

Applied Mathematics and Nonlinear Sciences

<https://www.sciendo.com>

Construction of a social audit platform based on big data for "industry price, quality and credit"

Rui Ma[†]

School of Financial Technology, Suzhou Industrial Park Institute of Service Outsourcing, Suzhou, Jiangsu, 215123, China

Submission Info

Communicated by Z. Sabir

Received January 9, 2022

Accepted July 5, 2022

Available online March 17, 2023

Abstract

As the latest product of the innovative integration of information technology, big data has had a huge impact on many fields. Big data technology has changed the traditional audit method, expanded the scope and field of auditing, and injected vitality into the national big data audit work. Full audit coverage has given social auditing a new mission and posed new challenges to creating mega data audit platforms. Driven by the overall goal of full audit coverage, the construction and innovation of a big data audit platform are particularly important. This subject combines data-related technologies and uses more mature data mining algorithms to try to establish an audit platform, and analyzes the three aspects of requirements, concepts and strategies for building an audit platform, dividing it into five modules: data center, acquisition, pre-processing, analysis and data visualization. The research results show that by predicting the revenue indicators of Company A, the overall error of the prediction results of the big data audit platform is controlled at 0.79%, which is better than 8.42% of the traditional audit model; by scoring the two audit models, the score of the big data audit method is 0.7024 higher than that of the traditional model, which is 0.3933, indicating that its accuracy, efficiency comprehensiveness, sustainability, and low risk are In this way, this paper explores a new effective audit path and method, which provides a good basis for the enterprise's "industry price, quality and trust payment" and guides the way forward for future mega Data audits.

Keywords: Big data; Full audit coverage; Data mining; Pre-processing; Visualization

AMS 2020 codes: 62-07

[†]Corresponding author.

Email address: mar489@126.com

1 Introduction

Since 2008 with the influence of the Internet and other media, the volume of data has increased significantly, the concept of big data was born, and related technologies are gradually applied to the reality of work [1-2]. To better comply with the trend of economic globalization, most changes in the application of analytical procedures are interspersed with the corresponding risk operations during the audit [3]. In the new era of auditing, the effectiveness of analytical procedures cannot be underestimated. It covers the review of financial statements, the assessment of estimated risks and the testing of risks. Its application in these periods allows to deepen the understanding of the reality of the audited organization, to find problems and to identify risks. However, with the widespread use of these procedures, some of their weaknesses and drawbacks have come to the fore, such as the possibility of hiding problems and lack of objectivity [4-5]. The emergence of the big data environment has created a better basis for improving such procedures.

The literature [6] argues that applying techniques will strengthen the predictive power of data analysis. The literature [7] investigates the advantages and barriers of incorporating mega data technologies into financial statement audits and identifies specific aspects of big data that may benefit auditors. The literature [8] analyzes the influence of mega data and cloud technologies on full audit oversight coverage and points out that big data and cloud technologies bring opportunities and challenges to full audit oversight coverage under audit planning and execution phases. The mega data audit platform can effectively improve the national governance capacity and comply with the progress requirements of a mega data audits. The big data audit platform can usually be refined into four subplatforms: acquisition, pre-processing, analysis, and visualization. When collecting data, an audit sampling model can be constructed based on the isolated point analysis method [9-11]. The literature [12] uses the DBSCAN clustering algorithm to cluster the audit sampling association rules and establish the association rules between audit samples and targets. Auditing is problematic regarding data consistency, integrity, identification, aggregation and confidentiality [13-14]. The literature [15] argues that the features of mega data are a very good fit for auditing and will be a strong alternative to traditional auditing models. However, by analyzing the application of big data technologies in auditing, the literature [16] found that they are not as commonly used as in other related fields, and the auditing field lags behind other research areas in using valuable big data technologies. Therefore, everyone should pay close attention to using big data technologies in the audit field. The mega data audit platform can effectively improve national governance and comply with the progress requirements of mega data auditing. Big data audit platforms can usually be refined into four subplatforms: acquisition, pre-processing, analysis, and visualization [17-18]. In the literature [19], theories and methods related to the network analysis of big data for auditing are described, and a framework for the construction of a network analysis platform for big data for auditing is proposed, outlining the network analysis method system of "point, line, surface and block". The literature [20-22] utilizes remote data auditing solutions to verify the integrity of stored data and mines the audit trail features according to mega data visualization technology. To this end, this paper analyzes the construction concept and path of a mega data audit platform in the context of the requirement of full audit coverage and the idea of building a mega data platform.

Firstly, in the process of designing and realizing the digital audit analysis platform, we mainly participated in the initial construction of the platform and proposed some functional design requirements; secondly, we introduced the design principles of the mega data audit analysis platform, B/S structure of web application, service-oriented SOA distributed application framework and J2EE standard, as well as related Oracle database and data mining technology from the technical level; finally, we introduce the application example of using the platform to collect and transform data, correlate and analyze around the audit focus, and focus on analyzing and discovering suspicious and decentralized verification from different perspectives. Finally, the use of mega data audit analysis

platform A company as an application case, the implementation of departmental revenue indicators audit as an example, the introduction of the use of the platform to collect conversion data, around the audit focus on correlation analysis, centralized analysis to find suspicious decentralized verification application examples, and from different perspectives to show the platform in the innovative audit mode, enhance the audit effectiveness of the fundamental role played by the platform, and then through the analysis of big data perspective of enterprise audit work in Then, by analyzing the problems in the concept, platform, personnel, and other aspects of enterprise audit work under the perspective of big data, we put forward optimization suggestions to strengthen the audit concept, improve the platform construction, and enhance the quality of personnel, in order to help the enterprise audit work adapt to the current mega data background, and then improve the quality of audit work in all aspects.

2 Big data audit algorithm modeling

2.1 Data analysis algorithm

Data analysis algorithms and data analysis tools are indispensable for analyzing and processing huge volume and many types of data. According to the different forms of data generated by different application scenarios and different forms of processing, the data analysis techniques and presentation methods to be applied are also different. When processing the data, it is important to understand the relevant types before choosing the relevant technical solutions to carry out the solution. Because of the wide variety of audit data itself and a large number of repetitions, two problems are proposed for the data characteristics, one is the prediction of financial audit indicators on the time series, and the other is how to use the algorithm to provide decisions on financial indicators and other issues, for which the convolutional neural network time series CNN-LSTM prediction model algorithm is selected to solve.

The control process of the LSTM (Long Short Term Memory) Network is according to the circular memory network with two additional states, the hidden state and the cellular state. The core function of the LTSM is that the cellular state of the information transmission pathway enables information to be transmitted continuously during sequence processing by using a "gate" structure to add and remove information, a sigmoid network layer and a multiplication for training. The "gate structure" learns which information is saved and which is forgotten. Sigmoid is equation (1):

$$f(x) = \frac{1}{e^{-x}} = \sigma(x) \quad (1)$$

Sigmoid works by making each number in the cell state have a value between 0 and 1 by operation. If the gate output is 0, the product of it and all vectors is 0 vector, and when multiplied by 1, it will give a 1 vector, with 0 representing "completely discarded" and 1 representing "completely retained", as shown in Figure 1.

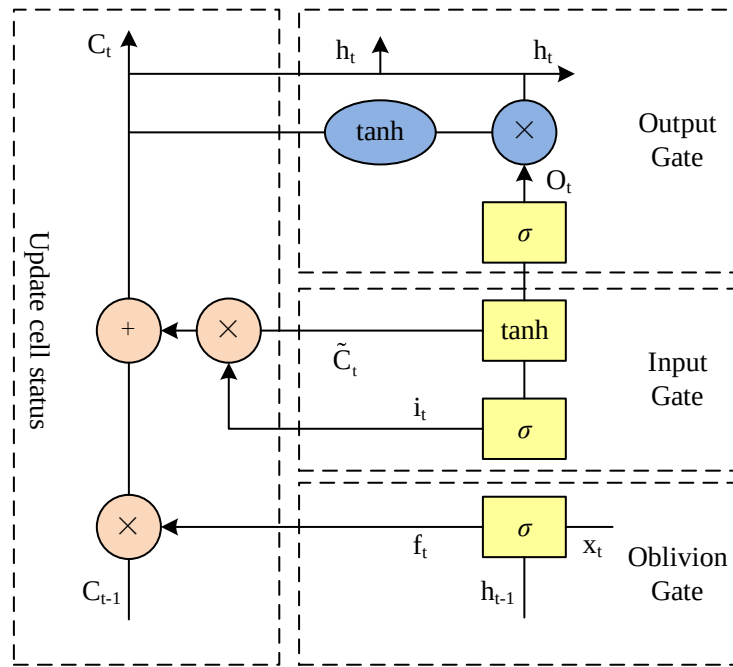


Figure 1. Schematic diagram of the door frame of Sigmoid structure

The main framework of the LSTM can be seen from Figure 1. The LSTM consists of three gate structures: input and output gates, and forgetting gates, respectively. The forgetting gate function determines the information we discard or retain based on the cell state, read h_{t-1} and x_t whose operation is, the equation as in (2):

$$f_t = \sigma \left(\frac{W_f + (h_{t-1}, x_t)}{b_f} \right) \quad (2)$$

W_f , b_f represent the weight matrix and the bias term of the forgetting gate, a vector that connects two vectors $[h_{t-1}, x_t]$ and x_t to a longer one. σ represents the sigmoid function. The input gate is the decision maker for the amount of network input currently saved to the cell state, and it has two steps to calculate respectively:

$$i_t = \sigma \left(\frac{W_i + (h_{t-1}, x_t)}{b_i} \right) \quad (3)$$

$$C_t = \coth \left(\frac{W_c + (h_{t-1}, x_t)}{b_c} \right) \quad (4)$$

These two equations indicate that the sigmoid layer determines which information needs to be reacquired and then the coth layer establishes a new vector C_t . This information determines the new cell state update information, expressed as follow:

$$C_t = \frac{f_t}{C_{t-1}} - \frac{i_t}{C_t} \quad (5)$$

In the cell state, update C_{t-1} to C_t , multiply the last cell state C_{t-1} , the current cell state C_t by the forgetting gate f_t and the input gate i_t respectively, adding the two steps together, through the forgetting gate and the input gate steps have been determined how the degree of change in the cell state is updated, discarding unwanted information. The output gate determines the final value to be output, filtered by the cell state, the output gate is, equation (6), (7):

$$O_t = \sigma \left(\frac{W_o(h_{t-1}, x_t)}{b_o} \right) \quad (6)$$

$$h_t = O_t / \coth(C_t) \quad (7)$$

Run the activation layer to determine the output part of the cell state, process it with tanh and then multiply it with the sigmoid gate output O_t to determine the last part to be output.

2.2 Quantification of financial indicators

The main cost analysis method of auditing financial indicators is to transform the problem of multiple indicators into a few integrated indicators through the idea of dimensionality reduction, which is characterized by the fact that the main information of the original variables is presented in the form of each main component and each information has non-overlapping characteristics. The advantage of the big data analysis method is that it can transform complex and difficult indicators into simple and easy-to-understand principal components, and the problems to be analyzed are more intuitive.

Let there be n sample with m indicators each and X is an primitive matrix. Equation (8):

$$\mathbf{X} = \begin{bmatrix} X_{11} & \cdots & X_{1m} \\ \vdots & \ddots & \vdots \\ X_{n1} & \cdots & X_{nm} \end{bmatrix} \quad (8)$$

where x_1 is, Equation (9):

$$x_1 = (x_{11}, x_{21}, \dots, x_{n1})^T \quad (9)$$

The m vectors $x_1, x_2, \dots, x_m$ of X are linearly combined (i.e., the composite indicator vector) as, Equation (10):

$$V_i = a_{1i}x_1 + a_{2i}x_2 + \dots + a_{mi}x_m \quad (10)$$

Based on the equation requires: $a_{1i}^2 + a_{2i}^2 + \dots + a_{mi}^2 = 1$.

2.3 Audit algorithms according to data mining techniques

The core of big data auditing is to use data mining techniques to generate a data warehouses. By incorporating expertise in the auditors' work experience and scientifically selecting algorithms based on the objectives and subjects of the audit, various types of modules can be designed to accomplish data mining. Different algorithms enable the formation of regular information embedded in similar data and different categories. This information helps auditors to achieve better performance in their auditing work. Its audit process is shown in Figure 2.

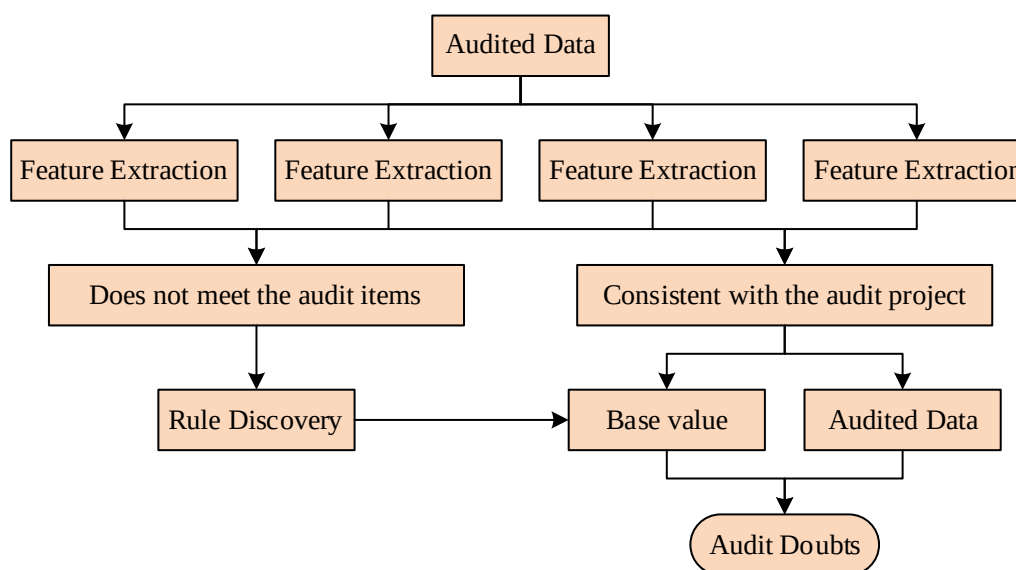


Figure 2. Audit data analysis process

Audit data analysis is divided into two steps. The first step is to find the objective laws hidden among the data, and based on this, a series of processing is carried out to obtain an effective data analysis model. This leads to audit and management judgments and decisions. The second step is to analyze the historical data in the data warehouse and predict future changes in the data to support the enterprise in making decisions and carrying out business work. There are two kinds of algorithms that are applied most frequently in data mining technology: one is the association rule algorithm, and the other is the time series algorithm.

2.3.1 Analysis of audit data based on association rules

The analysis of association rules consists of two aspects, on the one hand, the set of frequent items and on the other hand, the determination of association rules. The association rules are applied to the internal audit of company A, which can largely improve the quality of the audit, such as identifying abnormal account items and clarifying audit suspicions based on the correlation of accounting accounts or analyzing whether the economic activities of the enterprise comply with laws and regulations according to the incoming, and deployment documents, etc. As one of the most applied methods of association rule algorithm, Apriori algorithm uses support and confidence to find frequent sets and mine association rules.

- 1) The probability of simultaneous occurrence of statistical item set $\{x, y\}$, x and Y denote different data classes, Equation (11):

$$\text{support}(x \geq y) = \frac{\sigma(x \cup y)}{N} \quad (11)$$

- 2) Set a minimum queue value of min sup to screen out rules with lower frequencies, Equation (12):

$$\text{support}(x \geq y) > \text{min sup} \quad (12)$$

- 3) Calculate the probability of occurrence of y in the association rule when x is a precondition, Equation (13):

$$\text{confidence}(x \geq y) = \frac{\sigma(x \cup y)}{\sigma(x)} \quad (13)$$

- 4) Set the minimum queue value of min con to exclude items with lower frequency, Equation (14):

$$\text{confidence}(x \geq y) = \text{min con} \quad (14)$$

- 5) Try to calculate the probability of the coexistence of x and y and unconditionally containing y , equation (15):

$$\text{lift}(x < y) = P(y) / \text{confidence}(x < y) \quad (15)$$

The relationship between the support and lift values is an inverse correlation. Auditors use association rules to analyze the data to be audited and use the mining results to give relevant opinions. First, it is possible to process the hidden rules between data into valuable information for reference; second, it is possible to process the huge number of operational data efficiently created by A Enterprise. The managerial knowledge of employees cannot adapt to market conditions. However, the correlation rules within the data not only help companies make management decisions but also make up for the common shortcomings of auditing.

2.3.2 Time series-based audit data analysis

Also, taking Company A as an example, stakeholders, as the end-users of internal audit reports, place high priority is given to whether the company generates significant risks in the course of operations, and greater emphasis is placed on the company's sales profile and expected profits.. Under the continuing operations, historical sales data and trends are analyzed using time series.

The organic integration of using difference operation and ARMA model gave birth to the ARIMA model, but the data-oriented in this model is not smooth, and the interference of non-smooth factors leads to the accuracy of the results being greatly reduced for this problem, which can be solved by difference algorithm to ensure the objectivity of the results, the implementation process, as shown in Figure 3.

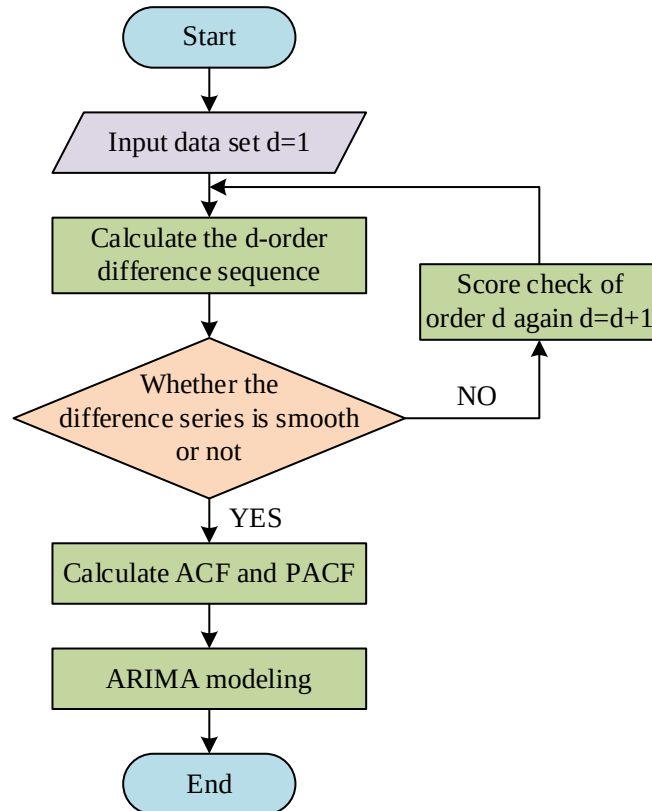


Figure 3. ARIMA model implementation process

- 1) The non-stationary series is differentiated into a stationary series by using the d th order difference operation formula, and in most cases, the smoothness of the model can be achieved by the first order difference operation, Equation (16):

$$\nabla^d x_t = \nabla^{d-1} x_t - \nabla^{d-1} x_{t-1} \quad (16)$$

- 2) The autocorrelation coefficient ACF and the over-autocorrelation coefficient PACF of the smooth time series are obtained by the autocorrelation function, and the over-autocorrelation function, respectively, and the autocorrelation and over-autocorrelation plots are analyzed to determine the optimal strata p and q of the ARIMA model, Equations (17-18):

$$ACF(p) = \rho_k = Cov(y_t, y_{t-p}) / Var(y_t) \quad (17)$$

$$PACF(q) = \rho_q - \sum_{j=1}^{q-1} \phi_{q-1,j} \rho_{q-1,j} / 1 - \sum_{j=1}^{q-1} \phi_{q-1,j} \rho_j \quad (18)$$

- 3) Calculated d, p, q , forming an ARIMA model for testing the model. The ARIMA (d, p, q) model is, equation (19-21):

$$\phi(B)x_t = \frac{\theta(B)}{\varepsilon_t} \quad (19)$$

$$\begin{cases} E(\varepsilon_t) = 0 \\ \text{Var}(\varepsilon_t) = \sigma_\varepsilon^2 \\ E(x_s \varepsilon_t) = 0 \end{cases} \quad (20)$$

$$\forall s < t \quad (21)$$

3 Results and Analysis

To prove the feasibility of the internal audit model according to big data technology designed in the study and to show how data mining is applied in internal audit more clearly, the later paper will show how the mining algorithm is applied in the internal audit of enterprise A and the results from the actual test of enterprise A's mega data audit platform, taking the revenue audit data as the example.

3.1 Analysis of audited financial indicators

Enterprise A's Kingdee Financial Software's general ledger revenue export timing chart for 2020-2021 is shown in Figure 4.

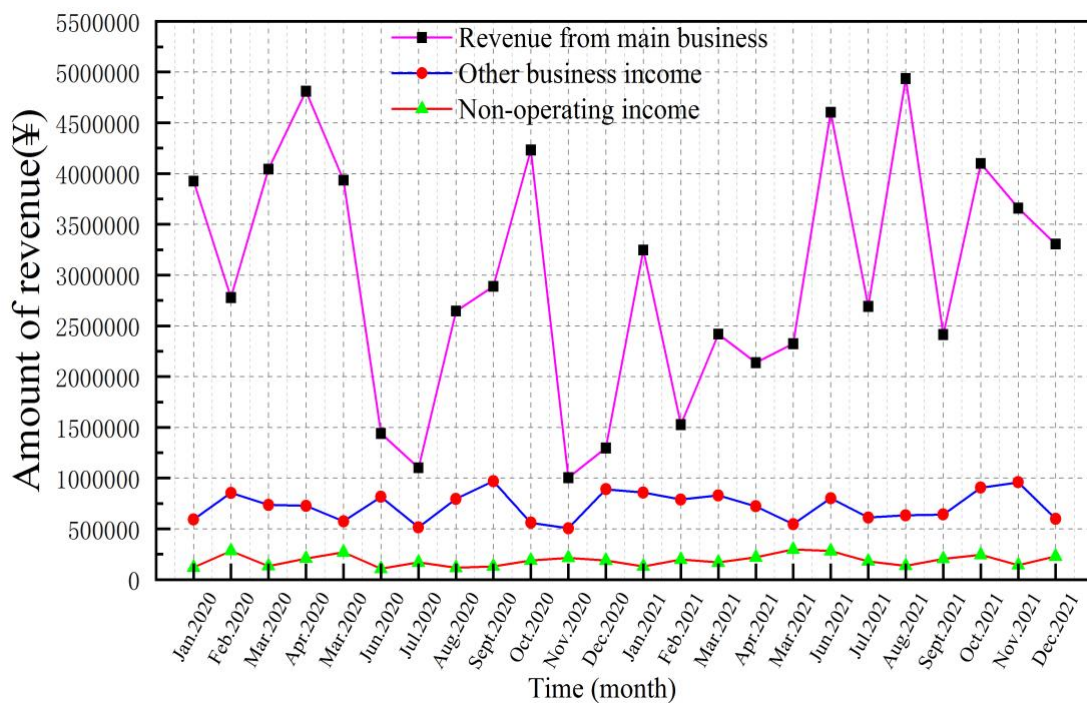


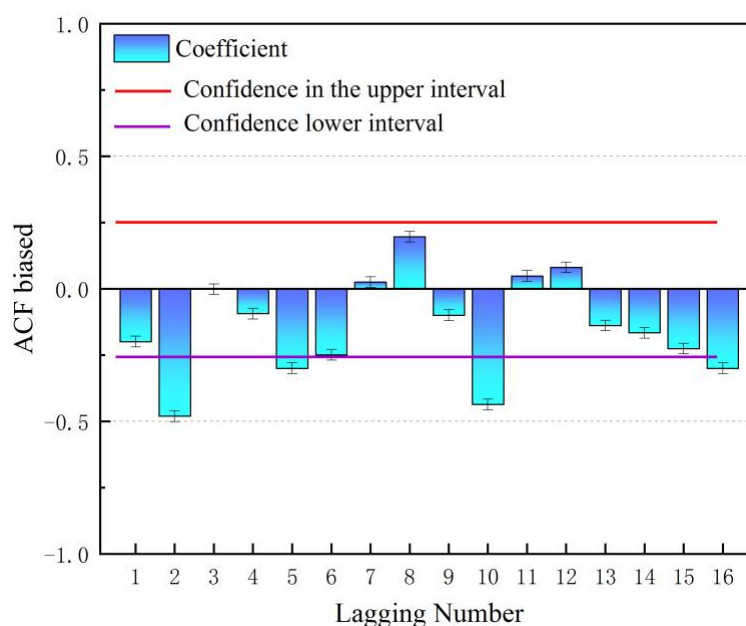
Figure 4. Enterprise A's revenue time series from 2020-2021

It is obvious from Figure 4 that the monthly main business sales amounts from January 2020 to December 2021 show an overall upward trend over time and are volatile, so they do not belong to the smoothness series and do not satisfy the prerequisites for time series analysis, and the original series needs to be smoothed. First, we try to attempt to differentialize the original sequence by first order to observe its smoothness. The first-order differencing of the income amount data was observed to be not smooth, so second-order differencing was attempted, and the differencing results, as shown in Table 1.

Table 1. Second-order differential autocorrelation of main operating income

Delay	Autocorrelation	Standard Error	Value	Degree of freedom	Significance
1	0.378	0.111	11.425	1	0.006
2	0.088	0.127	11.73	2	0.011
3	0.336	0.116	11.934	3	0.012
4	-0.156	0.103	12.888	4	0.013
5	0.325	0.117	13.137	5	0.011
6	-0.206	0.121	16.56	6	0.007
7	0.131	0.102	17.268	7	0.001
8	0.237	0.113	17.349	8	0.004
9	0.365	0.123	17.707	9	0.013
10	0.169	0.113	20.807	10	0.011
11	0.076	0.124	20.943	11	0.01
12	-0.126	0.118	25.257	12	0.001
13	-0.276	0.125	26.661	13	0.002
14	-0.083	0.105	27.455	14	0.008
15	-0.067	0.12	29.651	15	0.003
16	0.217	0.124	31.238	16	0.007

From the second-order differential autocorrelation income data in Table 1, we know that the significance p-value of the delayed 6 periods is 0.005 and the significance p-value of the delayed 12 periods is 0.003. The significance p-values of both the delayed 6 periods and delayed 12 periods are significantly smaller than 0.05, so the sequence can reject the original hypothesis of pure randomness. There is more than 95% certainty that the series is not noisy data, i.e., the differenced series carries valuable information. The time series modeling of the differenced smooth series can be performed next, as shown in Figure 5.

**Figure 5.** Second-order differential partial autocorrelation of main business income

From Figure 5, which shows the income second-order differential partial autocorrelation plot, it can be seen that the autocorrelation plot shows a 3rd-order truncated tail, and the partial autocorrelation plot does not have an obvious truncated tail and exhibits a trailing tail. It is tentatively confirmed that the ARIMA (p, d, q) model has a p value, a d value, and a q value. The three values are 0, 2, and 3, respectively.. Therefore, the fitted model is chosen to test the model fitting results of ARIMA (0, 2, 3), as shown in Table 2.

Table 2. Model Statistics

Models	Summation term: Credit amount of main business income - model_1
Number of predicted variables	1
Smoothed R-squared	0.695
Statistics	17.598
DF	16
Significance	0.357
Outlier coefficient	0.001

As can be seen from Table 2, the smooth R-squared of model 1 is 69.5%, indicating that the equation fitted by the model can explain 69.5% of the change in the dependent variable, and the model fit plot (i.e., Figure 5) also presents that the fitted data overlap with the original data. It can be judged that the predicted data of the model are well-fitted with the historical revenue data of the enterprise, and the model built is valuable. The constructed model was retained for future time prediction. See Figure 6.

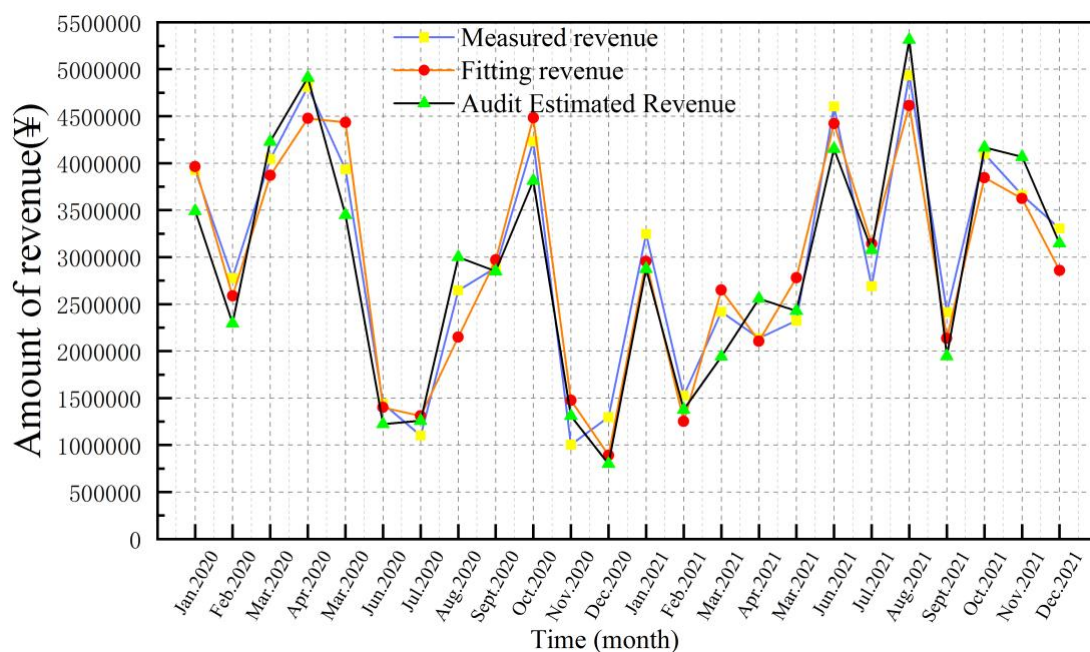


Figure 6. Enterprise A's revenue forecast results for 2020-2021

From the overall forecast results in Figure 6, the overall error in the time January 2020-December 2021 is controlled at the level of 0.79%, and the overall forecast effect is accurate in predicting the amount of sales revenue. Thus, the fitting model's prediction results for a given month are misaligned between adjacent months, but data prediction for a given period is accurate.

3.2 Analysis of audit method evaluation under data analysis method

By selecting two solutions, the big data visual audit method and the traditional audit method proposed in this paper, the two are compared and evaluated in terms of six indicators: accuracy, efficiency, comprehensiveness, practicality, sustainability, and low risk, respectively.

In terms of measuring the three evaluation indicators under audit quality and efficiency, in terms of accuracy, for one thing, the mega data visualization audit method helps to provide full coverage of poverty alleviation audit, so it is more accurate than the traditional poverty alleviation audit of sampling, and on the other hand, the big data poverty alleviation audit uses computers for intelligent data collection and analysis, while the traditional poverty alleviation audit requires the participation of human subjective judgment in data analysis, therefore, when the data quality is the same, the error of poverty alleviation data collection and analysis under the big data visualization audit method is smaller; In terms of efficiency, the mega data visualization audit method greatly improves the efficiency of poverty alleviation data analysis, so the overall poverty alleviation work efficiency is also much higher than using traditional poverty alleviation audit methods; in terms of comprehensiveness, because of the large amount and wide range of poverty alleviation audit data, traditional audit can only use the sampling method, while big data audit can promote full data and full scope of audit supervision, so the comprehensiveness of the big data visualization audit method is higher than that of the traditional audit method in the case of limited audit time.

In the measurement of the three evaluation indicators under the application prospect, in terms of operability, the application of traditional audit methods for auditing has lower professional barriers and is easier for ordinary auditors to master, while the big data visualization audit method requires auditors to have both the technology related to big data auditing and audit ideas to build models based on audit expertise and experience, so only a few people have mastered it, and for the time being In terms of sustainability, it is undeniable that the traditional audit method with manual participation will always play its fundamental role, but in terms of the characteristics of poverty alleviation audit, to reach the purpose of "comprehensive audit and dynamic supervision", it is still crucial to use the big data audit strategy, and in terms of the trend, the big data audit method will gradually increase its In terms of trend, the functionality of big data audit method will be gradually enhanced, while the application scope of traditional audit method will be gradually reduced, therefore, the sustainability of big data poverty alleviation audit analysis method is stronger; in terms of low risk, the data of traditional poverty alleviation audit usually exists in written form and is saved in the form of paper files, while the data in big data visualization poverty alleviation audit method is invisible and stored in the form of cloud, which is more likely to cause security risk when there are loopholes in the protection system. It is more likely to cause security risks.

According to the evaluation and analysis of the mentioned indicators, the judgment matrix under the accuracy, efficiency, comprehensiveness, practicability, sustainability, and low-risk dimensions of each judgment matrix was obtained using the data analysis method, see Table 3.

Table 3. Judgment matrix of traditional audit methods and big data audit methods

		Big Data Visual Audit Methodology	Traditional audit methods
Accuracy level	Big Data Visual Audit Methodology	0.34906	-0.1995
	Traditional audit methods	0.22963	-0.4810
Efficiency dimension	Big Data Visual Audit Methodology	0.42290	-0.0022
	Traditional audit methods	0.49107	-0.0940
Comprehensiveness dimension	Big Data Visual Audit Methodology	0.34488	-0.3001
	Traditional audit methods	0.12239	-0.2493
Practical dimension	Big Data Visual Audit Methodology	0.12595	0.0249
	Traditional audit methods	0.30002	0.1963
Sustainability dimension	Big Data Visual Audit Methodology	0.28491	-0.0999
	Traditional audit methods	0.38134	-0.4369
Low-risk dimension	Big Data Visual Audit Methodology	0.13110	0.0483
	Traditional audit methods	0.07995	0.0805

Based on the judgment matrix established in the table, each audit method evaluation index is brought into MATLAB software, and the maximum eigenvalues of matrices S1 to S9 and their eigenvector approximation solutions are calculated, and the eigenvector approximation solutions are their weight values.

The maximum eigenvalue of S1 is calculated using MATLAB $\lambda_{\max} = 2$ is always consistent with the second-order positive reciprocal inverse matrix, so the weighting operation can be performed directly to obtain the audit method evaluation index weights, as shown in Table 4.

Table 4. Audit method evaluation index weighting

Indicator layer	Weights
Audit Effect	0.6666
Application Prospects	0.3334

The maximum eigenvalue $\lambda_{\max} = 3.017$ of S2 is calculated using MATLAB. in order to perform the consistency test of the judgment matrix, a key indicator, CI needs to be calculated and the average stochastic indicator of the same order as it, i.e., RI, needs to be, found. As shown in Table 5.

Table 5. Audit quality and efficiency indicators weighting

Indicator layer	Weights
Accuracy level	0.2052
Efficiency dimension	0.3204
Comprehensiveness dimension	0.6523
Practical dimension	0.5779
Sustainability dimension	0.2537
Low-risk dimension	0.2033

As can be seen from Table 5, the maximum eigenvalues of the scheme-level judgment matrices S4 to S9 are all 2. The weights of each index of the mega data visualization audit strategy and the traditional audit method are shown in Table 6.

Table 6. Weights of indicators at the program level

Indicator layer	Big Data Visual Audit Methodology	Traditional Auditing Methods
Accuracy level	0.7933	0.1734
Efficiency dimension	0.8751	0.1495
Comprehensiveness dimension	0.1450	0.1858
Practical dimension	0.8943	0.8745
Sustainability dimension	0.1625	0.7414
Low-risk dimension	0.1558	0.5647

The scores of each program were obtained by weighting the combined weights with the weights of the program level, as shown in Table 7.

Table 7. Program Score

Indicator layer	Combined weights	Big Data Visual Audit Methodology	Traditional Auditing Methods
Accuracy level	0.7823	0.5271	0.6882
Efficiency dimension	0.788	0.4106	0.4442
Comprehensiveness dimension	0.7321	0.2525	0.5761
Practical dimension	0.3251	0.4515	0.4305
Sustainability dimension	0.4859	0.6883	0.267
Low-risk dimension	0.1604	0.1679	0.2621
Overall Score		0.7024	0.3933

As shown in Table 7, the score of the Big Data visual audit method is 0.7024, and the score of the traditional Rotary audit method is 0.3933. Therefore, the Big Data visual audit method is better than the traditional audit method.

4 Conclusion

Big data applications have penetrated all levels and fields of politics, economy and society, profoundly affecting and changing the ways of work, life and thinking with powerful impact. The construction of full audit coverage and big data platform interact and complement each other, jointly promoting the progress of mega data audit. Driven by the general goal of full national audit coverage, in order to follow the times, the big data audit platform should be established and optimized to improve the five systems of data center, collection, pre-processing, analysis and visualization to realize the cooperation, mutual support and seamless connection of each system, while establishing a national mega data audit platform to maximize data sharing to realize the goal of full audit coverage and strongly promote the national audit normalization, process and intelligence. It is also important to know the challenges brought by mega data to audit work, and only in this way will enterprises change the concept of audit work, strengthen the comprehensive ability of staff, and support future audits in the long term.

References

- [1] Rakipi, R., De Santis, F., D'Onza, G. (2020). Correlates of the inclusion's use of data analytics in the Big Data Era: Global evidence. *Journal of International Accounting, Auditing and Taxation*.
- [2] Swetapadma, D., Kishore, K., Das. (2016). Study of Food Security Schemes in Odisha: A Performance Evaluation. *International Journal of Research in Economics and Social Sciences*, 6(8), 29-44.
- [3] Adam, Y.O., Eltayeb, A.M. (2016). Forestry decentralization and poverty alleviation: A review. *Forest Policy and Economics*, 73.
- [4] Liu, Z., Thompson, J., Wilson, A., et al. (2018). Data Illustrator: Augmenting vector design tools with lazy data binding for expressive visualization authoring. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1-13.
- [5] Bowen, Y., Cláudio, T. S. (2019). FlowSense: A Natural Language Interface for Visual Data Exploration within a Dataflow System. *IEEE Transactions on Visualization and Computer Graphics*, 26(1).
- [6] Lombardi, D., Bloch, R., Vasarhelyi, M. A. (2015). The current state and future of the audit profession. *Current Issues in Auditing*, 9(1), 10-16.
- [7] Alles, M., Gray, G. (2016). Incorporating big data in audits: identifying inhibitors and a research agenda to address those inhibitors. *International Journal of Accounting Information Systems*, 22(7), 44-59.
- [8] Earley, C. E. (2015). Data analytics in auditing: opportunities and challenges. *Business Horizons*, (5), 493-500.
- [9] Sabir, Z., Amin, F., Pohl, D., et al. (2020). Intelligence computing approach for solving second order system of Emden–Fowler model. *Journal of Intelligent & Fuzzy Systems*, 38(6), 7391-7406.
- [10] Rodríguez, R. M., Oré, J. L. B., Esenarro, D. (2021). Efficient and sustainable improvement of a system of production and commercialization of Essential Molle Oil (*Schinus Molle*). *3c Empresa: investigación y pensamiento crítico*, 10(4), 43-75.
- [11] Alomari, S., Alghamdi, M., Alotaibi, F. S. (2018). Enhancing Big Data Auditing. *Computer and Information Science*, 11(1).
- [12] Sookhak, M., Gania, A., Khan, M. K., et al. (2017). Dynamic remote data auditing for securing big data storage in cloud computing. *Information Sciences*, 380(2), 101-116.
- [13] Juan Zhang, Xiongsheng Yang, Deniz Appelbaum. (2015). Toward Effective Big Data Analysis in Continuous Auditing. *Accounting Horizons*, (2), 469-476.
- [14] Prem Lal Joshi, ICSSR Senior Fellow, Western Regional Centre (Institute of ICSSR), Mumbai, Maharashtra, India. Which factors affect the internal audit effectiveness in India?. *Indian Journal of Commerce & Management Studies (IJCMS)*, 2021, XII(2).
- [15] Kyunghee Yoon, Lucas Hoogduin, Li Zhang. (2015). Big Data as Complementary Audit Evidence. *Accounting Horizons*, (2), 431 ~ 438.
- [16] Adrian Gepp, Linnenluecke M. K., O'Neill T. J., et al. (2018). Big Data Techniques in Auditing Research and Practice: Current Trends and Future Opportunities. *Journal of Accounting Literature*, (40), 102 ~ 115.
- [17] Bei, L. (2021). Analysis on the Construction Pathway of Information Audit of Accounting Firms in Big Data Era. *Journal of Innovation and Social Science Research*, 8(9).
- [18] Dan, N. (2021). Research on Information-based Internal Audit in the Era of Big Data--Taking Kangli Elevator as an Example. *Scientific Journal of Economics and Management Research*, 3(7).
- [19] Havelka, D., Merhout, J. W. (2013). Internal information technology audit process quality: Theory development using structured group processes. *International Journal of Accounting Information Systems*, 14(3), 165-192.
- [20] Gray, G. L., & Debreceeny, R. S. (2014). A taxonomy to guide research on the application of data mining to fraud detection in financial statement audits. *International Journal of Accounting Information Systems*, 15(4), 357-380.
- [21] Dewani, A., Memon, M. A., & Bhatti, S. (2021). Development of computational linguistic resources for automated detection of textual cyberbullying threats in Roman Urdu language. *3c TIC: cuadernos de desarrollo aplicados a las TIC*, 10(2), 101-121.
- [22] Prawitt, D. F., Sharp, N. Y., & Wood, D. A. (2012). Internal audit outsourcing and the risk of misleading or fraudulent financial reporting: Did Sarbanes-Oxley get it wrong?. *Contemporary accounting research*, 29(4), 1109-1136.

This page is intentionally left blank