial banks. Taking A Commercial Bank as an example, a detailed analysis of its internal audit practice, reveals the effectiveness and shortcomings of internal auditing in the big data environment. As a representative of rapidly rising commercial banks in recent years, A Bank's innovative practices in internal audit methods and processes have important reference significance for the entire industry. Based on the above analysis, this article aims to propose a series of strategies for optimizing unstructured data auditing methods by combining the characteristics and advantages of big data technology. These strategies aim to improve the efficiency and quality of audit work, promote the formation of unified audit work standards, better prevent and expose risky behaviors, and accelerate the process of audit informatization.

2. Correlation Theory

2.1. Theory Related to Data Assets

In the big data environment, with the rapid development of mobile Internet, social media, and cloud technology, the Internet and various industries have generated massive data, including structured, unstructured, and semi-structured data. The continuous advancement of big data technology has significantly improved the ability to analyze various types of data, especially unstructured data [1] such as document text, images, audio and video, etc. These data occupy an important position in the internal audit work of commercial banks. Unstructured data, due to its complexity and diversity, is difficult to represent directly in two-dimensional relationships, but it contains rich valuable information and audit clues. With the deepening of digital reform in commercial banks, the types of data involved in internal auditing are becoming increasingly diverse, including business contracts, meeting minutes, customer information, credit approval materials, security images, and other forms of data. Optimizing audit methods for unstructured data and utilizing big data and artificial intelligence technologies can not only improve audit efficiency and accuracy but also bring significant value to commercial banks in terms of operations and risk management. The analysis of unstructured data such as text, images, audio, and video requires specific technical methods, such as text segmentation, clustering analysis, image segmentation, feature extraction, etc., to uncover hidden information and potential risk points, providing strong support for audit decisions.

2.2. Data Audit Theory

In the context of big data, unstructured data audit theory is risk-oriented and aims to evaluate the

internal management and risk status of banks through information analysis methods. Audit objectives determine the focus of the pre-audit investigation, and the investigation results react to the realization of audit objectives, forming a mutual constraint relationship. The core of unstructured data auditing is to assess the risks throughout the entire data collection and analysis process, which requires auditors to have a deep understanding of the audited entity's business operations, database architecture, data distribution, and storage methods, as well as the relationships between databases. With the popularization of big data technology, the content of internal audits in commercial banks [2] has completely shifted towards information dominance. Innovative big data technology combined with on-site and off-site audit methods overcomes the limitations of traditional sampling auditing, achieving rapid collection, dynamic processing, data mining, and real-time analysis of massive data. This transformation has led to continuous updates in audit methods. The unstructured data audit method covers the entire process of data collection and analysis, from collecting data through various methods such as system export, database queries, and web crawlers, to using computer technology to assist in data processing, analysis, and suspicious point discovery. This effectively improves the efficiency and accuracy of audit work, promoting a profound transformation of audit mode from information technology to information-based auditing.

3. Research Method

3.1. The Current Status of Unstructured Data Audit Methods in Commercial Banks

At present, the unstructured data audit methods of commercial banks are undergoing rapid changes, and their data collection, analysis, audit scope, and model application are all showing new characteristics. In terms of data collection, the methods are divided into the on-site audit stage and the off-site audit stage. On-site auditing remains the core approach for obtaining evidence in unstructured data auditing. Using inquiry, consultation, and other means, rich audit evidence is obtained, including electronic financial system data, paper materials, and oral evidence, providing a solid foundation for audit opinions. As a supplement, the off-site audit stage collects extensive data through external official institutions, industry information databases, and corporate websites to broaden the audit perspective. At the level of data analysis, although commercial banks have adopted some computer-aided methods such as database management and EXCEL tools, the discovery of suspicious points in unstructured data auditing still heavily relies on manual analysis, which to some extent limits the efficiency and comprehensiveness of auditing. The commonly used analysis methods include inquiry, review, observation, and analysis. Although these methods are widely used in practice, they are limited by personnel and time costs and lack automatic warning mechanisms. With the deepening development of auditing work, there have been significant changes in the scope, functions, and time frame of auditing. The internal audit system [3] has gradually improved, not only expanding the scope of audit space, but also expanding the audit function from traditional error checking and rectification to consulting and management functions, and the focus of work has also extended to internal control audit [4]and risk management audit [5]. In addition, the audit time range has gradually shifted towards pre-audit and in-process audits, emphasizing risk prevention and monitoring, reflecting the important role of internal audits in internal control and risk management. The innovation of big data and computer technology has driven the innovation of audit methods in the application of audit models. The generalized audit model includes audit thinking and method models, which are risk-oriented and achieve control over key process points and prediction of potential risks through standardized and systematic audit processes. The introduction of the new model not only improves the accuracy and efficiency of auditing but also provides strong support for auditing work in the big data environment.

3.2. The Problems of Unstructured Data Audit Methods in Commercial Banks

This chapter elaborates on the multiple challenges currently faced by commercial banks in China in unstructured data auditing. Firstly, there are significant issues with the audit method model [6], mainly manifested as a low level of information technology auditing, insufficient innovation, and a low proportion of information system auditing. This is mainly due to the limited level of information technology development, which makes traditional sampling methods difficult to cope with massive system data, data analysis models difficult to adapt to complex data types, and insufficient attention to audit evidence in external networks, resulting in lagging audit work and passive risk analysis. In addition, there is a gap between the technical and business layers in the audit model. On-site audits focus more on structured data analysis than technical analysis, while off-site audits ignore the collection and utilization of external unstructured data, resulting in insufficient cooperation and experience sharing, which affects audit quality. Secondly, in terms of technical methods, the popularity and application level of information systems are limited, traditional audit methods still dominate, the level of information-based auditing needs to be improved, the built-in functions of audit systems need to be perfected, and the application of risk analysis methods and models needs to be more effective. Furthermore, in terms of management methods, there is a lack of quantitative rules and standards for audit data management, insufficient personnel for data quality management, and difficulty in ensuring management quality. Finally, in the face of complex and ever-changing internal and external environments, auditors need to expand their knowledge systems to cope with new technologies and risks. In summary, Chinese commercial banks need to comprehensively optimize and improve their methods, technical methods, management methods, and personnel allocation for unstructured data auditing.

4. Results and Discussion

4.1. Case Study

This article selects Bank A as the research case, mainly based on its industry representativeness and significant credit risk issues. As a rapidly growing financial institution, Bank A not only has a complete information system construction, providing a solid foundation for audit research in the big data environment but also has a long-term non-performing loan ratio higher than the industry average, highlighting the urgency of credit risk identification and prevention. Through in-depth understanding, it was found that Bank A has deficiencies in its internal audit, especially in the pre-loan, in-loan, and post-loan management of its loan business. Traditional sampling audit methods are difficult to cope with the huge volume of loan business and constantly changing business conditions. Therefore, based on big data technology, this article focuses on the unstructured data audit method in A Bank's credit business, aiming to improve audit efficiency and accuracy, timely detect and prevent risks, and reduce non-performing loan rates by optimizing audit processes and technical means. Through the analysis of a specific case of Bank A, this article proposes an optimized unstructured data audit method using case study methods and existing theories, to provide valuable reference and inspiration for China's commercial banks in the field of risk management..

4.2. Case Problem

Bank A's audit processes confront critical challenges, including incomplete data, inadequate analysis, and a lack of standardization. Within credit and post-loan management audits, ensuring data authenticity and comprehensiveness is challenging while neglecting unstructured data hinders

accuracy. Limited audit resources lead to superficial assessments and overlooked insights. Additionally, inadequate business risk management and missing audit rectification mechanisms weaken effectiveness. To tackle these issues, this article proposes a risk-focused optimization plan leveraging big data technology. By fusing on-site and off-site audits with data mining and natural language processing, this approach enables efficient collection and analysis of unstructured data, boosting accuracy and efficiency. It targets high-risk areas, enhancing audit precision and Bank A's risk management capabilities, thereby bridging the gap between audit resources and data volumes.

4.3. Optimization Measures

The optimization implementation framework of A Bank's unstructured data audit method in the big data environment aims to achieve comprehensive collection and analysis of unstructured data by constructing standardized processes and deeply integrating computer technology into the audit system [7]. This framework first establishes data collection standards, utilizing specific system interfaces, batch replication, and web crawling technologies [8] to extensively integrate internal and external data resources, ensuring the comprehensiveness and timeliness of data. Subsequently, an unstructured database was constructed to unify data formats, promote cross-departmental and cross-institutional information sharing, integrate internal business basic data with external environmental data, and improve the quality of credit evaluation. At the same time, banks make use of emerging information channels to enhance market sensitivity and judgment, explore external expert consultation mechanisms, and improve professional level. To improve this optimization research and promote practice, it is recommended to expand the audit scope to include management and consulting functions, emphasizing real-time supervision and continuous auditing; Innovate information audit technology to improve audit efficiency and accuracy; Cultivate comprehensive audit talents, enhance team professional competence and information operation ability; Improve mechanisms, standardize systems, formulate coordination rules, promote the modernization process of banking audit work, and jointly promote the innovation and application of unstructured data audit methods.

5. Conclusion

Against the backdrop of the rapid development of big data technology, this article delves into the current status and optimization strategies of unstructured data audit methods for commercial banks in China. By reviewing relevant literature and analyzing the case of Bank A, this article reveals the challenges faced by commercial banks in the process of internal audit informationization construction, especially the shortcomings in unstructured data auditing. Research has pointed out that as the market environment becomes increasingly complex and competitive pressure increases, the risks faced by banks are becoming increasingly difficult to predict and identify. Traditional audit methods are no longer able to meet current needs and urgently need to be optimized with the help of big data technology. Based on the specific case of Bank A, this article summarizes the current status of its unstructured data audit methods and analyzes the existing problems, including insufficient feasibility of audit models, lack of standardized processes, outdated technical methods, and the need to expand the knowledge system of auditors. Based on these issues, this article proposes optimization strategies for the unstructured data audit method of Bank A, aiming to improve audit efficiency and quality, and better cope with complex risks. The research contribution of this article is significant, not only emphasizing the core role of big data technology in optimizing unstructured data audit methods but also deeply analyzing the common problems faced by the banking industry in this field through the example of Bank A. We have proposed targeted optimization strategies aimed at providing practical and feasible reference solutions for A Bank and

commercial banks in China, promoting the dual improvement of audit efficiency and quality. At the same time, this article also actively calls on commercial banks to increase their attention to the application of unstructured data in auditing, advocating for continuous innovation in information audit technology and methods to adapt to the ever-changing financial environment. However, it should be pointed out that this study also has limitations, including the need for practical verification of the effectiveness of optimization methods, limitations in implementation conditions due to current technological levels, and uncertainties in future technological development. All of these require us to maintain continuous attention and in-depth research on the optimization of unstructured data audit methods. Overall, this article provides innovative ideas and practical suggestions for the informatization development of internal audits in commercial banks through in-depth research on the unstructured data audit methods of Bank A. It is expected to promote the optimization and upgrading of audit methods in the entire banking industry in the big data environment.

References

[1] Haan E D, Padigar M, Kihal S E, et al. Unstructured data research in business: Toward a structured approach. Journal of Business Research, 2024, 177. DOI: 10. 1016/j. jbusres. 2024. 114655.

[2] Pizzi S, Venturelli A, Variale M, et al. Assessing the impacts of digital transformation on internal auditing: A bibliometric analysis. Technology in Society, 2021, 67. DOI: 10. 1016/j. techsoc. 2021. 101738.

[3] Si Y. Construction and Application of Enterprise Internal Audit Data Analysis Model Based on Decision Tree Algorithm. Discrete Dynamics in Nature and Society, 2022, 22-25.

[4] Michailidis I D, Alexandridou K, Nerantzidis M, et al. Revisiting the linkage between internal audit function characteristics and internal control quality. The Journal of Operational Risk, 2022. DOI: 10. 21314/jop. 2021. 015.

[5] Marques R P. Continuous Assurance as Digital Transformation Enabler of Audit, Risk Management, and Business Compliance. IEEE Access, 2021.

[6] Pamungkas W A P, Gantyowati E. The Affecting Factors of Audit Quality with Earnings Surprise Benchmark Approach as a Measure of Audit Quality. European Journal of Business Management and Research, 2021, 6 (4): 230-236. DOI: 10. 24018/ejbmr. 2021. 6. 4. 983.

[7] Zhong H, Yang D, Shi S, et al. From data to insights: the application and challenges of knowledge graphs in intelligent audit. Journal of Cloud Computing, 2024, 13 (1). DOI: 10. 1186/s13677-024-00674-0.

[8] Pan J. Application of Web Crawler Technology Based on Python in Big Data Environment. International conference on Big Data Analytics for Cyber-Physical-Systems. Springer, Singapore, 2022. DOI: 10. 1007/978-981-16-7466-2_63.