



Jurnal Riset Akuntansi dan Keuangan

Journal homepage: <https://ejournal.upi.edu/index.php/JRAK/>



The Role of Big Data in Financial Reporting Quality: A Bibliometric Analysis

Panisha Mustikarini¹, Indira Januarti²

Faculty of Economics and Business, Diponegoro University

*Correspondence: E-mail: panishaaaa23@gmail.com

| ABSTRACT | INFO ARTIKEL |
|--|---|
| <p>This research focuses on the possibilities of Big Data Analytics (BDA) to increase greatly financial reporting quality (FRQ) and develop the knowledge base and conceptual framework inside this field of activity. The researcher investigated publications, including authors, journals, keywords, co-citations, thematic development, and topic trends, using a bibliometric method and the Biblioshiny tool. According to the analytical results, five primary clusters of profit management detection, financial statement manipulation theory, corporate governance, machine learning, and data mining techniques define the use of Big Data Analytics in the Financial Reporting Quality study. The key findings reveal that data mining and machine learning are the most often utilized and successful approaches for identifying deceptive behavior in financial reporting. Moreover, narrative report analysis helps BDA to enhance detecting skills. Artificial intelligence, deep learning, and machine learning are rising technologies projected to be the key subjects of future financial reporting quality studies.</p> <p>© 2025 Kantor Jurnal dan Publikasi UPI</p> | <p>Article History: <i>Submitted/Received 6 June 2025</i> <i>First Revised 10 June 2025</i> <i>Accepted 27 June 2025</i> <i>First Available online 1 August 2025</i> <i>Publication Date 7 August 2025</i></p> <hr/> <p>Keyword: <i>big data analytics (BDA); bibliometric analysis; deep learning; financial report quality (FRQ); mining of data.</i></p> |

1. INTRODUCTION

Big Data (BD) helps companies spot possibilities, learn from past mistakes, increase openness and transparency, and support innovation, efficiency, and competitiveness (Warren et al., 2015). Businesses must use analytical techniques and approaches to analyze Big Data (BD) (Inamdar et al., 2020). Businesses must employ analytical techniques and methods to analyze Big Data (BD) to gain these advantages. Big data analytics, or BDA, is the procedure for cleaning, evaluating, analyzing, influencing, and modeling data to extract important insights, identify patterns and trends, draw conclusions, and guide smarter decisions (Blazquez and Domenech, 2018). Intelligent machines, learning through machines, neural networks, and text mining, statistics, and databases are just a few of the approaches significant data analytics uses (Perols et al., 2017). Using real-time data analysis and processing, big data analytics (BDA) presents innovative ideas that can help the accounting profession by offering high-quality information quickly (Rusydziana et al., 2021).

Big Data Analytics (BDA) methods have been investigated in many studies to determine their advantages for companies, auditors, and other stakeholders using BDA approaches (Aboelfotoh et al., 2024). By creating plans and activities that increase corporate value, Big Data Analytics (BDA) applied in management accounting can help a company to operate better over the long run (Abdelhalim, 2024). On the other hand, recent studies have made growing use of BDA techniques to predict and spot fraud (Tang and Karim, 2019). Approaches for Identity, detection, and prediction of many kinds of fraud in financial reporting have been developed using machine learning (ML), data mining (DM), and deep learning (DL) algorithms known as Big Data Analytics (BDA) techniques (Aftabi et al., 2023; Hajek & Henriques, 2017; Omar et al., 2017)

Reliable and accurate information about the financial situation and state of a company comes from financial reporting. Reports have to satisfy quality standards so that stakeholders may assess the expected cash flows (Framework, 2018) and make informed investment decisions. Particularly expected cash flows, financial report quality (FRQ) is the accuracy of the reporting process in conveying information about the operations of an organization. Financial reporting fraud (FRF) (Azzam et al. 2024) Financial reporting fraud (FRF) results from corporations purposefully presenting significantly erroneous information or omitting important disclosures, therefore leading to mistaken decisions and possible crises. Big Data Analytics, however, play a major part in spotting such fraud and gives hope on the accuracy of financial statements. Financial reporting fraud (FRF) is purposefully distortion of financial statements meant to fool interested parties. Undetected financial reporting fraud (FRF) can cause major losses for stakeholders, therefore influencing both honest businesses and those who have dishonesty (Mandal et al., 2023) Through management of results, companies often engage in financial reporting fraud (FRF) (Tang and Karim, 2019).

Low-quality financial reports have significant adverse effects on stakeholders, so research on financial reporting quality (FRQ) employing big data analytics (BDA) approaches is necessary. Financial reporting quality (FRQ) (Teixeira & Rodrigues, 2022) is evaluated in this study using earnings management (EM) and financial reporting fraud (FRF). Given the relevance of BDA in evaluating FRQ and the dearth of thorough literature reviews on the topic, this paper aims to investigate the literature on Big Data Analytics (BDA) and Financial Reporting Quality (FRQ) with an eye toward use of BDA techniques in the investigation of Financial Reporting Fraud (FRF) and Earnings Management (EM). The research questions (RQs) are meant to direct our investigation and involve the reader in learning.

RQ.1 Which writers, sources, and countries in the field of research are most prolific?

RQ.2. Which nations and pieces of work most affect the research domain?

RQ.3: Which research domain terms do authors frequently use?

RQ. 4: What intellectual framework and research knowledge base exists?

RQ.5: Which basic patterns and conceptual framework apply in the topic of research?

RQ.6. How have Big Data Analytics (BDA) approaches and major challenges affected the themes in this field over time?

RQ.7: What holes in the body of current research have been found that might point future research paths?

A comprehensive bibliometric analysis was carried out to systematically examine 360 peer-reviewed journal publications indexed in the Scopus database from January 2006 to December 2024, focusing on the application of Big Data Analytics (BDA) techniques in assessing Financial Reporting Quality (FRQ). This approach enables the study to achieve its research objectives and effectively address the research questions.

By giving researchers, professionals, and legislators current information and critical insights on how Big Data Analytics (BDA) is used to investigate Financial Reporting Quality (FRQ) via a comprehensive bibliometric analysis (performance analysis and science mapping), this study advances current knowledge. Assessing the research environment and conceptual framework of this field, it highlights the leading and most highly referred authors and papers and the major contributing journals and countries. It offers research-based evidence to help discover appropriate data analysis (BDA) ways for constructing models that detect and anticipate financial reporting fraud (FRF) and earnings management (EM), as well as track changes in these approaches.

This paper emphasizes the need for data analysis (BDA) applications in FRQ research, therefore enabling the incorporation of sophisticated BDA for academic reasons. It also looks at significant developments and points for future study in this field in specific ways. The study fully described and shared the technique and results using the most recent PRISMA criteria "Preferred Report Items for Systematic Reviews and Meta-analysis," recommended by (Page et al., 2021) facilitating the replication and understanding of the research. The findings of this investigation will aid regulators, investors, analysts of markets, and auditors. This paper proves that big data analysis (BDA) could be a helpful substitute for evaluating financial reporting (FRQ) quality and lowering investment risk.

In recent years, several academics have conducted literature assessments examining the usage of (BDA), and its methodology and procedures to a number of accounting and commercial industries. Big Data Analytics (BDA) has emerged as a critical tool for organizations and industries, providing insights that enhance both operational efficiency and strategic decision-making. Its applications span a wide range of functional areas, including supply chain management, business process optimization, audit decision-making, human resource management, risk management, and the development of dynamic capabilities (Ahmad, 2019; Fauzi et al., 2023; Fosso Wamba & Mishra, 2017; Gupta & Mehta, 2024; Rialti et al., 2019; Santhikumar et al., 2022; Tambuskar et al., 2024) Furthermore, BDA has been implemented across diverse industries such as healthcare, manufacturing, financial services, e-commerce, and tourism, highlighting its versatility and potential to transform organizational practices (Agrawal et al., 2022; Akter & Wamba, 2019; Galetsi & Katsaliaki, 2020; Pattnaik et al., 2024). Research also identifies platforms, technologies, and emerging trends associated with BDA, emphasizing the opportunities and challenges that accompany its adoption (Inamdar et al., 2020; Mohamed et al., 2020)

Beyond operational and strategic benefits, BDA plays an increasingly significant role in detecting and preventing fraudulent activities. Techniques such as artificial intelligence (AI), machine learning (ML), deep learning (DL), data mining (DM), and text mining (TM) have been successfully employed to identify financial fraud, including credit card fraud, thereby strengthening organizational controls and decision-making processes (Btoush et al., 2023; Hilal et al., 2022). Evaluations across various sectors, including during the COVID-19 pandemic, demonstrate the effectiveness of BDA in real-world applications, enabling organizations to maintain resilience and adapt to rapidly changing environments (Al-Sai et al., 2022)

To address gaps in prior research, this study applies a comprehensive bibliometric analysis to explore the relationship between BDA and Financial Reporting Quality (FRQ). The analysis maps the intellectual structure of the field, identifies emerging trends, and provides guidance for future research directions. Unlike previous studies, this research employs broader search strategies to capture all relevant studies linking BDA approaches with FRQ, with a particular focus on earnings management (EM) and financial reporting fraud (FRF). By integrating advanced analytical techniques and systematically reviewing the literature, the study offers a robust foundation for understanding the impact of BDA on financial reporting practices and informs both scholars and practitioners in the field.

2. METHODOLOGY

This research applies bibliometric analysis in combination with Big Data Analytics (BDA) to assess Financial Reporting Quality (FRQ). Bibliometric methods are frequently employed in literature reviews due to their objectivity, quantitative foundation, and ability to minimize potential biases. By systematically analyzing a large collection of scientific publications, this approach enables the identification of research trajectories, extraction of key insights regarding the development of the field, recognition of emerging challenges, understanding of critical issues within a specific domain, and guidance toward the most relevant studies before engaging in detailed examination (Lamboglia et al., 2020). As a result, this research is frequently employed in contemporary management, business, and accounting courses (Donthu et al., 2021).

Donthu et al. (2021) and Lamboglia et al. (2020) A full bibliometric analysis was conducted using two main approaches. Science mapping employs techniques such as co-citation, co-word, thematic evolution, bibliographic coupling, and co-authorship analyses to identify key themes and reveal structural and cognitive connections within the field. Performance analysis, on the other hand, uses bibliographic data such as articles, keywords, and annual citation counts—to assess the productivity, impact, and contributions of authors, journals, countries, and institutions (Aria et al. 2020; Donthu et al. 2021). Such systematic methods have been shown to effectively track the flow of information, utilize evidence-based data transparently, and fully comply with PRISMA 2020 reporting guidelines (Page et al., 2021).

Identifying the Right Keywords (Search Terms)

Defining the research scope constitutes the initial step in selecting suitable keywords. Drawing on previous studies (Shahana et al., 2023; Soltani et al., 2023) the primary search terms chosen were "Big Data Analytics" and "Financial Reporting Quality," along with their related measurement approaches, reflecting the researcher's focus on examining Financial Reporting Quality (FRQ) methods. To maximize coverage of relevant literature, multiple iterations were conducted to construct the most comprehensive search query possible. The final query employed

the TITLE-ABS-KEY function, searching within article titles, abstracts, and keywords in the researcher database.

Data Collection and Preliminary Findings

The data was imported from a regular Scopus database on December 31, 2024 the end date of the search to guarantee the scientific rigor and completeness of peer-reviewed scientific papers used in numerous bibliometric publications. Figure 1 illustrates that the initial search conducted using the above-specified keywords produced 1,088 results in the researcher database.

Initial Results Refinement (Adopting Criteria for inclusion or exclusion)

Using criteria for inclusion or exclusion allows the first search results to be better and thus facilitates the identification of pertinent information by the researcher. Published January 2006 until December 2024, 938 papers with the field of research filter "Business, Management and Accounting," "Economics, Econometrics and Finance," "Social Sciences," "Decision Sciences," "Computer Science," The initial search results were automatically filtered by the database to include only peer-reviewed journal articles published in English, ensuring that the analysis focused on high-quality scholarly sources., ensuring that the analysis focused exclusively on high-quality scholarly publications. The automatic filtering method helped 393 documents to be deleted; the remaining 545 documents were under review for filtering (LIMIT-TO, "ar"). As the PRISMA diagram (Figure 1) clearly shows, 360 were generally kept.

Analysis of Collected Data

At this stage, a bibliometric analysis was conducted on a total of 360 peer-reviewed articles, comprising journal articles authored by researchers and experts in the field. The analysis involved two main dimensions: performance analysis, which assesses the productivity, influence, and citation patterns of authors, institutions, and journals; and scientific mapping, which identifies the intellectual structure, thematic developments, and collaborative networks within the research landscape. To support this investigation, Biblioshiny the graphical interface of the bibliometrix package in R was employed. This tool enabled a systematic, data-driven exploration of the bibliographic information, generating meaningful insights into the evolution, impact, and emerging trends of the studied topic.

3. RESULT AND DISCUSSION

Performance Analysis, Or Descriptive Bibliometric Analysis:

Descriptive Results

Main Information summarizes the studied papers. Published first in this field of research in 2006, 360 journal items total between 2006 and 2024 were printed in 205 publications. There were 891 contributors to these publications; just 43 papers were produced by one author, while the other 50 papers were produced by author collaboration.

The conclusion is presented as a cohesive paragraph, summarizing the study's key findings concisely and clearly. It highlights the contributions of the research, including the development of new theoretical insights and the potential directions for future studies. Additionally, the conclusion discusses the implications of the research, addressing both theoretical significance and practical applications, thereby demonstrating the overall value and impact of the study. The

conclusion should be an answer to the research question, and not expressed in statistical sentences. This underlines the multidisciplinary nature of this field of research since the analysis of Financial Reporting Quality (FRQ) using Big Data Analytics (BDA) requires knowledge.

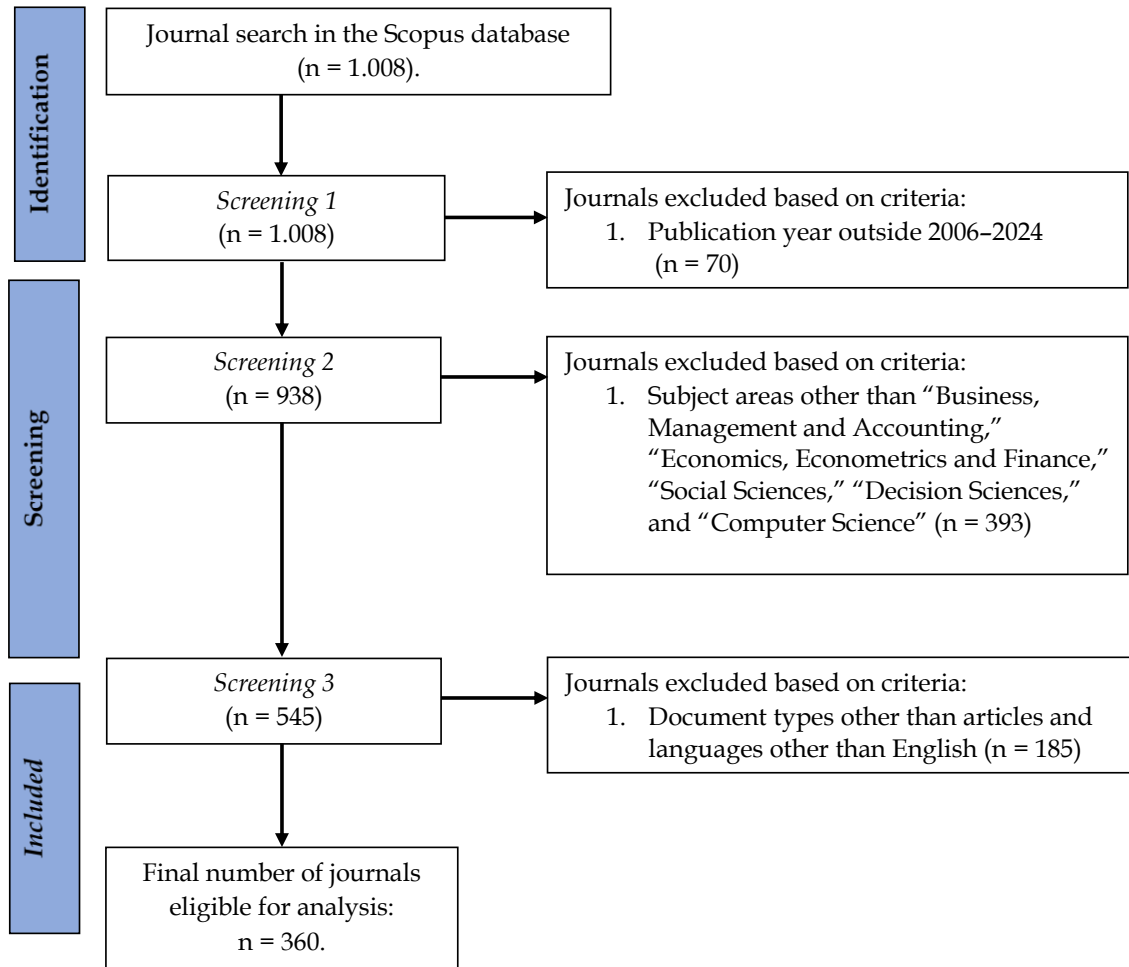


Figure 1. PRISMA Flow Diagram

Number of Publications Annually

Figure 2 displays annual academic article production in the field of research based on the annual published count of papers. First, Kirkos et al., (2007) used Big Data Analytics (BDA) techniques to address fraud in financial reporting. The annual range of publications between 2010 and 2019 shows constant variations between 10 and 20 articles. However, after 2020, scientific publishing output has skyrocketed and is still rising; it peaked in 2024 with roughly 70 papers. This growing tendency indicates more scholarly attention on applying BDA for FRQ. Reflecting fast development in creating and implementing big data analytics (BDA) methodologies to assess financial reporting quality (FRQ), the number of publications reached approximately 147 articles between 2020 and 2023. Table 1 shows the annual growth rate of articles of 26.62% and the strong upward trend line of the graph, therefore confirming the evidence of a notable increase in interest in this field in recent years.

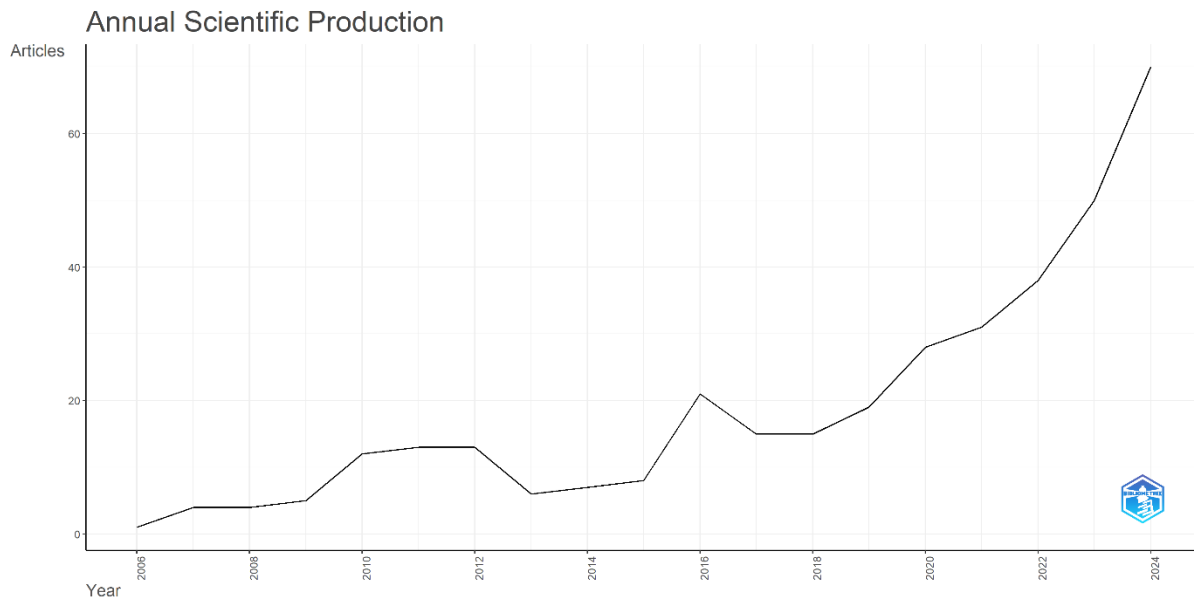


Figure 2.
Annual Research Productivity

Most Influential Journals

The corpus in this study is spread across 86 journals, with 53 articles (42.74%) published in the top 15 journals, each containing at least two articles. Figure 3 shows this research topic's most productive journal sources (RQ1). The Australian Journal of Management is the most dominant journal with 10 articles (8.40%), followed by Sustainability (Switzerland) with 9 articles (7.56%), and the Journal of Accounting and Public Policy with 8 articles (6.72%). Furthermore, Accounting Review and Decision Support Systems each contributed 7 articles (5.88%). This distribution indicates that research in this field is interdisciplinary, reflected in the diversity of journals actively publishing related publications.

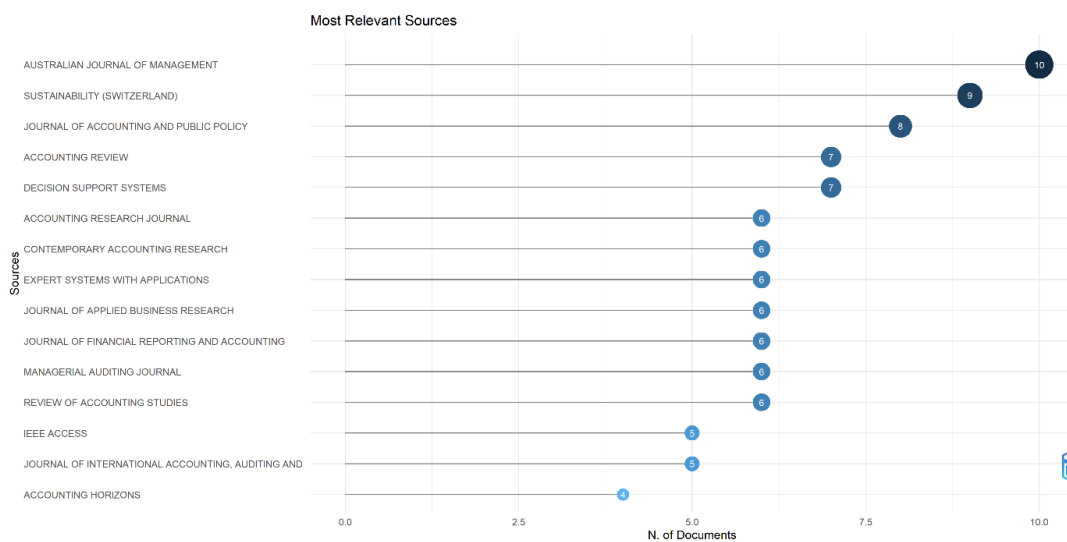


Figure 3.
Most Influential Journals

Most Influential Authors

Table 1 presents the ten most productive authors in the analyzed dataset addressing Research Question 1, each contributing a minimum of three publications either individually or collaboratively. The most prolific author is Bansal, Manish, with a total of nine publications, accounting for 7.56 percent of the dataset, followed by Malikov, Kamran with five publications representing 4.20 percent, while Thomas, Wayne B. and Valaskova, Katarina each contributed four publications, making up 3.36 percent of the total output. The remaining authors, including Bose, Indranil, Chen, Fu-Hsiang, Durana, Pavol, Jan, Chyan-Long, Na, Hyung Jong, and Nagar, Neerav, each contributed three publications, representing 2.52 percent of the dataset. The distribution of these contributions highlights the prominence of highly productive authors while also reflecting the broader engagement of multiple researchers, indicating active scholarly participation and collaboration. These findings demonstrate how both leading and moderately contributing authors play a significant role in advancing research on Big Data Analytics applications in Financial Reporting Quality, shaping the intellectual structure of the field and offering insights into patterns of research activity and collaboration among academics.

Most Productive And Influential Countries

Table 3 presents the ten countries with the highest productivity and citation impact (addressing RQ1 and RQ2), each contributing a minimum of three publications. Collectively, these countries account for approximately 87% of the total documents in the dataset. The United States submitted the most papers (52; 14.4%), followed by China (50; 13.9%). These two countries published 102 papers, accounting for roughly 28.3% of all publications. India ranks third with 29 (8.1%) papers, followed by the United Kingdom in fourth place with 24 (6.7%) papers. South Korea is in fifth place with 14 (3.9%) publications, while Australia and Indonesia each contribute 11 (3.1%). Canada (9), Spain (7), and Greece (6) round out the top ten most productive countries. Regarding citation impact (TC), the United Kingdom ranks first with 2,381 citations, indicating the significant impact of their publications despite having fewer documents than the United States or China. The United States follows with 2,212 citations, followed by China (963) and India (667). Interestingly, Greece ranks fifth in influence with 557 citations despite having only six papers. Other nations having high citation counts are Germany (165), France (159), South Korea (184), Hong Kong (176), and the Czech Republic (225). These results reveal that certain nations with low publishing counts can yet contribute significantly academically. This confirms that, in addition to quantity, the quality and relevance of research also greatly determine a country's level of scientific influence.

Tabel 1.
Top 10 Most Influential Authors

| Authors | Articles | Articles Fractionalized |
|---------------------|----------|-------------------------|
| Bansal, Manish | 9 | 7.58 |
| Malikov, Kamran | 5 | 1.92 |
| Thomas, Wayne B. | 4 | 1.42 |
| Valaskova, Katarina | 4 | 1.08 |
| Bose, Indranil | 3 | 0.83 |
| Chen, Fu-Hsiang | 3 | 1.17 |
| Durana, Pavol | 3 | 0.83 |
| Jan, Chyan-Long | 3 | 2.50 |
| Na, Hyung Jong | 3 | 1.00 |

Nagar, Neerav

3

1.33

Source: Authors contributions using Biblioshiny App (2025)

Table 3.
The Top Ten Highly Production and Cited Countries

| Country | Articles | Articles% | Country | TC |
|----------------|----------|-----------|----------------|------|
| Usa | 52 | 14.4 | United Kingdom | 2381 |
| China | 50 | 13.9 | Usa | 2212 |
| India | 29 | 8.1 | China | 963 |
| United Kingdom | 24 | 6.7 | India | 667 |
| Korea | 14 | 3.9 | Greece | 557 |
| Australia | 11 | 3.1 | Czech Republic | 225 |
| Indonesia | 11 | 3.1 | Korea | 184 |
| Canada | 9 | 2.5 | Hong Kong | 176 |
| Spain | 7 | 1.9 | Germany | 165 |
| Greece | 6 | 1.7 | France | 159 |

Source: Authors contributions using Biblioshiny App (2025)

Most Frequent Authors' Keywords

The researcher corpus contains 171 occurrences of keywords from various research documents. Table 4 shows the 10 most frequently occurring keywords (RQ3), each of which appears at least 10 times. Figure 4 depicts the collection of words from all these keywords. The more frequently a keyword appears, increase the font size between terms in the word cloud. The frequency of keywords shows the number of articles that use that word as the primary study topic. The total number of occurrences of the 10 most popular keywords is 171 times (100%). The keyword most commonly used by authors across the analyzed publications is "finance.", with 30 occurrences (17.5%), followed by "crime" with 27 occurrences (15.8%), and 'data mining' with 23 occurrences (13.5%). Additionally, keywords such as 'financial statement frauds' and 'decision trees' appear 18 times (10.5%) and 16 times (9.4%), respectively. These findings indicate that most studies use data mining, decision trees, and machine learning approaches, especially to uncover financial practices, financial crimes, and fraud in financial statements. In addition, keywords such as 'fraud detection,' Additionally extensively employed are NN and SVM in exploring and developing analytical methods for detecting fraud as an indicator of financial reporting quality (FRQ).

Table 4.
Top 10 Most Frequent Authors' Keywords

| Words | Occurrences | % Of Occurrences |
|----------------------------|-------------|------------------|
| Finance | 30 | 17.5% |
| Crime | 27 | 15.8% |
| Data Mining | 23 | 13.5% |
| Financial Statement Frauds | 18 | 10.5% |
| Decision Trees | 16 | 9.4% |
| Fraud Detection | 13 | 7.6% |
| Neural Networks | 12 | 7.0% |
| Financial Statements | 11 | 6.4% |
| Machine Learning | 11 | 6.4% |

Table 2.
Top 10 Publications by Global Citation Count

| Authors | Total Citations | TC per Year |
|---------------------------|-----------------|-------------|
| Sivaraja et al., (2017) | 1447 | 160.78 |
| Li, F. (2010) | 792 | 49.50 |
| Hennes et al., (2008) | 570 | 31.67 |
| Kirkos et al., (2007) | 473 | 24.89 |
| Feldman et al., (2010) | 432 | 27.00 |
| Ravisankar et al., (2011) | 383 | 25.53 |
| Larcker et al., (2012) | 372 | 26.57 |
| Humpherys et al., (2011) | 253 | 16.87 |
| Hajek et al., (2017) | 224 | 24.89 |
| Perols et al., (2011) | 212 | 14.13 |

Source: Authors contributions using Biblioshiny App (2025)

Science Map: Bibliometric Network Analysis

Science mapping creates a network or bibliometric map that outlines the conceptual, intellectual, and social structure of a certain discipline (Aria & Cuccurullo, 2017; Zupic & Čater, 2015) so exploring the relationship between document attributes that is, source, author, keywords, references, and citations). This section addresses Research Questions 4, 5, and 6, providing insights and analysis relevant to each query.

Co-Citation Analysis

Co-citation analysis can reveal a study domain's knowledge base, intellectual structure, and theme clusters (Donthu et al., 2021; Zupic & Čater, 2015). The co-citation of authors assumes that the more frequently two writers are mentioned in other publications, the more probable their scientific contributions will have thematic similarities (Fahimnia et al., 2015). Because the co-citation network in this study had many references, the analysis, i.e., authors with acceptable link strength. The image depicts the author's co-citation analysis results, which were rendered using Biblioshiny.

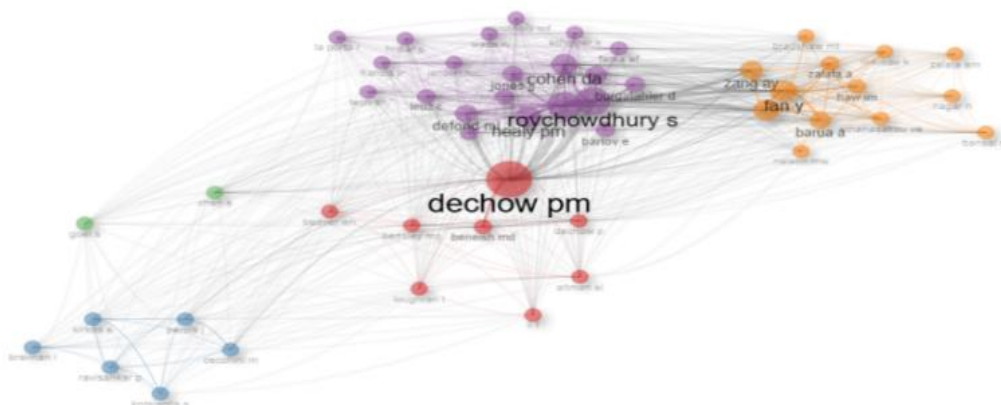


Figure 5.
Co Citation Network

According to the co-citation network visualization, the intellectual organization of this field is divided into five main groups (RQ4). The red group focuses led by Dechow PM. The purple group highlights theories and models in financial statement manipulation, with figures such as Roychowdhury S and Cohen DA. The orange group investigates the role of governance and managerial characteristics in fraudulent reporting. The blue group is considering the use of machine learning to detect fraud, while the green group uses traditional data mining approaches for the same purpose. Each group represents a different thematic focus in fraudulent financial reporting research.

Co-Word Analysis

Many points of view are provided by co-word analysis, which also clarifies the significant problems and ideas of a given topic (Lamboglia et al., 2020). Donthu et al., (2021) build a conceptual framework of a particular field using a network of co-occurring keywords co-word analysis a technique used to study the actual content of papers. This is based on the concept usually found in articles; keywords imply closely linked ideas (Aria et al., 2020). His method helps one to identify these groups that is, keywords that mirror a focus point of interest or a specific theme/topic of inquiry in the studied corpus (Aria et al., 2020). A fractional counting approach helped the researcher identify thirty-three of the 712 keywords in her co-word network (Figure 6).



Figure 6.
Co Occurrence Network

Figure 6 shows the co-occurrence network of six main clusters, each marked with a different color to represent a distinct theme. The blue cluster, the largest, containing terms like “finance,” “data mining,” “crime,” “fraud detection,” and “financial statement frauds.” The orange cluster highlights recent advances in sophisticated machine learning for fraud detection, including “graph neural networks,” “adversarial machine learning,” and “contrastive learning.” The green cluster reflects a regional focus, combining “machine learning,” “artificial intelligence,” and “China.” The red cluster links algorithmic approaches with sustainability and financial systems, featuring terms such as “artificial neural network,” “algorithm,” and “sustainable development.”

The purple cluster addresses methodological and market-related aspects of fraud detection, using terms like “capital market” and “detection method.” Lastly, the brown cluster represents statistical model applications in detecting financial reporting manipulation, with terms like “earnings management,” “regression analysis,” and “feature selection.” This study presents six clusters demonstrating the applicability of several Big Data Analytics (BDA) approaches to evaluate financial reporting quality (FRQ), highlighting a significant need in this field of research.

Theme Evolution Over Time

Theme Evolution Over Time is the relationships between many themes, demonstrating how these themes converge or diverge into several topics. Thus, it provides an interesting overview of the field's historical evolution and a glimpse of potential future research directions (Aria et al., 2020). Based on the concurrent occurrence of author keywords, researchers created a theme evolution map from 2006 to 2024 using Biblioshiny. Three-time slices 2006–2018, 2019–2022, and 2023–2024 were used in the analysis to investigate developments in BDA methods for financial statement fraud detection. Visualizing the thematic progression of BDA applications for financial statement fraud detection, Figure 7 The rectangles' dimensions reflect the frequency of keywords.

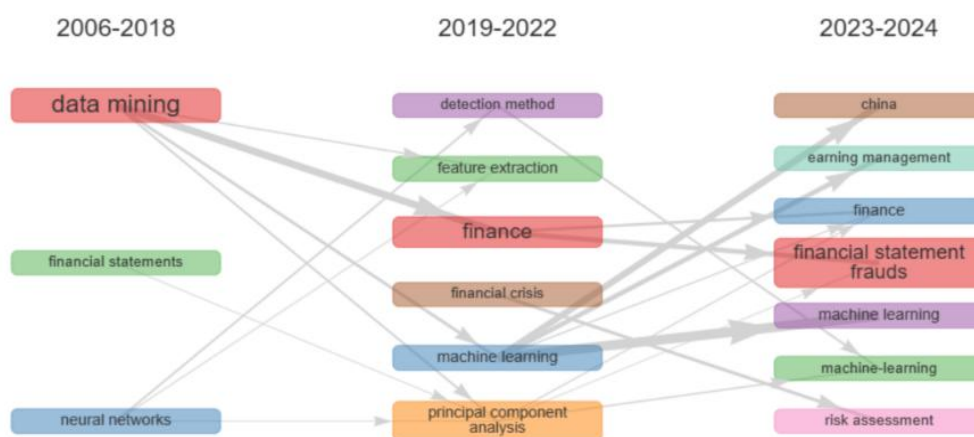


Figure 7.
Thematic Evolution

During the first period (2006–2018), academics focused on data mining techniques, with neural networks and financial statements as supporting keywords. Research during this period explored how data mining techniques were used to assess financial statements and develop early detection systems for financial fraud.

From 2019 to 2022, there was a significant shift in the research focus. Linking important terminology such feature extraction, detection techniques, principal component analysis, financial crises, and machine learning, the issue of finance became the core emphasis. This shows a change from conventional data-based methods to increasingly sophisticated, contemporary algorithm-based ones. Applying techniques, including machine learning and advanced statistical analysis, marks an evolution in the approach to detecting anomalies and financial statement manipulation (RQ5).

After 2022, specifically in the 2023–2024 period, research explicitly focused on the theme of financial statement frauds, which became the central core and was connected to various keywords such as machine learning, machine-learning (with spelling variations), earnings management, finance, risk assessment, and country-specific focuses like China. This indicates that the literature is beginning to narrow its focus to more specific and applied aspects, including in a geographical context. Detection methods are being refined by applying risk-based assessment techniques and focusing on complex earnings management practices.

These results show that the availability of massive volumes and kinds of data and continuous technological developments help explain the continual increase in the acceptance of BDA approaches. Among the several methods of neural networks that benefit one another are deep neural networks (DNN), CNN, and RNN, both types of neural networks. Amazingly creative and powerful prediction and detection systems were developed to fight dishonest financial reporting practices during and after crises. Interestingly, clever and efficient prediction and detection systems were developed to offset dishonest financial reporting methods during and following crises. These models apply ensemble learning methods among other unsupervised and supervised learning methods. Other research employing modern BDA techniques like feature selection, machine learning, and data mining algorithms has also found and anticipated financial statement fraud and earnings manipulation.

Trend Topics

Depending on the results in Figure 8, these sentences relate to applying (BDA) in investigating FRF in numerous techniques. Emerging among the first and most often used approaches in recognizing and forecasting financial reporting fraud (FRF) in 2012 were neural networks (NN). Until 2020, neural networks (NN) remained a significant focus. Researchers include Kirkos et al., (2007) and Liou (2008) , who developed financial fraud-detecting models using NN and other data mining (DM) techniques. Inspired by Ragothaman & Lavin (2008), who used NN to forecast financial information manipulation and revenue restatement, Ragothaman & Lavin (2008) utilized NN to predict financial information manipulation and revenue restatement, followed by Ravisankar et al., (2011) and Zhou, & Kapoor (2011) created similar models for spotting financial reporting fraud (FRF). In 2014, financial reporting, support vector machines (SVM), and financial statements were among the subjects trending. Applied generally in classification and regression to forecast and identify FRF, the SVM algorithm is a supervised machine learning (ML) method (Jan, 2018; Omar et al., 2017). Combining SVM with other DM techniques, the researchers Dutta et al. (2017), Hsieh et al. (2012), and Nan et al. (2012) discovered earnings management (EM) and financial statement restatements.

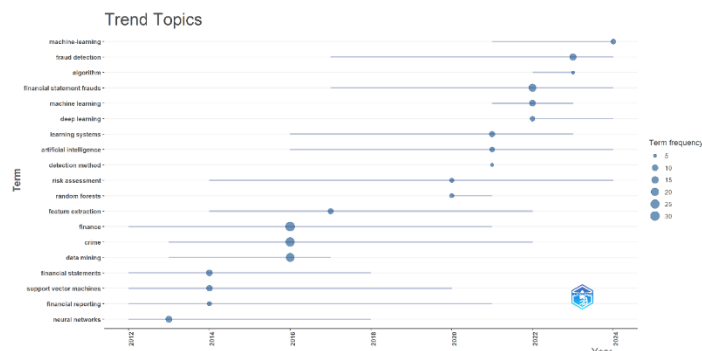


Figure 8.
Trend Topics

From 2016 through 2019, data mining, crime, and finance attracted growing interest. From personal approaches to a more all-encompassing, contextual approach covering financial and criminal issues, this age saw developments in fraud detection measures. Moreover, improved statistical methods and feature extraction-based approaches began to be applied, therefore showing development in the processing and analysis of financial data. 2020 brought innovative ideas such as random forests, risk assessment, and detection techniques that deviate from conventional knowledge to use modern algorithms and more exact risk evaluation processes. This also shows BDA's flexibility in risk management, fraud detection, and evaluation of the possibility of systemic collapse.

Deep learning (DL) and machine learning approaches dominated learning systems during 2021 and 2022. This field moves toward developing intelligent learning algorithm-based artificial intelligence fraud prediction systems. Inquiry models for financial statement fraud are developed using ML and DL algorithms by Ali et al. (2023), Li et al. (2023) and Rahman et al. (2023), Topics include algorithms, financial statement frauds acquired the stage in 2023 and 2024 and showed greatest frequency RQ6, machine learning, fraud detection, and tools. Furthermore, phrases like artificial intelligence (AI) and deep learning suggest that advancements are oriented toward maximizing AI-based BDA approaches for audit, reporting, and financial control needs. Several recent research by Alam et al. (2021), Li et al. (2023), Xiuguo & Shengyong (2022) constructing prediction models using the MD&A component of annual reports and testing AEM using several DL methods) confirm this tendency.

Not least of all, artificial intelligence (AI) is most likely to remain the most current developing topic, breaking through in 2023 and the future. Choi et al. (2022) use artificial intelligence to enhance the accuracy of anomalous audit cost forecasts, and Rahman & Zhu (2023) in developing a fraud categorizing system founded on artificial intelligence. These findings show that ML, DL, and artificial intelligence algorithms increasingly predominate in research trends. This implies that developing and applying these techniques give great possibility for future financial statement fraud detection investigations.

4. CONCLUSION

This study provides meaningful theoretical and practical insights into the evolution of literature on the application of Big Data Analytics (BDA) in examining Financial Reporting Quality (FRQ). Through a bibliometric analysis using Biblioshiny, 360 articles published between 2006 and 2024 were identified from the Scopus database. The results reveal rapid growth in this field, which has developed through a multidisciplinary approach and shows a high annual growth rate. The Australian Journal of Management emerges as the most relevant publication outlet, Bansal Manish is identified as the most productive author, and the United Kingdom holds the position of the most influential country. Co-citation analysis categorizes the literature into five major themes: earnings management detection as an indicator of reporting quality, theoretical and model-based studies on financial statement manipulation, the influence of governance and managerial traits on fraudulent reporting, the use of machine learning for detecting financial statement fraud, and data mining approaches. Dominant keywords include machine learning (ML), earnings management (EM), and data mining (DM), with recent developments showing increased adoption of advanced algorithms such as deep learning (DL) and artificial intelligence (AI), particularly in addressing misleading reporting in the post-financial crisis period.

The findings reinforce that BDA serves as an effective alternative for detecting financial statement manipulation, where algorithms like random forest, decision tree, and natural language processing can support regulators, auditors, and investors in improving fraud detection. Suggested directions for future research include extending investigations to developing countries, integrating BDA with Benford's Law, applying these techniques to the financial sector, and analyzing narrative disclosures such as sustainability reports. Nonetheless, this study faces certain limitations, including a focus on English-language journals and reliance on a single database. Future studies could benefit from combining bibliometric and narrative systematic literature review methods to examine various algorithm types, FRQ measurements, and theoretical perspectives, thereby providing a deeper understanding of the effectiveness and overall dynamics of BDA in the FRQ context.

5. REFERENCES

- Abdelhalim, A. M. (2024). How management accounting practices integrate with big data analytics and its impact on corporate sustainability. *Journal of Financial Reporting and Accounting*, 22(2), 416–432. <https://doi.org/10.1108/JFRA-01-2023-0053>
- Aboelfotoh, A., Zamel, A. M., Abu-Musa, A. A., Frendy, F., Sabry, S. H., & Moubarak, H. (2024). Examining the ability of big data analytics to investigate financial reporting quality: a comprehensive bibliometric analysis. *Journal of Financial Reporting and Accounting*. <https://doi.org/10.1108/JFRA-11-2023-0689>
- Aftabi, S. Z., Ahmadi, A., & Farzi, S. (2023). Fraud detection in financial statements using data mining and GAN models. *Expert Systems with Applications*, 227, 120144. <https://doi.org/10.1016/j.eswa.2023.120144>
- Agrawal, R., Wankhede, V. A., Kumar, A., Luthra, S., & Huisingh, D. (2022). Big data analytics and sustainable tourism: A comprehensive review and network based analysis for potential future research. *International Journal of Information Management Data Insights*, 2(2), 100122. <https://doi.org/10.1016/j.jjime.2022.100122>
- Ahmad, F. (2019). A systematic review of the role of Big Data Analytics in reducing the influence of cognitive errors on the audit judgement. *Revista de Contabilidad-Spanish Accounting Review*, 22(2), 187–202. <https://doi.org/10.6018/rccsar.382251>
- Akter, S., & Wamba, S. F. (2019). Big data and disaster management: a systematic review and agenda for future research. *Annals of Operations Research*, 283(1–2), 939–959. <https://doi.org/10.1007/s10479-017-2584-2>
- Al-Sai, Z. A., Husin, M. H., Syed-Mohamad, S. M., Abdin, R. M. S., Damer, N., Abualigah, L., & Gandomi, A. H. (2022). Explore Big Data Analytics Applications and Opportunities: A Review. *Big Data and Cognitive Computing*, 6(4), 157. <https://doi.org/10.3390/bdcc6040157>
- Alam, N., Gao, J., & Jones, S. (2021). Corporate failure prediction: An evaluation of deep learning vs discrete hazard models. *Journal of International Financial Markets, Institutions and Money*, 75, 101455. <https://doi.org/10.1016/j.intfin.2021.101455>

- Ali, A. Al, Khedr, A. M., El-Bannany, M., & Kanakkayil, S. (2023). A Powerful Predicting Model for Financial Statement Fraud Based on Optimized XGBoost Ensemble Learning Technique. *Applied Sciences (Switzerland)*, 13(4), 2272. <https://doi.org/10.3390/app13042272>
- Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4), 959–975. <https://doi.org/10.1016/j.joi.2017.08.007>
- Aria, M., Misuraca, M., & Spano, M. (2020). Mapping the Evolution of Social Research and Data Science on 30 Years of Social Indicators Research. *Social Indicators Research*, 149(3), 803–831. <https://doi.org/10.1007/s11205-020-02281-3>
- Azzam, M. E. A. Y., Alsayed, M. S. H., Alsultan, A., & Hassanein, A. (2024). How big data features drive financial accounting and firm sustainability in the energy industry. *Journal of Financial Reporting and Accounting*, 22(1), 29–51. <https://doi.org/10.1108/JFRA-03-2023-0125>
- Blazquez, D., & Domenech, J. (2018). Big Data sources and methods for social and economic analyses. *Technological Forecasting and Social Change*, 130, 99–113. <https://doi.org/10.1016/j.techfore.2017.07.027>
- Btoush, E. A. L. M., Zhou, X., Gururajan, R., Chan, K. C., Genrich, R., & Sankaran, P. (2023). A systematic review of literature on credit card cyber fraud detection using machine and deep learning. *PeerJ Computer Science*, 9. <https://doi.org/10.7717/PEERJ-CS.1278>
- Choi, S. U., Lee, K. C., & Na, H. J. (2022). Exploring the deep neural network model's potential to estimate abnormal audit fees. *Management Decision*, 60(12), 3304–3323. <https://doi.org/10.1108/MD-07-2021-0954>
- Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, 133, 285–296. <https://doi.org/10.1016/j.jbusres.2021.04.070>
- Dutta, I., Dutta, S., & Raahemi, B. (2017). Detecting financial restatements using data mining techniques. *Expert Systems with Applications*, 90, 374–393. <https://doi.org/10.1016/j.eswa.2017.08.030>
- Fahimnia, B., Sarkis, J., & Davarzani, H. (2015). Green supply chain management: A review and bibliometric analysis. *International Journal of Production Economics*, 162, 101–114. <https://doi.org/10.1016/j.ijpe.2015.01.003>
- Fauzi, M. A., Kamaruzzaman, Z. A., & Abdul Rahman, H. (2023). Bibliometric review on human resources management and big data analytics. *International Journal of Manpower*, 44(7), 1307–1327. <https://doi.org/10.1108/IJM-05-2022-0247>
- Fosso Wamba, S., & Mishra, D. (2017). Big data integration with business processes: a literature review. *Business Process Management Journal*, 23(3), 477–492. <https://doi.org/10.1108/BPMJ-02-2017-0047>
- Framework, C. (2018). Conceptual framework for financial reporting. *IFRS Foundation*.
- Galetsis, P., & Katsaliaki, K. (2020). A review of the literature on big data analytics in healthcare.

Journal of the Operational Research Society, 71(10), 1511–1529.
<https://doi.org/10.1080/01605682.2019.1630328>

- Gupta, S., & Mehta, S. K. (2024). Data Mining-based Financial Statement Fraud Detection: Systematic Literature Review and Meta-analysis to Estimate Data Sample Mapping of Fraudulent Companies Against Non-fraudulent Companies. *Global Business Review*, 25(5), 1290–1313. <https://doi.org/10.1177/0972150920984857>
- Hajek, P., & Henriques, R. (2017). Mining corporate annual reports for intelligent detection of financial statement fraud – A comparative study of machine learning methods. *Knowledge-Based Systems*, 128, 139–152. <https://doi.org/10.1016/j.knosys.2017.05.001>
- Hennes, K. M., Leone, A. J., & Miller, B. P. (2008). The importance of distinguishing errors from irregularities in restatement research: The case of restatements and CEO/CFO turnover. *Accounting Review*, 83(6), 1487–1519. <https://doi.org/10.2308/accr.2008.83.6.1487>
- Hilal, W., Gadsden, S. A., & Yawney, J. (2022). Financial Fraud: A Review of Anomaly Detection Techniques and Recent Advances. *Expert Systems with Applications*, 193, 116429. <https://doi.org/10.1016/j.eswa.2021.116429>
- Hsieh, T.-J., Hsiao, H.-F., & Yeh, W.-C. (2012). Mining financial distress trend data using penalty guided support vector machines based on hybrid of particle swarm optimization and artificial bee colony algorithm. *Neurocomputing*, 82, 196–206. <https://doi.org/https://doi.org/10.1016/j.neucom.2011.11.020>
- Inamdar, Z., Raut, R., Narwane, V. S., Gardas, B., Narkhede, B., & Sagnak, M. (2020). A systematic literature review with bibliometric analysis of big data analytics adoption from period 2014 to 2018. *Journal of Enterprise Information Management*, 34(1), 101–139. <https://doi.org/10.1108/JEIM-09-2019-0267>
- Jan, C. L. (2018). An effective financial statements fraud detection model for the sustainable development of financial markets: Evidence from Taiwan. *Sustainability (Switzerland)*, 10(2). <https://doi.org/10.3390/su10020513>
- Kirkos, E., Spathis, C., & Manolopoulos, Y. (2007). Data Mining techniques for the detection of fraudulent financial statements. *Expert Systems with Applications*, 32(4), 995–1003. <https://doi.org/10.1016/j.eswa.2006.02.016>
- Lamboglia, R., Lavorato, D., Scornavacca, E., & Za, S. (2020). Exploring the relationship between audit and technology. A bibliometric analysis. *Meditari Accountancy Research*, 29(5), 1233–1260. <https://doi.org/10.1108/MEDAR-03-2020-0836>
- Li, F. (2010). The information content of forward- looking statements in corporate filings-A naïve bayesian machine learning approach. *Journal of Accounting Research*, 48(5), 1049–1102. <https://doi.org/10.1111/j.1475-679X.2010.00382.x>
- Li, R., Liu, B., Yuan, X., & Chen, Z. (2023). A bibliometric analysis of research on R-loop: Landscapes, highlights and trending topics. *DNA Repair*, 127. <https://doi.org/10.1016/j.dnarep.2023.103502>

- Liou, F. M. (2008). Fraudulent financial reporting detection and business failure prediction models: A comparison. *Managerial Auditing Journal*, 23(7), 650–662. <https://doi.org/10.1108/02686900810890625>
- Mandal, A., & Amilan, S. (2023). Preventing financial statement fraud in the corporate sector: insights from auditors. *Journal of Financial Reporting and Accounting*. <https://doi.org/10.1108/JFRA-02-2023-0101>
- Mohamed, A., Najafabadi, M. K., Wah, Y. B., Zaman, E. A. K., & Maskat, R. (2020). The state of the art and taxonomy of big data analytics: view from new big data framework. *Artificial Intelligence Review*, 53(2), 989–1037. <https://doi.org/10.1007/s10462-019-09685-9>
- Nan, X., Sun, X., Li, Y., & Hou, T. (2012). Weighted-Support Vector Machine based earnings management detection during IPOs. *Journal of Information and Computational Science*, 9(9), 2607–2617.
- Omar, N., Johari, Z. A., & Smith, M. (2017). Predicting fraudulent financial reporting using artificial neural network. *Journal of Financial Crime*, 24(2), 362–387. <https://doi.org/10.1108/JFC-11-2015-0061>
- Page, M. J., McKenzie, J. E., Bossuyt, P., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The prisma 2020 statement: An updated guideline for reporting systematic reviews. *Medicina Fluminensis*, 57(4), 444–465. https://doi.org/10.21860/medflum2021_264903
- Pattnaik, D., Ray, S., & Raman, R. (2024). Applications of artificial intelligence and machine learning in the financial services industry: A bibliometric review. *Heliyon*, 10(1). <https://doi.org/10.1016/j.heliyon.2023.e23492>
- Perols, J. L., Bowen, R. M., Zimmermann, C., & Samba, B. (2017). Finding needles in a haystack: Using data analytics to improve fraud prediction. *Accounting Review*, 92(2), 221–245. <https://doi.org/10.2308/accr-51562>
- Ragothaman, S., & Lavin, A. (2008). Restatements Due to Improper Revenue Recognition: A Neural Networks Perspective. *Journal of Emerging Technologies in Accounting*, 5(1), 129–142. <https://doi.org/10.2308/jeta.2008.5.1.129>
- Rahman, M. J., & Zhu, H. (2023). Predicting accounting fraud using imbalanced ensemble learning classifiers – evidence from China. *Accounting and Finance*, 63(3), 3455–3486. <https://doi.org/10.1111/acfi.13044>
- Ravisankar, P., Ravi, V., Raghava Rao, G., & Bose, I. (2011). Detection of financial statement fraud and feature selection using data mining techniques. *Decision Support Systems*, 50(2), 491–500. <https://doi.org/10.1016/j.dss.2010.11.006>
- Rialti, R., Marzi, G., Ciappei, C., & Busso, D. (2019). Big data and dynamic capabilities: a bibliometric analysis and systematic literature review. *Management Decision*, 57(8), 2052–2068. <https://doi.org/10.1108/MD-07-2018-0821>

- Rusydiana, A. S., Laila, N., Tubastuvi, N., Ibrahim, M. A., & Marlina, L. (2021). Energy efficiency in oic countries: Sdg 7 output. *International Journal of Energy Economics and Policy*, 11(1), 74–81. <https://doi.org/10.32479/ijeep.10499>
- Santhikumar, R., Kartillkayani, K., Mishra, M. K., Thota, S., Beschi, I. S., & Mishra, B. P. (2022). Utilization of Big Data Analytics for Risk Management. *4th International Conference on Inventive Research in Computing Applications, ICIRCA 2022 - Proceedings*, 1559–1565. <https://doi.org/10.1109/ICIRCA54612.2022.9985709>
- Shahana, T., Lavanya, V., & Bhat, A. R. (2023). State of the art in financial statement fraud detection: A systematic review. *Technological Forecasting and Social Change*, 192. <https://doi.org/10.1016/j.techfore.2023.122527>
- Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*, 70, 263–286. <https://doi.org/10.1016/j.jbusres.2016.08.001>
- Soltani, M., Kythreotis, A., & Roshanpoor, A. (2023). Two decades of financial statement fraud detection literature review; combination of bibliometric analysis and topic modeling approach. *Journal of Financial Crime*, 30(5), 1367–1388. <https://doi.org/10.1108/JFC-09-2022-0227>
- Tambuskar, D. P., Jain, P., & Narwane, V. S. (2024). An exploration into the factors influencing the implementation of big data analytics in sustainable supply chain management. *Kybernetes*, 53(5), 1710–1739. <https://doi.org/10.1108/K-07-2022-1057>
- Tang, J., & Karim, K. E. (2019). Financial fraud detection and big data analytics – implications on auditors' use of fraud brainstorming session. *Managerial Auditing Journal*, 34(3), 324–337. <https://doi.org/10.1108/MAJ-01-2018-1767>
- Teixeira, J. F., & Rodrigues, L. L. (2022). Earnings management: a bibliometric analysis. *International Journal of Accounting and Information Management*, 30(5), 664–683. <https://doi.org/10.1108/IJAIM-12-2021-0259>
- Warren, J. D., Moffitt, K. C., & Byrnes, P. (2015). How big data will change accounting. *Accounting Horizons*, 29(2), 397–407. <https://doi.org/10.2308/acch-51069>
- Xiuguo, W., & Shengyong, D. (2022). An Analysis on Financial Statement Fraud Detection for Chinese Listed Companies Using Deep Learning. *IEEE Access*, 10, 22516–22532. <https://doi.org/10.1109/ACCESS.2022.3153478>
- Zhou, W., & Kapoor, G. (2011). Detecting evolutionary financial statement fraud. *Decision Support Systems*, 50(3), 570–575. <https://doi.org/10.1016/j.dss.2010.08.007>
- Zupic, I., & Čater, T. (2015). Bibliometric Methods in Management and Organization. *Organizational Research Methods*, 18(3), 429–472. <https://doi.org/10.1177/1094428114562629>