

Audit Data and Management Platform Based on Intelligent Algorithm

Yiqi Yang

Hunan Agricultural University, Changsha, Hunan, China
Hauyangyiqi@163.com

Abstract: Due to the uncertainty of the external economic environment and the changes caused by new technologies in the new era, there are still many deficiencies in the current financial statement audit. The use of new technology can optimize the audit process, reduce audit practice risks and improve audit efficiency. Therefore, it is necessary to study the intelligent audit of financial statements. The purpose of this paper is to study the audit data analysis and management platform based on intelligent algorithms. This paper focuses on the construction purpose and content of the audit data analysis and management platform, and analyzes the specific requirements of some functional modules in the platform according to the audit workflow and business requirements. At the same time, this paper also describes the overall design architecture of the platform, which mainly uses the isolated point mining method for enterprise intelligent audit. This paper takes it as an effective method to analyze the problems existing in the internal audit of enterprises, and constructs the audit model of benefit-cost matching, and verifies the effectiveness of the algorithm through an example. There are four isolated points in the gross profit rate of catering of an enterprise.

Keywords: Intelligent Algorithm; Audit Data; Analysis Management; Management Platform.

1. Introduction

In the audit of financial statements, there are a lot of simple, trivial and repetitive work, which are all done manually [1]. With the development of modern technology, the scale of the company is continuously expanding, and the business scope of the company is also constantly expanding, which has virtually increased the workload and audit time of financial statement auditors, leading to the decline of audit efficiency. In this simple, trivial and repetitive work, proper use of intelligent technology to replace manual work can effectively improve the quality of audit work; Secondly, when the audit is conducted at the end of the year, the company does not need to hire many interns to do some simple and repetitive work, which can reduce the company's operating costs; Using intelligent technology can enable auditors to get rid of tedious work and engage in more high-tech auditor work. Intelligent audit is not only a combination of technology and methods, but also an innovation of audit ideas [2-3]. This not only provides new development opportunities for auditors, but also further promotes the theory and practice of auditors. In the audit, business intelligence is applied to the audit, using the extraction of business data, real-time response to data changes in the data warehouse, and various data analysis methods to analyze the business data of enterprises, so as to achieve the purpose of monitoring economic activities [4]. In this way, the post-audit can be transformed into prior and in-process supervision, the centralized audit can be transformed into real-time audit, and the risk warning can be implemented in the audit process, so as to improve the efficiency of audit [5].

With the continuous growth of the content and business data of the communication enterprise audit, the traditional audit method is mainly offline audit, the audit work is heavy and temporary, the number of auditors is scarce, and the support of data market and analysis model is lacking. Andre é M. Carrington put forward an efficient management method based on data domain of multiple systems in view of various problems faced by traditional audit methods. On this basis, he

defined the data of multiple systems uniformly and defined it as a unified database. At the same time, the platform also has certain scalability, which can gradually achieve the coverage of data, and can provide support for more management areas and business systems. On this basis, he uses the K-means method to find the center point of a group of continuous attribute clusters, and fuzzies it with triangular fuzzy numbers. On this basis, combined with the minimum fuzzy degree theory, a visual fuzzy decision system is established [6]. The existing audit big data analysis methods mainly include the lack of mining the audit information hidden in the big data, the detection of more fraud technologies in the context of informatization and networking, and the quality defects of the audit big data. The association analysis algorithm in data mining technology is an effective method to extract knowledge from a large amount of data. It can complete, hide, clean up and reduce defect data, and analyze these data with the help of experts or analysts, so as to obtain the corresponding audit results. Based on this, Fouad Ben Abdelaziz first reviewed the world's relevant literature, and then reviewed the world's relevant literature; This paper introduces the background, current situation and challenges of relevant research in the world; It mainly expounds some basic principles and methods of data mode, mode transformation, mode establishment, etc; Be responsible for the collection and arrangement of audit data; Pre-process the audit data and explain the algorithm; A new correlation analysis method is proposed. The activation value of hidden nodes and the extraction of audit rules are analyzed; On this basis, an audit data processing method based on association analysis is presented; Finally, the quality of audit data and audit risk are discussed. This paper summarizes the current research situation in the field of audit data processing in the world. The results show that, on the basis of the analysis of association rules, he proposed a correlation analysis method based on association rules, which can identify a large amount of audit data, and can use probabilistic methods to describe the occurrence of association events at the same time or

successively [7].

This paper discusses the current situation of the audit data analysis and management platform and the main problems faced by auditors, and introduces the business intelligence technology and data mining programming technology, aiming at clarifying the actual needs of building the audit data analysis and management platform and the feasibility of analyzing and developing software systems. Then, this paper proposes a financial audit mining model based on outlier mining algorithm, and gives the concrete implementation method of the model. By analyzing the theory of business intelligence and studying the application of relevant technologies, applying business intelligence technology to complex financial data audit and establishing an audit data analysis and management platform can make auditors more convenient to audit and reduce the workload.

2. Audit Data and Management Platform Based on Intelligent Algorithm

2.1. Online Analytical Processing Technology

On-line Analytical Processing (OLAP) is a very critical analysis method in data warehouse. Online analysis and processing is a complex analysis method based on large-scale data. It can analyze and process data in data warehouse in multi-dimensional and efficient ways from multiple perspectives [8-9].

The main feature of OLAP is that it can respond quickly to various complex queries, and has the ability of multidimensional modeling. Its main technologies include: multidimensional data model, multidimensional analysis, multidimensional query and display, data cube, etc. [10-11]. The main characteristics of online analytical processing are: fast speed, and the ability to respond to the analysis requests of most users in very little time; Users do not need a program to determine specific new operations in the way that users expect, and perform different logic and statistical analysis related to the application in the way that users expect; Multi-dimensional, providing users with multi-dimensional observation and analysis; Timeliness, no matter how large the number of data is or where it is stored, you can get timely information and manage these large amounts of information [12-13]. The use of OLAP technology has greatly expanded the ability of data warehouse to process data, but in this era of information explosion, the limitations of OLAP technology have also become increasingly prominent. In order to better adapt to data processing and analysis, people have begun to conduct in-depth research on data mining technology [14-15].

2.2. Outlier Algorithm

(1) Outlier definition

Definition: In the S set, only when the distance between the object with p in S is larger than d, will an isolated point appear [16].

(2) Measurement of distance

In distance-based isolation point recognition methods, absolute value, Euclidean value and Chebyshev value are usually used. The so-called Manhattan distance is a geometric academic term used to express the sum of absolute lengths between two coordinates [17]. Its definition is:

$$dij = \sum_{k=1}^m |xik - xjk| \quad (1)$$

In the formula, m is the number of attributes of the data object, and xij is the value of the jth attribute of the i-th object [18].

Euclidean distance, also known as Euclidean distance, refers to the real distance between two points on two dimensions. The formula is:

$$dij = \sqrt{\sum_{k=1}^m (xik - xjk)^2} \quad (2)$$

(3) Distance-based outlier detection algorithm

The distance-based outlier detection algorithm is mainly used to calculate the distance between each object and other objects based on the number of outliers that users want to find before this, and then calculate the distance between each object and all other objects, so that the M objects with the largest distance sum are outliers [19].

Definition: xi and xj represent the objects in a data set X, n represents the number of objects, dij represents the spacing of xi and xj, then the spacing matrix R of X is as follows:

$$R = \begin{bmatrix} d11 & d12...d1n \\ d21 & d22...d2n \\ \dots & \dots \\ dn1 & dn2...dnn \end{bmatrix} \quad (3)$$

Definition: xi is the ith data object in the dataset, and the deviation degree Devi of xi is defined as:

$$Devi = \sum_{j=1}^n dij \quad (4)$$

That is, Devi is the sum of row i in matrix R.

From the definition, we can see that the larger Devi is, the more likely object i is to become an isolated point. If M is the number of outliers the user wants to obtain, then the largest outlier is outliers [20].

2.3. Overview of Platform Requirements

The functional requirements of the audit data analysis and management platform are as follows:

(1) The platform uses the data collection server to collect data related to the budget implementation of the financial department on a regular basis. It can also collect data in other electronic formats, and automatically schedule data processing templates according to the audit data plan. The collected financial data is processed and converted into data in standard format;

(2) The model can be used to monitor all businesses and financial data in the financial information system in real time, and also to set up auditors appropriately, so as to achieve the early warning function of financial data;

(3) Based on the project implementation plan of the financial budget implementation audit, the auditors select a corresponding audit mode for the corresponding audit items. This platform will automatically extract the basic data of the audit background, so as to realize multidimensional and multi-level audit query analysis.

(4) This platform will feed back the audit questions analyzed by each module to the auditors, who will further implement the audit questions after the analysis.

3. Investigation and Audit Data and Management Platform Based on Intelligent Algorithms

3.1. Development Language and Tools

(1) Development language

The languages needed for the development of the system include C #, Microsoft, SQL, MDX, DMX, and ADOMD NET API, etc. In DMX, you can complete the structure of data mining, establish data mining models, test models and forecast data, and combine DMX language with practical applications. NET API interface is used to connect with analysis services server, and MDXprediction statement is used to predict data based on data mining. MDX provides the ability to process multi-dimensional data. It can embed MDX declarations into applications, enabling them to interact with multi-dimensional databases, and users can perform corresponding conditional retrieval according to their own needs and needs. The results will be displayed in a chart or table.

(2) Development tools

SQL SERVER 2008 is not only a relational database and data warehouse server, but also an Analysis Services server. Through the connection with SQLSERVER2008, users can use the software to mine, predict and analyze data in multi-dimension.

3.2. Audit Data and Management Platform Architecture

The audit data analysis and management platform include data warehouse, multidimensional analysis (OLAP) and data mining technologies. With these technologies, data can be collected and integrated, and data can be analyzed, so as to make a reasonable assessment and prediction, and provide a basis for auditors to make decisions. OLAP is a kind of confirmatory analysis. It can analyze multidimensional data sets based on data warehouse at different levels and dimensions, so that auditors can have a comprehensive understanding of the overall situation, observe the development trend, and select the key points. You can use drilling to find some clues before you can conduct an audit. Data mining is a kind of mining analysis. It uses various mining algorithms to let auditors extract the information and knowledge hidden in the massive financial data. Data mining improves the application of data from low-level and simple queries to mining more knowledge from data to help people make decisions. The audit data analysis platform in this paper mainly includes four modules: audit data collection, audit analysis, suspicious problem handling, and audit collaboration. The overall functional framework is shown in Figure 1:

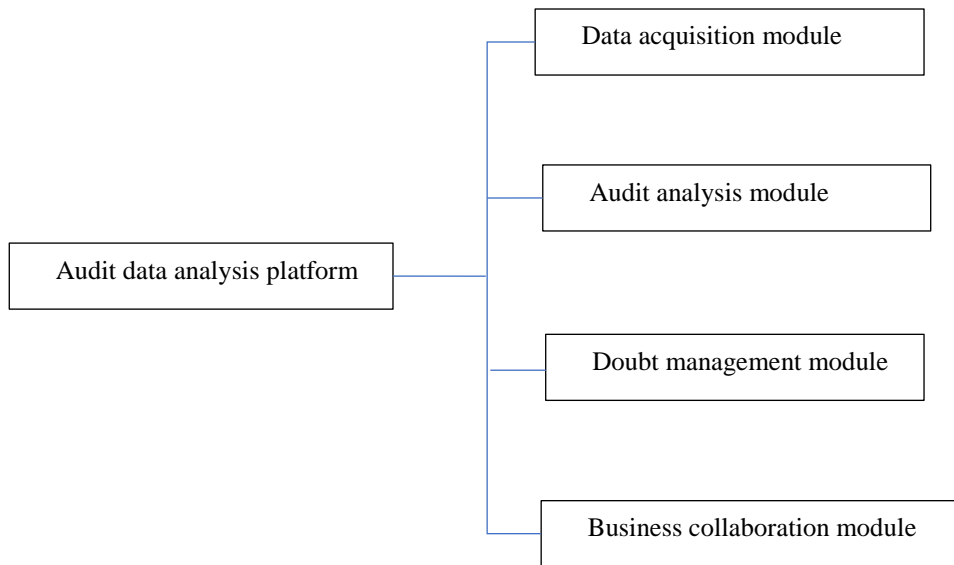


Figure 1. Platform functional architecture

3.3. Financial Audit Mining Model Based on Outlier Mining Algorithm

During the audit process, outlier mining technology can be used to obtain abnormal data, which is very important for the audit work and may be an important breakthrough. During the audit process, the specific model of outlier mining technology for anomaly detection is shown in Figure 2.

(1) Determine audit requirements: analyze the characteristics of the audited unit according to the assigned audit tasks and requirements, clarify the audit direction and audit focus, understand the content to be audited, and transform it into an expression that can use outliers for data mining.

(2) Data understanding and preparation: understand and analyze the data structure of the audited entity, select the appropriate data source according to the needs of the audit business, identify the data that can be analyzed using outliers, transform them, and load them into the data warehouse for mining.

(3) Modeling and outlier mining: on this basis, the proposed outlier mining method is modeled.

(4) Evaluation: After building the model, we should analyze and evaluate the operation results of the model, and test and evaluate the knowledge and rationality of the operation results. If the results cannot accurately reflect the matters to be audited, then we need to revise the model several times until the model can meet the needs of the auditors.

(5) Release: The results of mining should be presented to

auditors in an intuitive, concise and easy-to-understand way through visual technology, so that auditors can quickly

determine audit clues and make audit judgments.

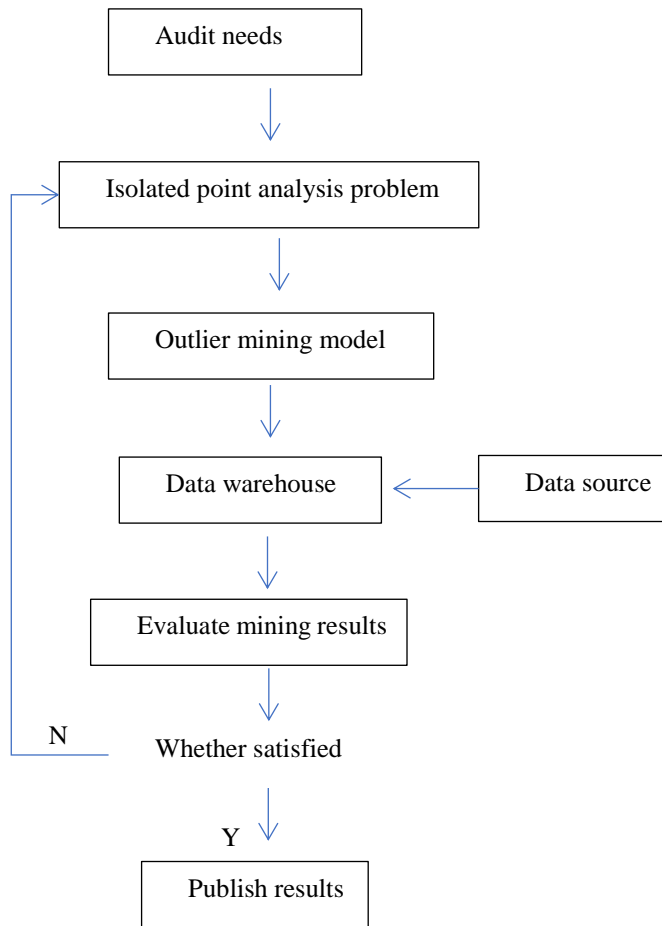


Figure 2. Financial audit mining model based on outlier mining algorithm

4. Audit Data and Management Platform Based on Intelligent Algorithm

4.1. Implementation of Audit Function

(1) OLAP-based audit process

Create a database for analysis: create a database in the Analysis Server for analysis, and connect it with the source database by ODBC (Open Database Connectivity) or OLEDB (Object Linking and Embedding, Database), and use it as the data source of the Analysis database.

Create Dimension: in Shared Dimension, create a dimension for analyzing data through the Dimension Wizard or Dimension Editor.

Create a cube: Create a cube through the Cube Wizard or Cube Editor. After building a cube, you can use the analysis manager to browse the data directly, use Excel or Web as the front-end analysis tool to analyze the data, and develop client applications to analyze and process the data.

(2) OLAP analysis

OLAP has effectively processed complex analysis operations, and its focus is on assisting decision-making, while providing users with an intuitive and easy-to-understand retrieval result. In programming, its implementation method is as follows:

```
using Microsoft.AnalysisServices.AdomdClient;
```

```
using System.Data.SqlClient;
connection=new AdomdConnection();
connection.ConnectionString="Data Source=localhost;
Initial
Catalog=Data Warehouse ";
connection. Open ();
AdomdCommand cmd=connection.CreateCommand();
Cmd. CommandText=MDX query statement;
//execute the command and display the prediction result
AdomdDataReader reader=cmd.ExecuteReader();
if (reader. Read ())
string predictedGeneration=reader. GetValue
(0).ToString();
(3) Report management
```

The report is a kind of XML (Extensible Markup Language) format data. It needs to go through the steps of variable setting, cell editing, SQL statement setting, file parsing and so on to realize the display of the query results and tables.

4.2. Data Mining Model and Data

In this paper, the revenue and cost of a company's restaurant are extracted from a database. After conversion, cleaning, verification and loading into the data warehouse, the profit of the restaurant can be obtained. In order to facilitate calculation, we use a positive exponential linear normalization method to normalize these data. The formula is:

$$y_{i,j} = \frac{x_{i,j} - \min(x_j)}{\max(x_j) - \min(x_j)} \quad (5)$$

The two-dimensional scatter diagram of the obtained data is shown in Figure 3:

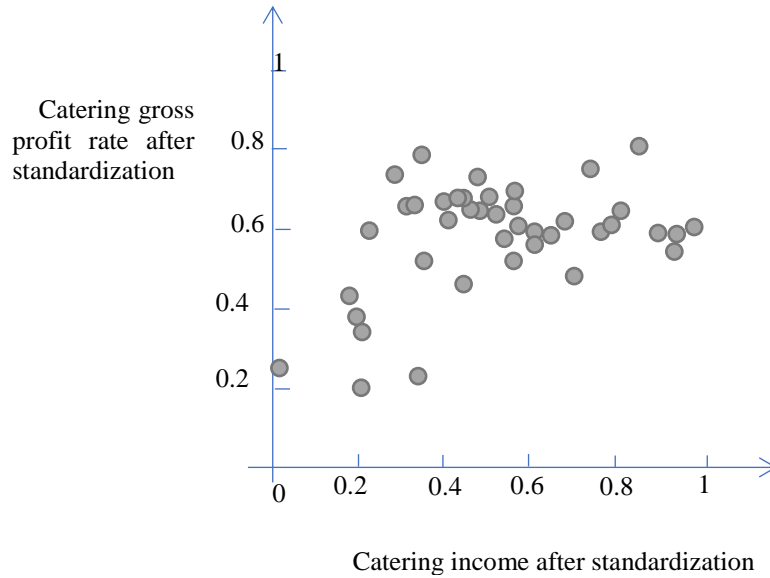


Figure 3. Scatter Chart of Standardized Catering Revenue and Catering Gross Profit Rate

The x-axis is the food and beverage income after standardization, and the y-axis is the food and beverage gross profit rate after standardization. Using the above distance-based outlier algorithm, a total of four outliers have been mined, as shown in Table 1 and Figure 4.

Table 1. Results of outlier mining

Date	Income after standardization	Catering gross profit rate after standardization
November 2022	0.9	0.81
December 2022	0.05	0.21
January 2023	0.2	0.2
February 2023	0.33	0.25

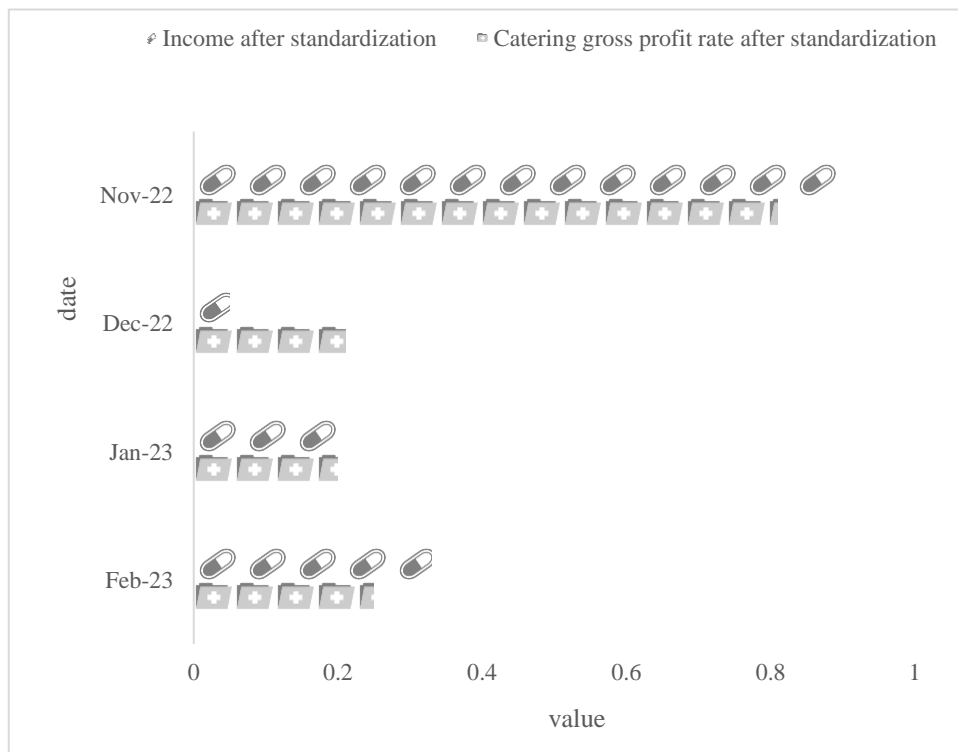


Figure 4. Mining results of distance-based outlier algorithm

Through the analysis of these three isolated points, it can be seen that three of these three isolated points are because

the operation is in the off-season and the marketing strategy of reducing prices to attract customers has been adopted, resulting in a relatively low gross profit rate of catering. Another isolated point is that part of the revenue of the previous month was deferred to November 2022, resulting in a relatively high income in November 2022 and a relatively high gross profit rate. To this end, we submitted an audit report to the company.

5. Conclusions

Nowadays, AI, big data and other new technologies have been widely used, and the application of new technologies in audit analysis is also imminent. The audit data analysis and management platform studied in this paper has promoted the informatization of the audit work to a certain extent, thus further improving the quality and efficiency of the audit work, and the audit method has also become more convenient and flexible. However, we also found that there are still some defects in the platform itself and its use. Therefore, we have some prospects for the future development and use of the platform: on this basis, we will further strengthen the security of the platform's network and data. In order to enhance the prevention of network and data security risks, the data of the platform can be synchronized and hot standby by two computers, and the audit data analysis and management platform can be run using the government intranet or private network, which has stronger confidentiality and higher security performance. At the same time, a complete emergency plan should be prepared to respond quickly and efficiently in case of safety problems. If possible, you can also find a professional data confidentiality engineer to achieve professional protection. Therefore, in the follow-up work, we should do a good job in the network and data security related to the platform.

References

- [1] Joseph Johnston, Joseph H. Zhang: Auditor Style and Financial Reporting Similarity. *J. Inf. Syst.* 35(1): 79-99 (2021)
- [2] Diane J. Janvrin, Maureen Francis Mascha, Melvin A. Lamboy-Ruiz: SOX 404(b) Audits: Evidence from Auditing the Financial Close Process of the Accounting System. *J. Inf. Syst.* 34(3): 77-103 (2020)
- [3] Jakob Mökander, Maria Axente: Ethics-based auditing of automated decision-making systems: intervention points and policy implications. *AI Soc.* 38(1): 153-171 (2023)
- [4] Inioluwa Deborah Raji, Joy Buolamwini: Actionable Auditing Revisited: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products. *Commun. ACM* 66(1): 101-108 (2023)
- [5] R. Hariharan, G. Komarasamy, S. Daniel Madan Raja: Implementation of ID-based Audit Protocols to Enhance Security and Productivity. *Comput. Syst. Sci. Eng.* 46(1): 873-882 (2023)
- [6] Andre M. Carrington, Douglas G. Manuel, Paul W. Fieguth, Tim Ramsay, Venet Osmani, Bernhard Wernly, Carol Bennett, Steven Hawken, Olivia Magwood, Yusuf Sheikh, Matthew McInnes, Andreas Holzinger: Deep ROC Analysis and AUC as Balanced Average Accuracy, for Improved Classifier Selection, Audit and Explanation. *IEEE Trans. Pattern Anal. Mach. Intell.* 45(1): 329-341 (2023)
- [7] Fouad Ben Abdelaziz, Souhir Neifar, Khamoussi Halioui: Multilevel optimal managerial incentives and audit fees to limit earnings management practices. *Ann. Oper. Res.* 311(2): 587-610 (2022)
- [8] Oscar Danilo Gavilánez Alvarez, Glen Dario Rodriguez Rafael: VoIP security auditing model based on COBIT 4.1. *Int. J. Secur. Networks* 17(2): 63-76 (2022)
- [9] Sergio Davalos, Ehsan H. Feroz: A textual analysis of the US Securities and Exchange Commission's accounting and auditing enforcement releases relating to the Sarbanes-Oxley Act. *Intell. Syst. Account. Finance Manag.* 29(1): 19-40 (2022)
- [10] Dimitrios Kydros, Michail Pazarskis, Athanasia Karakitsiou: A framework for identifying the falsified financial statements using network textual analysis: a general model and the Greek example. *Ann. Oper. Res.* 316(1): 513-527 (2022)
- [11] Hyun Woong (Daniel) Chang, Steven Kaszak, Peter C. Kipp, Jesse C. Robertson: The Effect of iXBRL Formatted Financial Statements on the Effectiveness of Managers' Decisions When Making Inter-Firm Comparisons. *J. Inf. Syst.* 35(2): 149-177 (2021)
- [12] David Broneske, Anna Drewes, Bala Gurumurthy, Imad Hajjar, Thilo Pionteck, Gunter Saake: In-Depth Analysis of OLAP Query Performance on Heterogeneous Hardware. *Datenbank-Spektrum* 21(2): 133-143 (2021)
- [13] Bhagyaraj Thangavel, Sabarathinam Srinivasan, Kathamuthu Thamilaran, Guisheng Zhai, Nallappan Gunasekaran: Dynamical Analysis of T-S Fuzzy Financial Systems: A Sampled-Data Control Approach. *Int. J. Fuzzy Syst.* 24(4): 1944-1957 (2022)
- [14] Margareta Gardijan, Zrinka Lukac: The financial efficiency of small food and drink producers across selected European Union countries using data envelopment analysis. *Eur. J. Oper. Res.* 291(2): 586-600 (2021)
- [15] Emna Mnif, Anis Jarboui, M. Kabir Hassan, Khaireddine Mouakhar: Big data tools for Islamic financial analysis. *Intell. Syst. Account. Finance Manag.* 27(1): 10-21 (2020)
- [16] Amrita Chatterjee: Financial inclusion, information and communication technology diffusion, and economic growth: a panel data analysis. *Inf. Technol. Dev.* 26(3): 607-635 (2020)
- [17] Manas Tripathi, Arunabha Mukhopadhyay: Financial Loss due to a Data Privacy Breach: An Empirical Analysis. *J. Organ. Comput. Electron. Commer.* 30(4): 381-400 (2020)
- [18] Dat Thanh Tran, Alexandros Iosifidis, Juho Kannianen, Moncef Gabbouj: Temporal Attention-Augmented Bilinear Network for Financial Time-Series Data Analysis. *IEEE Trans. Neural Networks Learn. Syst.* 30(5): 1407-1418 (2019)
- [19] Rui Pedro Brito, Pedro Júdice: Asset classification under the IFRS 9 framework for the construction of a banking investment portfolio. *Int. Trans. Oper. Res.* 29(4): 2613-2648 (2022)
- [20] Ahmed Jasim Hameed, Anfal S. Shareef, Sameer Imad Shaban: Effect of Fair Value Based on IFRS 13 on the Qualitative Characteristics of Accounting Information: An Exploratory Study in the Iraqi Environment. *J. Cases Inf. Technol.* 24(2): 1-12 (2021).