The current issue and full text archive of this journal is available at
https://e-journal.unair.ac.id/jraba
Jurnal Riset Akuntansi dan Bisnis Airlangga Vol. 8. No. 2 (2023) 1521-1542 ISSN 2548-1401 (Print) ISSN 2548-4346 (Online)

FINANCIAL FRAUD DETECTION AND MACHINE LEARNING ALGORITHM (UNSUPERVISED LEARNING): SYSTEMATIC LITERATURE REVIEW

Nadia Husnaningtyas[1]
Totok Dewayanto[2]

ABSTRACT

This research aims to assess the usage of unsupervised learning in detecting financial fraud across various financial industries by identifying cognitive constructs, benefits, economic optimization, and challenges associated with fraud detection necessitating innovative approaches for effective detection. This study conducts Systematic Literature Review following PRISMA protocol for article selection of 27 journal articles published between 2010 and 2023, sourced from Scopus database. The analysis discloses that unsupervised learning has been implemented across diverse financial sectors, including online payments, insurance, and prominently in banking, especially for identifying anomalies in credit card transactions. K-Means is the most popular method used in unsupervised learning. Nevertheless, there are ongoing challenges that require solutions to ensure the efficacy of machine learning implementation, encompassing issues like class imbalance and the complexity of fraudulent activities. In theoretical terms, this research provides an understanding of cognitive concepts, benefits and applications, challenges, and practical recommendations in the use of unsupervised learning for financial fraud detection. This is useful for practical implementation, benefiting industry practitioners in selecting appropriate models with datasets that have the potential to enhance detection system accuracy and reduce financial losses due to fraud.

Keyword: **Fraud, Detection, Machine Learning, Unsupervised Learning**

ABSTRAK

Penelitian ini bertujuan untuk menilai penggunaan *unsupervised learning* dalam mendeteksi kecurangan keuangan di berbagai sektor keuangan dengan mengidentifikasi konsep kognitif, manfaat, optimalisasi ekonomi, dan tantangan yang terkait dengan deteksi kecurangan yang memerlukan pendekatan inovatif untuk pendeteksian yang efektif. Studi ini menggunakan *Systematic Literature Review* dengan mengikuti protokol PRISMA dalam pemilihan artikel dari 27 jurnal yang diterbitkan antara tahun 2010 dan 2023 yang diperoleh dari database Scopus. Analisis ini mengungkapkan bahwa *unsupervised learning* telah diterapkan di berbagai sektor keuangan, termasuk pembayaran online, asuransi, dan juga perbankan, khususnya dalam mengidentifikasi anomali transaksi kartu kredit. *K-Means* adalah metode yang paling populer yang digunakan dalam *unsupervised learning*. Walaupun demikian, masih terdapat tantangan yang memerlukan solusi untuk memastikan efektivitas implementasi *machine learning*, diantaranya masalah ketidakseimbangan kelas dan kompleksitas aktivitas kecurangan. Secara teoretis, penelitian ini memberikan pemahaman konsep kognitif, manfaat dan penerapan, tantangan, dan rekomendasi praktis dalam penggunaan *unsupervised learning* untuk deteksi kecurangan keuangan. Hasil ini juga berguna untuk diterapkan dalam praktiknya yang bermanfaat bagi praktisi industri untuk memilih penggunaan model dengan dataset yang sesuai, yang berpotensi meningkatkan akurasi sistem deteksi dan mengurangi kerugian keuangan akibat kecurangan.

Kata kunci: **Kecurangan, Deteksi, Machine Learning, Unsupervised Learning**

ARTICLE INFO

***Article History*:**

Received 04 June 2023

Accepted 06 October 2023

Available online 30 November 2023

Introduction

Financial fraud is a pervasive global challenge that demands our collective attention in an increasingly digitized world. The prevalence of fraud indicates that a method is needed to detect or prevent fraud from occurring (Widhiastuti and Kumalasari, 2020). As reported in the 2022 ACFE study, incidents of asset misappropriation account for a staggering 86% of reported fraud cases. This involves both theft and the improper utilization of internal corporate authority. Efforts to

[1] Correspondence Author : Master Graduate Student at Universitas Diponegoro, Semarang,Telp. 085641504949, Email : nadiahusnaningtyas@gmail.com
[2] Second Author : Lecturer at Universitas Diponegoro, Semarang, Email : totokdewayanto@lecturer.undip.ac.id

Jurnal Riset Akuntansi dan Bisnis Airlangga Vol. 8. No. 2 (2023) 1521-1542
ISSN 2548-1401 (Print) ISSN 2548-4346 (Online)

address this particular type of fraud encompass various actions such as scrutinizing credit card expenditures, conducting unannounced audits on authorized employees, analyzing trends in vendor payments, investigating unfavorable balances within performance reports, and other relevant measures (Singleton and Singleton, 2010). Conflicts of interest related to weaknesses in the organizational system, conflicts of interest in the procurement process, and conflicts of interest in the process of granting power and authority tend to be associated with fraudulent actions (Koerniawati, 2021). To effectively counter these fraudulent activities, the ongoing development of more advanced and efficient strategies remains crucial. In this context, a shift in operations from paper-based to entirely paperless digital will be required by digital transformation, with the expectation that fraud such as financial statement manipulation will be minimized (Normasari and Sekar Mayangsari, 2022). The complexity and sophistication of these fraudulent activities demand innovative approaches to detection and prevention. This study recognizes the urgent need to address this pressing issue.

The advancements in information technology, besides being utilized for work efficiency, have enabled fraudulent actors to manipulate electronic data (Tiwari et al., 2020). Consequently, the adoption of technological innovations as tools for financial fraud detection has become imperative. Technology not only holds the promise of enhancing the effectiveness of fraud detection but also presents opportunities to reduce detection costs and expedite the analysis process (Pugliese et al., 2021). These technologies can identify patterns that are difficult for humans to detect, thus helping to minimize the risk of fraud (Raghavan and Gayar, 2019). Among these technological marvels, machine learning stands out, equipped to identify anomalies and enable computers to learn autonomously. Machine learning enables computers to learn without explicit programming (Mitchell, 1997). Machine learning has found extensive applications across diverse domains, spanning document classification, computer vision, natural language processing, and, notably, fraud detection (Constâncio et al., 2023).

Machine learning technology, particularly, has exhibited remarkable accuracy in detecting financial fraud compared to conventional human-centric methods (Saeidi S.P., 2020). This underscores the urgency for its widespread adoption by both companies and regulatory bodies to embrace machine learning as a fraud detection tool for enhanced effectiveness and efficiency. The algorithms are able to be applied to examine transactional data and recognize suspicious fraud patterns, such as unusual transactions or recurring transaction patterns (Zhang et al., 2021). Its application is used to detect credit card fraud activities, bank credit administration, and online banking transactions (Carneiro et al., 2015; Gyamfi and Abdulai, 2019; Kumari et al., 2014).

Machine learning operates under various classifications, including supervised learning, unsupervised learning, and reinforcement learning (Sánchez-Aguayo et al., 2021). Among these, supervised learning and reinforcement learning are the most commonly employed techniques in fraud detection compared to unsupervised learning (Ali et al., 2022; Ashtiani and Raahemi, 2022; Constâncio et al., 2023; Sánchez-Aguayo et al., 2021). Unsupervised learning, however, stands out as it leverages unlabeled data to uncover patterns and anomalies (Ali et al., 2022). This

Jurnal Riset Akuntansi dan Bisnis Airlangga Vol. 8. No. 2 (2023) 1521-1542
ISSN 2548-1401 (Print) ISSN 2548-4346 (Online)

involves programming the machine to recognize specific patterns or trends within financial data and provide outcomes aligned with fraud detection objectives. Previous research has identified the accuracy levels of machine learning in detecting fraud and highlighted the commonly used detection tools (Ali et al., 2022; Constâncio et al., 2023; Mangala and Soni, 2023).

Despite numerous studies on financial fraud detection, a critical research gap emerges in the application of unsupervised learning techniques, particularly clustering algorithms, to address the complexities of fraud patterns. Ali et al. (2022) conclude that unsupervised learning, a subset of machine learning entailing clustering techniques, remains underrepresented in recent literature. Therefore, this technique still requires further research to identify the most effective methods in addressing increasingly complex and evolving fraud scenarios. Clustering techniques are particularly valuable for uncovering latent relationships and similarities, making them well-suited to handle instances of fraud (Ali et al., 2022). For that reason, exploring unsupervised learning becomes imperative to fully grasp its potential and advantages in the context of financial fraud detection. This research aims to bridge this gap by providing a comprehensive examination of the application of unsupervised learning in financial fraud detection. This urgency aligns with the benefits that can be extended to users leveraging machine learning to select the most applicable classification. This study employs a systematic literature review method and focuses on reviewing unsupervised learning within machine learning as applied to the detection of financial fraud. This method allows for a comprehensive examination of existing research and provides valuable insights into the efficacy of unsupervised learning within the context of fraud detection. Furthermore, it enables researchers to assess the influence of utilizing unsupervised learning in uncovering fraudulent financial transactions. Notably, this approach has not been extensively utilized in prior research on this topic.

To address the research gap, this study adopts a systematic approach to examine the application of unsupervised learning techniques, specifically clustering algorithms, in the domain of financial fraud detection. This study primarily aims to assess the effectiveness of unsupervised learning in detecting financial fraud. The focus is on analyzing the development and performance of unsupervised learning methods in identifying fraudulent financial transactions and considering their impact on industry-wide fraud detection. Additionally, the exploration encompasses both the advantages and limitations of unsupervised learning in addressing complex fraud scenarios. Conducted through 27 key articles from 2010 to 2023, the objective is to offer valuable insights into the applicability of unsupervised learning, enhancing fraud detection strategies. This research has substantial implications, enriching theoretical understanding and offering practical guidance. From a theoretical perspective, this research enhances understanding of the effectiveness, applications, cognitive concepts, benefits, challenges, and practical recommendations in the field of unsupervised learning, thereby contributing to the expanding body of knowledge in financial fraud detection. Based on this knowledge, practically, this research can be beneficial to at least three parties: regulatory authorities, business entities, and technology developers. The regulatory authorities can gain an understanding of the importance of financial fraud detection in day-to-day transactions and consider it to

Jurnal Riset Akuntansi dan Bisnis Airlangga Vol. 8. No. 2 (2023) 1521-1542
ISSN 2548-1401 (Print) ISSN 2548-4346 (Online)

be applied to all economic sectors. Businesses entities can also evaluate the implementation of technologies like machine learning to strengthen internal controls against financial fraud that could potentially have adverse effects on customers, clients, or even management itself. Moreover, this offers valuable insights to technology developers, aiding them in tackling unresolved issues through the enhancement of accuracy and efficiency within fraud detection systems. This, in turn, may lead to the mitigation of financial losses resulting from fraudulent activities.

Literature Review

Financial Fraud Detection

Financial fraud detection is a crucial process that involves identifying fraudulent activities within an organization (Ikhsan et al., 2022). Its significance lies in ensuring public trust in financial reports and the organization's performance, as well as preventing significant financial losses (Pourhabibi et al., 2020). This process entails data analysis, information gathering, and inquiry into financial documents and transactions (Adali and Kizil, 2017). One commonly used method in fraud detection is data analysis (Bănărescu, 2015). Data analysis helps uncover patterns or anomalies in financial transactions that serves as indicators of fraudulent activities (Kshetri, 2018). However, detecting fraud is a complex challenge, as it often occurs covertly, and perpetrators employ intricate strategies to conceal their actions (Mustika et al., 2021). The integration of technology, particularly data analytics, has proven effective and efficient in identifying fraudulent activities within company transaction data (Kock et al., 2017). Leveraging data analytics to detect fraud offers several advantages, including the ability to detect patterns and anomalies that are challenging for manual methods (Bănărescu, 2015).

Machine Learning in Fraud Detection

Machine learning, a subfield of artificial intelligence, is a crucial role in enhancing fraud detection processes (Goldberg and Holland, 1988). Machine learning algorithms are trained to identify patterns, make predictions, and classify data based on historical data (Mitchell, 1997). The machine learning process involves several steps: data preparation, feature engineering, model selection, training, evaluation, and deployment (Sarker, 2021). These steps ensure that machine learning models are developed, trained, and tested effectively to detect fraud. There are three primary types of machine learning: supervised learning, unsupervised learning, and reinforcement learning (Pugliese et al., 2021). Each type serves a specific role in solving various problems.

The Role of Unsupervised Learning

Unsupervised learning, in particular, plays a significant role in fraud detection by uncovering hidden patterns and anomalies within data without the need for labelled data (Sengupta et al., 2020). This technique is gaining importance in the field due to its potential to detect complex fraudulent behaviours effectively. While supervised and reinforcement learning have been extensively studied in fraud detection, unsupervised learning, with its ability to identify hidden patterns in financial data, has garnered less attention (Ashtiani and Raahemi, 2022).The research

Jurnal Riset Akuntansi dan Bisnis Airlangga Vol. 8. No. 2 (2023) 1521-1542
ISSN 2548-1401 (Print) ISSN 2548-4346 (Online)

focus is shifting toward the development and utilization of unsupervised learning techniques in fraud detection.

Research Method

This research utilized a Systematic Literature Review (SLR) methodology in the direction of enhancing the synthesized findings from the past research assessment. SLR is a research method that entails systematically identifying, assessing, and comprehensively interpreting all prior studies pertinent to the research question, subject area, or phenomenon being examined (Kitchenham and Charters, 2007). Identifying the addressed and examined issues is often an important component of SLR, serving to maintain the coherence and focus of the research. Research questions are designed with the assistance of the PICO framework (Population, Intervention, Comparison, and Outcomes) (Eldawlatly et al., 2018). Table 1 present the PICO framework and research questions.

Table 1. PICO Framework and Research Question

Structure	Description
Population	Financial fraud detection
Intervention	Application of Unsupervised Learning
Comparison	n/a
Outcomes	1. Understanding the Cognitive Construct of Financial Fraud Detection Using Unsupervised Learning 2. Understanding the Benefits and Economic Optimization of Financial Fraud Detection Using Unsupervised Learning 3. Understanding the Challenges of Financial Fraud Detection Using Unsupervised Learning

Research Question		
ID	**Inquiry**	**Motive**
RQ1	How is the cognitive construct regarding financial fraud detection using unsupervised learning?	Identifying the cognitive constructs of financial fraud detection using unsupervised learning
RQ2	What are the benefits and economic optimizations of financial fraud detection using unsupervised learning?	Identifying the benefits and economic optimization of financial fraud detection using unsupervised learning
RQ3	What are the challenges of financial fraud detection using unsupervised learning?	Identifying the challenges of financial fraud detection using unsupervised learning

Source: Researcher Analysis

The essential aspect of SLR involves a meticulous process of selecting research samples. Typically, this involves three key stages in the process of literature collection: identification, screening, and final assessment of samples included (depicted in Figure 1). This investigation focused on the Scopus database, chosen for its elevated article standards and extensive publication coverage. A broad time range for scientific articles was maintained to ensure comprehensive coverage of the research subject. The initial phase involved identifying relevant keywords to align with the research topic. To address the specific focus on unsupervised learning for financial fraud detection, the researchers formulated appropriate search terms using Boolean operators like "OR" and "AND." The search parameters employed for this

Jurnal Riset Akuntansi dan Bisnis Airlangga Vol. 8. No. 2 (2023) 1521-1542
ISSN 2548-1401 (Print) ISSN 2548-4346 (Online)

systematic literature review included: TITLE-ABS-KEY (("machine learning" OR "unsupervised learning" OR "clustering") AND ("fraud detection" OR "fraud" OR "accounting fraud detection")). This yielded a total of 3,283 articles spanning from 2010 to 2023. The scope was then refined by selecting the subject area "business, management, and accounting" and filtering for document types "article" and "final." Consequently, 50 articles were initially identified. Following a thorough assessment, 23 articles were excluded due to non-English languages (2) and relevance (21), resulting in the inclusion of 27 articles for detailed analysis within the timeframe of 2010 to 2023.

Once the final selection of articles is established, a descriptive analysis is carried out for the systematic literature review (SLR). This analysis encompasses various verification criteria, including examining publication trends over a specific time period, assessing publication types, identifying the most prestigious journals, authors, affiliations, and the countries of publication originate. In addition, the reference rates of the journals are examined to define their impact. Scopus database is employed to rank the often-referenced articles to classify the most significant publications. This approach facilitates the identification of publications that hold substantial importance and are frequently referenced by other researchers engaged in similar investigations, see figure 1.

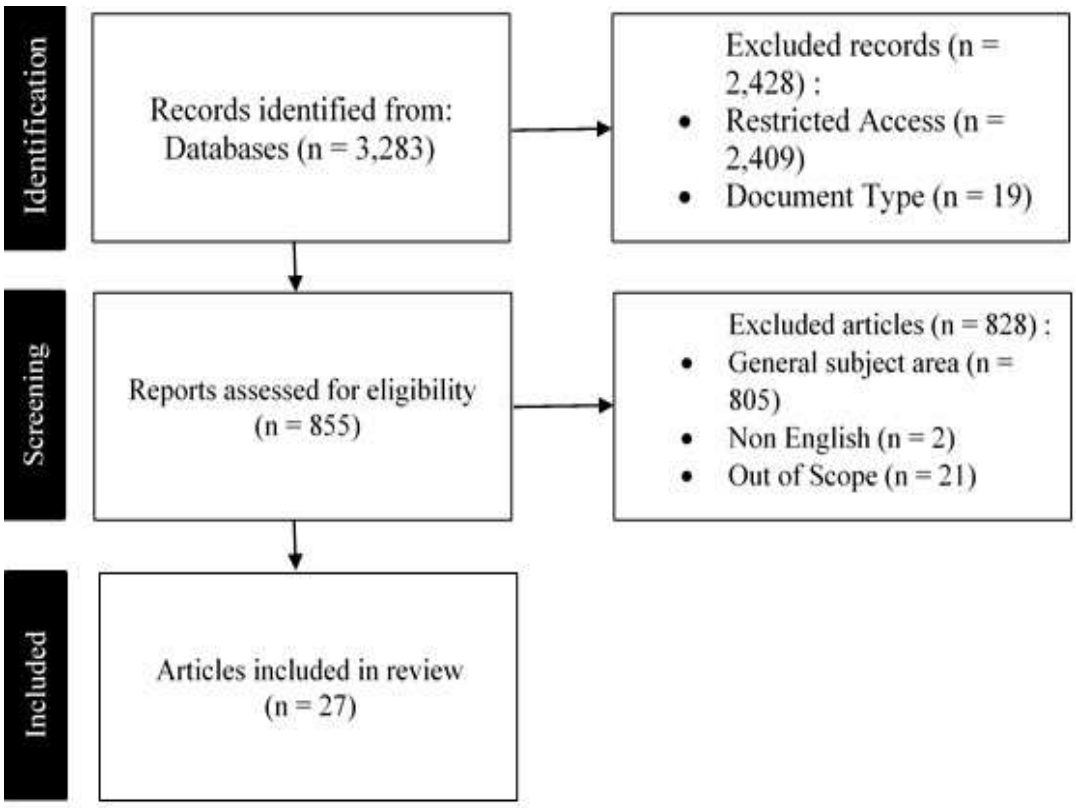


Figure 1. Sampling Procedure

Jurnal Riset Akuntansi dan Bisnis Airlangga Vol. 8. No. 2 (2023) 1521-1542
ISSN 2548-1401 (Print) ISSN 2548-4346 (Online)

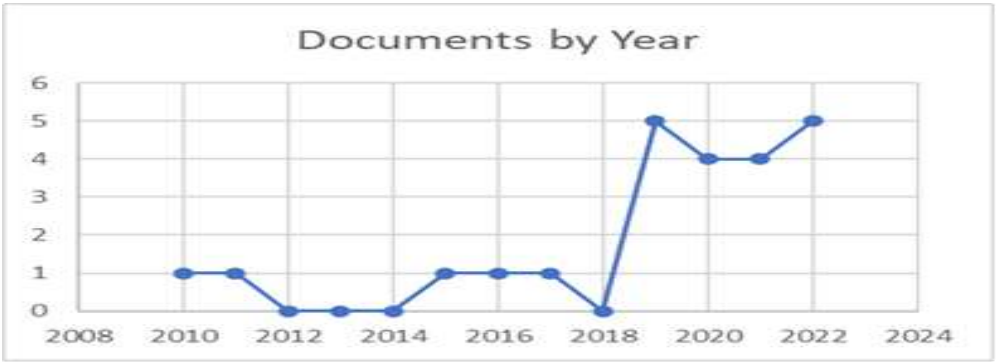
potential receipt of waste levies that have an impact on strategies for increasing Original Region Income. Furthermore, concluding means concluding from the results of the analysis that has been carried out previously following the problem formulation and research objectives. The results of this research will be validated using the data source triangulation method by comparing the primary and secondary data that have been obtained.

Results

The initial phase of the comprehensive performance analysis entails the rigorous examination and descriptive assessment of the 27 identified article sources. During this careful examination, we cover various critical aspects of general performance analysis. Firstly, it includes an assessment of the annual publishing volume, thereby explaining how scholarly output has changed over time. Closely analyzing and counting the articles in each journal provides valuable insights into the distribution of research across different publication platforms. Furthermore, an investigation into the order of publications is conducted predicated upon their citation counts, affording insights into the scholarly impact and influence of each individual article. Lastly, a categorization of articles predicated upon their countries of origin is carried out, facilitating an examination of the geographical contributions to the overarching research landscape. This multifaceted and systematic analysis offers a comprehensive and scientific perspective on the performance and impact of the selected articles.

Growth of Unsupervised Learning Application Literature

Graph 1 provides an insightful analysis of the quantity of publications pertaining to the intersection of machine learning and financial fraud. It's noteworthy that, during the dissemination item screening phase, no specific time restrictions were applied. The initial publication in this domain emerged in 2010, with a single article, followed by another in 2011. Over the subsequent three years, no publications matching the precise search criteria for unsupervised learning classification were identified, suggesting potential alternative categorizations in use during that period. The classification subsequently reappeared between 2015 and 2017, each year featuring a single article. However, there was a notable decline in 2018, with no publications matching the specified criteria, followed by a resurgence in the last five years. It's important to note that, as of August 2023, there may be limitations in capturing the most recent publications. Nevertheless, publications from 2023 have been included to provide insight into the sustained research interest in this subject.



Graph 1. Annual publishing volume

Jurnal Riset Akuntansi dan Bisnis Airlangga Vol. 8. No. 2 (2023) 1521-1542
ISSN 2548-1401 (Print) ISSN 2548-4346 (Online)

Continuing the analysis, it's apparent that there was a decline in both 2021 and 2022 in terms of publications related to machine learning and financial fraud. This fluctuation in research output underscores the dynamic nature of the field and may reflect shifting research priorities or trends within the academic community. Despite these variations, the inclusion of publications from 2023 in the analysis demonstrates an ongoing and sustained interest in the subject matter. This comprehensive examination of publication trends not only offers a historical perspective on research activity but also highlights the need for continued exploration and investigation in the intersection of machine learning and financial fraud detection, as evidenced by the recent resurgence in research output.

The Most Significant Affiliation on Unsupervised Learning Application

The research presents a comprehensive analysis of publications distributed across 16 distinct journals. Within this domain, it becomes evident that a predominant share of articles finds its home in the journal "Risks". This specific journal is esteemed for its research concentration on various dimensions of financial fraud risks, with a notable emphasis on the insurance industry. Another prominent contributor is the journal "Advances In Science Technology And Engineering Systems," which showcases a wide-ranging scope of interest spanning finance and engineering. Remarkably, both of these journals feature an equal share of four articles each, reflecting their significance in this research domain. Notably, Graph 2 provides an illustrative depiction of the cumulative count of publications pertaining to machine learning and financial fraud, with a pinpoint focus on these two influential journals.

Graph 2. Number of Articles in Each Journal

Jurnal Riset Akuntansi dan Bisnis Airlangga Vol. 8. No. 2 (2023) 1521-1542
ISSN 2548-1401 (Print) ISSN 2548-4346 (Online)

Continuing the examination, this distribution of articles across journals underscores the central role of "Risks" and "Advances In Science Technology And Engineering Systems" in disseminating research on machine learning and its application in tackling financial fraud. These journals serve as primary platforms for scholarly contributions in this field. The diverse range of topics covered within "Risks", particularly in the context of financial fraud risks within the insurance industry, demonstrates its multidisciplinary approach to addressing pertinent issues. Similarly, "Advances In Science Technology And Engineering Systems" exhibits a broad scope of interest, encompassing finance and engineering, which aligns with the interdisciplinary nature of research in machine learning for financial fraud detection. This insightful distribution of articles provides valuable guidance for researchers seeking to explore and contribute to this evolving field of study.

Comprehensive Reference Analysis

Table 2 provides a comprehensive overview of publications ranked by their citation counts, with a focus on works that have received more than 10 citations. In total, this analysis encompasses 28 publications with a collective citation count of 384. Notably, the study conducted by Thiprungsri and Vasarhelyi (2011) emerged as the most cited work, amassing a substantial number of citations. It is imperative to highlight that this study revolves around the critical intersection of machine learning and financial fraud detection, reflecting the significance of this research domain in garnering scholarly attention. Additionally, the second-ranking publication in terms of citations further underscores the enduring interest and relevance of research exploring the connection between machine learning and financial fraud, exemplifying its enduring impact on the academic landscape.

Table 2. Research Publications and References

No.	Title	Authors	No. of Reference
1	Cluster analysis for anomaly detection in accounting data: An audit approach	Thiprungsri, Sutapat., Vasarhelyi, Miklos A. Nian, Ke., Zhang,	69
2	Auto insurance fraud detection using unsupervised spectral ranking for anomaly	Haofan., Tayal, Aditya., Coleman, Thomas., Li, Yuying	66
3	Internal fraud risk reduction: Results of a data mining case study	Jans, Mieke., Lybaert, Nadine., Vanhoof, Koen	55
4	Big data-based fraud risk management at Alibaba	Chen, Jidong., Tao, Ye., Wang, Haoran., Chen, Tao	44
5	Data engineering for fraud detection	Baesens, Bart., Höppner, Sebastiaan., Verdonck, Tim	40
6	Detecting anomalies in financial statements using machine learning algorithm: The case of Vietnamese listed firms	Lokanan, Mark., Tran, Vincent., Vuong, Nam Hoai	24
7	Multi-level clustering-based outlier's detection (MCOD) using self-organizing maps	Li, Menglu., Kashef, Rasha., Ibrahim, Ahmed	14
8	Detection of Auction Fraud in Commercial Sites	Anowar, Farzana., Sadaoui, Samir	14
9	Computational intelligent hybrid model for detecting disruptive trading activity	Zhai, Jia., Cao, Yi., Yao, Yuan., Ding, Xuemei., Li, Yuhua	12

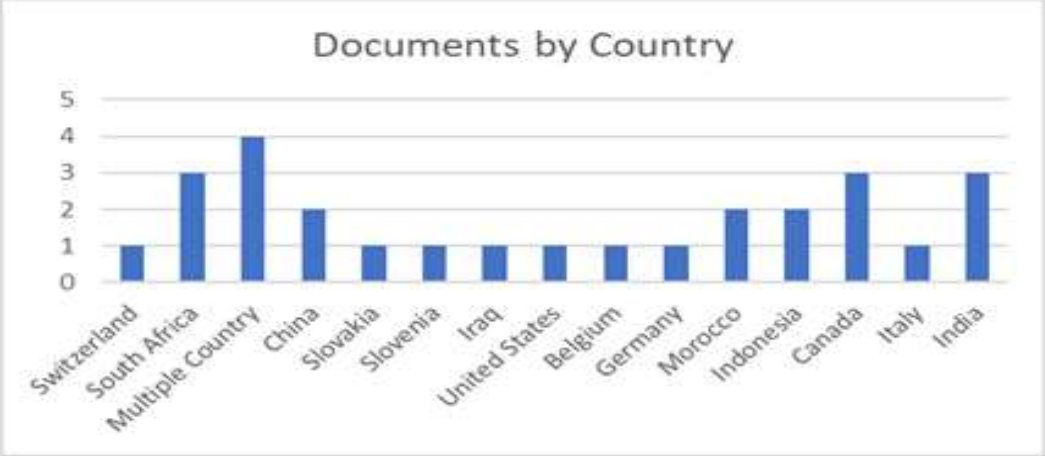
Source: Processed by the Author

Jurnal Riset Akuntansi dan Bisnis Airlangga Vol. 8. No. 2 (2023) 1521-1542
ISSN 2548-1401 (Print) ISSN 2548-4346 (Online)

Moreover, this citation analysis highlights the lasting influence and significance of research in the crossroads of machine learning and financial fraud detection. It demonstrates the enduring scholarly interest in this field, as reflected in the considerable citations these publications have received. This enduring impact establishes machine learning in financial fraud detection as a critical area of research, highlighting its ongoing importance.

Publications Trend by Geographical Distribution

Graph 3 provides an insightful breakdown of the geographical distribution of articles concerning financial fraud detection using machine learning. Among the 27 articles included in the analysis, the focus is predominantly directed towards 15 distinct countries. It is noteworthy that research in this domain extends across 23 individual countries, demonstrating the global reach and relevance of this research topic. Evidently, a significant portion of the studies, comprising 4 publications, falls under the category of "Multiple Country," suggesting that these investigations encompass data or scenarios involving multiple nations. South Africa, Canada, and India emerge as key contributors to this body of knowledge, each accounting for 3 publications. Additionally, research efforts have been channeled into countries such as China, Morocco, and Indonesia, resulting in 2 publications from each of these regions. This diverse geographic distribution underscores the international significance of machine learning applications in financial fraud detection, with researchers from various nations actively engaging in this field.



Graph 3. Categorization of articles based on countries

Overall, the findings illustrate a global collaborative effort in understanding and combatting financial fraud, with researchers pooling data and insights from multiple countries. This cooperative approach reflects the shared commitment to addressing the challenges posed by financial fraud on a global scale, fostering international cooperation and knowledge exchange in the pursuit of effective fraud detection solutions.

Jurnal Riset Akuntansi dan Bisnis Airlangga Vol. 8. No. 2 (2023) 1521-1542
ISSN 2548-1401 (Print) ISSN 2548-4346 (Online)

Discussion

Within this section, we provide a detailed account the outcomes of our analysis and synthesis of the data review, all of which are closely aligned with the research inquiries that were meticulously formulated at the outset of this study. Our thorough examination of the chosen articles serves as the cornerstone for revealing the conclusions drawn from the literature review.

RQ1 : How is the Cognitive Construct Regarding Unsupervised Learning-Based Fraud Detection Structured?

This section discusses a comprehensive exploration of the various algorithms used in fraud detection across diverse industries. To effectively identify patterns and anomalies associated with fraudulent activities, machine learning approaches employed to identify patterns and anomalies are categorized based on the dataset used to determine the suitable machine learning classification (Mqadi et al., 2021). This dataset comprises historical or real-time updated transaction records, a crucial component in the ongoing battle against potential fraud (Lokanan et al., 2019). The continuous evolution of machine learning techniques has empowered fraud detection systems to directly capture intricate fraud signals from extensive user behavior data and network activity. This capability enables real-time analysis through machine learning algorithms, leading to precise predictions of both fraudulent users and potentially fraudulent transactions (Chen et al., 2015). Pattern recognition involves assessing patterns, classes, or clusters of suspicious actions either routinely or to match assumed inputs (Vadakara and Kumar, 2019). As highlighted by Ali and Ahmed (2019), there is a scarcity of research concerning clustering classification. Within the scope of this systematic literature review that prioritizes the utilization of unsupervised learning classification, several articles also adopt alternative classifications, as outlined in table 4. This multifaceted approach to classification underscores the dynamic nature of fraud detection techniques and the ongoing efforts to enhance their effectiveness and adaptability in combating financial fraud across various sectors and scenarios.

Unsupervised anomaly detection operates without a training dataset and relies on the assumption of significantly fewer anomalies compared to normal instances, making cluster analysis a suitable choice for tasks like fraud and anomaly detection due to its ability to circumvent this challenge (Thiprungsri and Vasarhelyi, 2011). Clustering represents an unsupervised learning method, where data is examined without predefined labels (for instance, "fraudulent" or "non-fraudulent") (Kachigan, 1991). Clustering analysis is a specific technique within the broader concept of unsupervised learning (Li et al., 2020). Clustering analysis manages data without labels, addresses data scarcity, and operates effectively on datasets with diverse configurations. This is an advantage of unsupervised learning over supervised learning. Despite the substantial potential of unsupervised learning through the utilization of data without labels, a considerable number of individuals still choose to employ labeled data due to the evaluation criteria of supervised learning, which emphasizes accuracy and prediction precision.

Jurnal Riset Akuntansi dan Bisnis Airlangga Vol. 8. No. 2 (2023) 1521-1542
ISSN 2548-1401 (Print) ISSN 2548-4346 (Online)

The tables 3 and 4 showcase a diverse range of machine learning algorithms categorized into unsupervised and supervised techniques. These algorithms, whether employed individually or in tandem, are tailored to address distinct fraud areas. They offer a versatile toolkit for detecting financial fraud with precision and efficiency. In the following sections, we explore these algorithms further, highlighting their specific applications in the evolving field of fraud detection.

Table 3. Data of Unsupervised Learning Algorithms

No.	Technique	Fraud Area
1	• Bayesian Anomaly Detection	• Insurance Claims
2	• K-nearest Neighbors Distance • Local Outlier Factor (LOF) • Isolation Forests	• Debt Card Transaction
3	• Generative Adversarial Network (GAN)	• Credit Card Fraud
4	• Multi-Level Outlier Detection Algorithm (MCOD)	• Online Transaction in Business Application
5	• Local Outlier Factor, Isolation Forest, Bagged Decision Trees	• Online Transaction Banking Fraud
6	• K-Means	• Fraudulent Calls in Telecommunication Industry
7	• DBSCAN Algorithm	• Sale of Bus Tickets by Conductors
8	• K-Means (Cluster Analysis)	• Group Life Insurance Claims
9	• Latent class clustering algorithm	• Internal Fraud in a Case Company's Procurement Data
10	• Graph Analysis Algorithms	• Cryptocurrency Transaction
11	• Spectral Ranking Method	• Auto Insurance Claims

Source: Processed by the Author

In examining the table, it becomes evident that unsupervised learning techniques, such as Bayesian Anomaly Detection, Generative Adversarial Networks (GAN), and clustering methods like K-means and DBSCAN, are frequently utilized in a range of fraud areas. These techniques are particularly valuable for identifying irregular patterns and anomalies in diverse datasets, making them well-suited for applications like detecting fraudulent transactions, insurance claims, and online banking fraud. On the other hand, supervised learning techniques, including Support Vector Machines (SVM), Decision Trees, Random Forests, and Neural Networks, find extensive use in various fraud detection tasks. These algorithms are employed for tasks such as credit card fraud detection, assessing the creditworthiness of firms, and detecting disruptive trading behavior in financial markets.

The array of methods outlined in table 4 illustrates how machine learning can effectively respond to various fraud scenarios within the financial sector. Researchers and industry experts can gain valuable insights from this range of approaches, helping them choose the best-suited algorithms for their specific fraud detection requirements. This diversity mirrors the ever-changing nature of financial fraud and the necessity for flexible tools to combat it efficiently. By comprehending and utilizing this spectrum of techniques, researchers and professionals in the industry can create more resilient and customized fraud detection systems to protect financial assets effectively.

Jurnal Riset Akuntansi dan Bisnis Airlangga Vol. 8. No. 2 (2023) 1521-1542
ISSN 2548-1401 (Print) ISSN 2548-4346 (Online)

Table 4. Data of Supervised Learning Algorithms

No.	Technique	Fraud Area
1	• Factorisation Machines	• Credit Card Fraud
2	• Support Vector Machine (SVM) • Logistic Regression (LR) • Decision Tree (DT) • Random Forest	• Credit Card Fraud
3	• Neural network	• Payment Systems on E-Commerce, Banking, and Financial Service
4	• Support Vector Machine (SVM) • Artificial Neural Network (ANN) with MultiLayer Perceptron (MLP) • Random Forest	• Shill Bidding in Online Auctions
5	• Support Vector Machine (SVM) • Random Forest (RF) • Navie Bayes • Deep Convolution Neural Network (DCNN)	• Anomalous Financial Reports
6	• Support Vector Machine (SVM) • Hidden Markov Model (HMM)	• Disruptive Trading Behavior in Financial Markets
7	• Logistic Regression	• Mobile Internet and Finance
8	• DeepFool algorithm	• Health Insurance Claims
9	• Multi-Task Learning	• Loan1 Application Assessments
10	• Classification Tree	• Corruption Crimes
11	• Gradient Boosting Algorithms	• Credit Card Fraud
12	• Random Forest • K-Nearest Neighbor (KNN) • Supprot Vector Machine (SVM) • Naive Bayes • RPART • NNET • Latent Dirichlet Allocation (LDA) • Quadratic Discriminant Analysis (QDA)	• Health Care Claims
13	• Logistic Regression • Artificial Neural Networks (ANN) • Fuzzy Logic • Ensemble-based Methods	• The Credit Worthiness of a Firm's Quarterly Financial Report
14	• e EKMC (Ensemble of kNN using MetaCost)	• Fraud Detection
15	• Aggrandized random forest	• Credit Card Transactions/ATM Transactions
16	• Classification Tree	• Online Banking Transactions

Source: Processed by the Author

RQ2 : What are the Benefits and Economic Optimizations of Implementing Unsupervised Learning in Financial Fraud Detection?

The practical application of fraud activities varies across different industry sectors. This section intends to respond to the research question (RQ2) by showcasing the application of machine learning in diverse fraudulent activities based on chosen articles. The objective is to provide an insightful analysis of how machine learning techniques have been leveraged to combat fraud, drawing from the selected articles. Upon conducting an extensive literature review, it becomes evident that fraudulent activities within the financial sector can be broadly categorized into several key areas, including but not limited to credit card fraud, insurance claims fraud, fraudulent online transactions, and various other types of financial fraud. These distinct categories of fraudulent activities will be explored in greater detail in the subsequent subsections, shedding light on the multifaceted nature of fraud detection and prevention across different sectors.

Jurnal Riset Akuntansi dan Bisnis Airlangga Vol. 8. No. 2 (2023) 1521-1542
ISSN 2548-1401 (Print) ISSN 2548-4346 (Online)

Credit Card Fraud

Due to the high volume of credit card transactions, it becomes challenging to detect financial fraud occurrences (Mytnyk et al., 2023). This situation contributes to data imbalance. Consequently, this imbalance leads to a scarcity of representation for fraudulent data in contrast to legitimate transactions, potentially affecting the efficacy of models and analyses employing this dataset. Ngwenduna and Mbuvha (2021) propose an innovative unsupervised learning approach that harnesses Generative Adversarial Networks to generate high-quality data, thereby addressing the scarcity of fraudulent data representation. Moreover, when constructing a credit rating model, the unique characteristics of credit samples often result in a shortage of scores for minority class samples. In other words, when dealing with a substantial number of genuine samples, this can introduce a bias in machine learning models during the training phase (Alothman et al., 2022). This finding underscores the importance of tackling data imbalance to enhance the accuracy and reliability of fraud detection methods in the realm of credit card transactions.

Insurance Claims

In the practical context within the corporate insurance sector, when initiating a business process designed to prevent fraud, it is normal that no instances of fraudulent claims have been detected yet, emphasizing the necessity of employing the classification of unsupervised learning. The absence of legitimate labels of fraud places a greater stress on the process of attribute engineering. In numerous data mining scenarios, acquiring labels comes with high costs and consumes a significant amount of time, unless it's practically unfeasible (Nian et al., 2016). A short time difference between the policy start date and the claim serves as an indicator of potential fraudulent behavior. Hence, Vosseler (2022) delves more profoundly into the creation of relevant features associated with fraud to capture possible indications of deceptive claims.

Bayesian Histogram-based Anomaly Detector exhibits linear complexity in relation to the sample size and quantity of attributes, enabling a direct interpretation of single anomaly ratings. Moreover, it efficiently handles both continuous and discrete features, eliminating the need for bin partitioning in the latter case. Furthermore, there exists a spectral ranking method for anomalies, which includes an unsupervised approach suitable for situations involving costly or unavailable labeled data, the ability to address various types of features, including categorical ones, and its technique in spectral ranking that showcases its efficacy in producing anomaly rankings without the requirement for labels. Moreover, the scalability of SRA, its demonstrated excellence in fraud detection, and its flexibility in adapting to diverse similarity measures all enhance the significance of SRA in the realm of anomaly detection tasks. The size and complexity of health insurance programs necessitate the use of analytical methods for audit processes to detect complex patterns (Ekin et al., 2021). In the insurance sector, artificial intelligence functions as a versatile tool that holds the promise of delivering positive outcomes in a range of areas, including pricing,underwriting, marketing, claims management, and post-sales services (Amerirad et al., 2023).

Jurnal Riset Akuntansi dan Bisnis Airlangga Vol. 8. No. 2 (2023) 1521-1542
ISSN 2548-1401 (Print) ISSN 2548-4346 (Online)

Online Transaction

Fraud in online transactions occurs when criminals manage to take control of someone's internet-based bank account and move funds from it (Vanini et al., 2023). Similarly, anomalies in cryptocurrency transactions might involve fraudulent activities, money laundering, or other illicit actions. The decentralized and relatively anonymous nature of online transactions often leads criminals to exploit this system for unlawful purposes (Vičič and Tošić, 2022). This situation poses a machine learning challenge due to imbalanced data and intricate fraud complexities (Vanini et al., 2023). Detecting unauthorized online transactions can rely on spotting patterns through outlier detection (Li et al., 2020).

The majority of payment systems utilize the database types, with well-documented papers describing the migration of schemas and data, including algorithms and rules (Orche and Bahaj, 2020). The multi-level outlier detection algorithm (MCOD) utilizes unsupervised multi-level learning to array the data and pinpoint continuous and efficient outliers, aiming to enhance profitability or elevate business outcomes. Vanini et al. (2023) employed The Local Outlier Factor (LOF), suitable for detecting outlier, assigning a degree of outlier to each observation depending on how far it is from the closest group of nearby observations. The goal is to spot anomalies in non-uniform data. Graph analysis algorithms can unveil intricate patterns and relationships in data represented as graphs or networks, proving beneficial for identifying elusive patterns that traditional analysis methods might miss (Vičič and Tošić, 2022).

Other Types of Financial Fraud

In addition to the previously explained types of fraud in the financial sector, there exist various other forms of fraudulent activities within the financial domain. These encompass fraud linked to fraudulent transactions carried out via phone calls or other telecommunication channels (Jabbar and Suharjito, 2020), deceptive practices in bus ticket sales (Wihartiko and Wihartika, 2019), and instances of internal corporate fraud (Jans et al., 2010). By analyzing the patterns present in Call Detail Records, which contain the records of customer conversations, including details like the number source and its destination, starting timestamps, and call durations, it's possible to detect suspicious or irregular patterns within these communication records. The K-Means clustering technique is connected to group Call Detail Record (CDR) data based on comparable patterns, yielding higher accuracy (Jabbar and Suharjito, 2020).

Fraudulent activities in bus ticket sales can be categorized based on geographic (spatial) and chronological (temporal) factors. Identifying such patterns or data clusters characterized by these attributes can aid in identifying unusual or suspicious behaviors within the established patterns. The applied algorithm employs spatio-temporal clustering (Wihartiko and Wihartika, 2019). Addressing the risk of internal corporate fraud encompasses a comprehensive approach of diminishing fraudulent practices in the procurement process by merging detection and prevention strategies. Notably, the utilization of latent class clustering permits the consideration of overlapping groups, thereby allowing for the examination of outliers where an observation might not align with any particular group ((Jans et al., 2010).

Jurnal Riset Akuntansi dan Bisnis Airlangga Vol. 8. No. 2 (2023) 1521-1542
ISSN 2548-1401 (Print) ISSN 2548-4346 (Online)

RQ3 : What are the Challenges of Implementing Unsupervised Learning in Detecting Financial Fraud?

This section addresses the challenges arising in conducting fraud detection using machine learning to address the research question. The primary challenges in implementing machine learning for fraud detection encompass class imbalance and fraud complexity. Both of these challenges require machine learning model developers to design precise strategies. This aims to achieve an effective, efficient, and accurate model for detecting various types of fraudulent transactions.

Class imbalance occurs when numeral data instances among the same category significantly outweighs the others (Ahmed and Mahmood, 2015). Regarding fraud detection, valid transactions have a tendency to outnumber fraudulent ones (Slabber et al., 2023). Optimizing the rating effect of the model on imbalanced data has become the focus of machine learning algorithm selection to avoid misclassification of fraud classes (Alothman et al., 2022). Within the scope of credit card fraud detection, the minority class identified as fraudulent transactions only constitutes 0.02% or less of the data (Alothman et al., 2022). As a result, machine learning models tend to predict transactions as legitimate because the majority of the data consists of valid transactions (Mqadi et al., 2021). In this scenario, accuracy measurements become unreliable as they introduce bias (Ngwenduna and Mbuvha, 2021). This can result in subpar effectiveness in identifying fraudulent transactions and a high rate of false positives, where fraudulent transactions are misclassified as legitimate (Ahmed and Mahmood, 2015). Datasets often tend to be imbalanced, and evaluation can be carried out using various advanced oversampling, undersampling, and hybrid-sampling methods while comparing their performance across multiple classification algorithms (Niranjan et al., 2019). Imbalanced dataset diminishes the predictive efficacy of the classifiers (Anowar and Sadaoui, 2020).

Fraud complexity is another challenge in fraud detection using machine learning. Fraud often involves various strategies and techniques to deceive detection systems. Fraudsters continually develop new methods of fraud, which make it difficult for existing models to detect them (Baesens et al., 2021). The majority of financial institutions afford complex security solutions to the fraudsters who adjust their approaches over a period (Deepika and Senthil, 2019). Fraud patterns can be highly diverse and change over time, making it challenging to understand these patterns and adapt to changes in fraud tactics (Jabbar and Suharjito, 2020). One of the datasets that makes modeling complex is non-stationary financial data (Zhai et al., 2017). Anomalies in accounting data also change over time (Thiprungsri and Vasarhelyi, 2011). The complexity is influenced by the dataset used as it requires considering more comprehensive attributes in the fraud detection process (L. Chen et al., 2022). Algorithms must be able to adapt to various types of fraud, including changes in data patterns and outlier behavior over time (Li et al., 2020). Fraud detection must occur in real-time, so algorithms must operate quickly and efficiently (Wihartiko and Wihartika, 2019). Complexity is also closely related to the interpretability of modeling results, which can be challenging to analyze as decision-masing tools (gwenduna and Mbuvha, 2021). Algorithms must be able to identify unusual and unpredictable patterns, which can be a challenge (Vanini et al., 2023). More complex algorithms can provide better performance (de Blasio et al., 2022).

Jurnal Riset Akuntansi dan Bisnis Airlangga Vol. 8. No. 2 (2023) 1521-1542
ISSN 2548-1401 (Print) ISSN 2548-4346 (Online)

Conclusion

This study reveals the function of unsupervised learning in financial fraud detection. There is some algorithms used in fraud detection across industries, categorizing machine learning approaches based on datasets for suitable classification. Unsupervised anomaly detection, like cluster analysis, is emphasized for its ability to handle instances where anomalies are significantly fewer than normal cases, operating effectively without predefined labels. It highlights credit card fraud, insurance claims, and online transaction fraud as specific cases. In unsupervised learning classification, K-Means is the most popular method. Challenges that still need to be lectured for the effective implementation of machine learning include class imbalance and the complexity of fraud. In conclusion, this work emphasizes the significance of detecting financial fraud detection with machine learning, especially in the context of developing transaction types that lack predetermined labels for classification as financial fraud. Furthermore, the findings of the study emphasize the potential of machine learning.

Limitation

It is important to acknowledge certain limitations inherent to this study. The research only reviewed a small number of articles, specifically 27 articles, due to the highly specific focus on the usage of unsupervised learning as a rarely explored tool in the literature, despite its advantages in detecting unlabeled data. The article sources were limited to the Scopus database to ensure the quality of high-impact articles. Additionally, the data collection process was restricted to August 2023, excluding recent publications that might be relevant to the evolving field of unsupervised learning and financial fraud detection. Consequently, recent publications that may have relevance to the evolving landscape of unsupervised learning and financial fraud detection were not included in our analysis. Despite these limitations, we are confident that this SLR will continue to serve as a valuable resource in the times ahead.

Suggestions

Based on the limitations of this study as previously explained, future research suggestions include comprehensively examining the technologies that can be utilized for financial fraud detection. Each detection tool has its advantages and uniqueness in its application. To broaden this scope, future literature can compare the implementation of technologies beyond machine learning, such as deep learning and blockchain, thus making substantial contributions to this field by creating innovative strategies, algorithms, and techniques. These advancements have the potential to enhance the effectiveness of financial fraud detection, reduce risks, and strengthen the security of transactions and operations across various industries like insurance, finance, banking, and trading. To expand the reviewed articles, future research can also consider adding other databases in the selection process. In addition to Scopus, other databases like Web of Science and Google Scholar can be considered for inclusion. It is important to emphasize that as these databases expand, it becomes

Jurnal Riset Akuntansi dan Bisnis Airlangga Vol. 8. No. 2 (2023) 1521-1542
ISSN 2548-1401 (Print) ISSN 2548-4346 (Online)

crucial to maintain rigorous inclusion and exclusion criteria to create a comprehensive literature review. Furthermore, the time limit should be adjusted to account for the rapid and extensive technological advancements, ensuring that the data remains up-to-date. Beyond addressing the limitations highlighted in this study, this research strongly advocates for conducting further studies related to the advancement of technology-based approaches customized to the unique characteristics of datasets, particularly as effective tools for combating financial fraud.

Implication

This research contributes to both knowledge and practicality. While unsupervised learning has found prominence in areas like credit card fraud and online transactions, its potential extends to uncharted territories in fraud detection, risk management, and AI technology. This knowledge can empower businesses including organizations, regulatory bodies, and technology developers to design more effective financial fraud detection strategies, addressing the ever-evolving complexities of fraudulent activities.

Reference

ACFE. 2022. Occupational Fraud 2022: A Report to the nations. *Association of Certified Fraud Examiners.* Pp. 1–96.

Adali, S. and Kizil, C. 2017. A Research on the Responsibility of Accounting Professionals to Determine and Prevent Accounting Errors and Frauds: Edirne Sample. *EMAJ: Emerging Markets Journal*. Vol. 7 No. 1. Pp.2158–8708.

Ahmed, M. and Mahmood, A. N. 2015. Novel Approach for Network Traffic Pattern Analysis using Clustering-based Collective Anomaly Detection. *Annals of Data Science*. Vol. 2 No. 1. Pp. 111–130.

Ali, R. and Ahmed, S. 2019. Machine Learning Applications in Accounting and Finance: A Review. *Journal of Accounting & Marketing.* Vol. 8 No. 2. Pp.285–290.

Ali, A., Abd Razak, S., Othman, S. H., Eisa, T. A. E., Al-Dhaqm, A., Nasser, M., Elhassan, T., Elshafie, H., and Saif, A. 2022. Financial Fraud Detection Based on Machine Learning: A Systematic Literature Review. *Applied Sciences (Switzerland).* Vol. 12 No. 9637. Pp.1-24.

Alothman, R., AliTalib, H., and Mohammed, M. S. 2022. Fraud Detection Under the Unbalanced Class Based on Gradient Boosting. *Eastern-European Journal of Enterprise Technologies.* Vol. 2 No.1. Pp. 6–12.

Amerirad, B., Cattaneo, M., Kenett, R. S., and Luciano, E. 2023. Adversarial Artificial Intelligence in Insurance: From an Example to Some Potential Remedies. *Risks.* Vol. 11 No. 20. Pp. 1-17.

Jurnal Riset Akuntansi dan Bisnis Airlangga Vol. 8. No. 2 (2023) 1521-1542
ISSN 2548-1401 (Print) ISSN 2548-4346 (Online)

Anowar, F. and Sadaoui, S. 2020. Detection of Auction Fraud in Commercial Sites. *Journal of Theoretical and Applied Electronic Commerce Research.* Vol. 15 No. 1. Pp. 81–98.

Ashtiani, M. N., and Raahemi, B. 2022. Intelligent Fraud Detection in Financial Statements Using Machine Learning and Data Mining: A Systematic Literature Review. *IEEE Access.* Vol. 10 No.2.Pp. 72504–72525.

Baesens, B., Höppner, S., and Verdonck, T. 2021. Data engineering for fraud detection. *Decision Support Systems.* Vol. 150 No.2. Pp.113-129.

Bănărescu, A. 2015. Detecting and Preventing Fraud with Data Analytics. In *Procedia Economics and Finance.* Vol. 32. No.3. Pp.1827-1836.

Carneiro, E. M., Dias, L. A. V., Cunha, A. M. Da, and Mialaret, L. F. S. 2015. Cluster Analysis and Artificial Neural Networks: A Case Study in Credit Card Fraud Detection. *Proceedings - 12th International Conference on Information Technology: New Generations.* Pp. 122–126.

Chen, J., Tao, Y., Wang, H., and Chen, T. 2015. Big data based fraud risk management at Alibaba. *Journal of Finance and Data Science.* Vol. 1 No. 1. Pp. 1–10.

Chen, L., Jia, N., Zhao, H., Kang, Y., Deng, J., and Ma, S. (2022). Refined analysis and a hierarchical multi-task learning approach for loan fraud detection. *Journal of Management Science and Engineering.* Vol. 7 No. 4. Pp. 589–607.

Constâncio, A. S., Tsunoda, D. F., Silva, H. de F. N., Silveira, J. M. da, and Carvalho, D. R. 2023. Deception detection with machine learning: A systematic review and statistical analysis. *PloS One.* Vol. 18 No. 2. Pp. 1-31.

de Blasio, G., D'Ignazio, A., and Letta, M. 2022. Gotham city. Predicting 'corrupted' municipalities with machine learning. *Technological Forecasting and Social Change.* Vol. 184 No.2.Pp. Pp. 1-27.

Deepika, S. and Senthil, S. 2019. Credit card fraud analysis using robust space invariant artificial neural networks (RSIANN). *International Journal of Recent Technology and Engineering.* Vol. 8 No. 2. Pp. 6413–6417.

Ekin, T., Frigau, L., and Conversano, C. 2021. Health care fraud classifiers in practice. *Applied Stochastic Models in Business and Industry.* Vol. 37. No. 6. Pp. 1182–1199.

Eldawlatly, A., Alshehri, H., Alqahtani, A., Ahmad, A., Al-Dammas, F., and Marzouk, A. 2018. Appearance of Population, Intervention, Comparison, and Outcome as research question in the title of articles of three different anesthesia journals: A pilot study. *Saudi Journal of Anaesthesia.* Vol. 12 No. 2. Pp. 283–286.

Goldberg, D. E., and Holland, J. H. 1988. Genetic Algorithms and Machine Learning. *Machine Learning.* Vol. 3 No.2. Pp. 95–99.

Gyamfi, N. K., and Abdulai, J. D. 2018. Bank Fraud Detection Using Support Vector Machine. *2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference, IEMCON 2018.* Pp. 37–41.

Ikhsan, W. M., Ednoer, E. H., Kridantika, W. S., and Firmansyah, A. 2022. Fraud Detection Automation Through Data Analytics and Artificial Intelligence. *Riset.* Vol. 4. No.2. Pp.103–119.

Jurnal Riset Akuntansi dan Bisnis Airlangga Vol. 8. No. 2 (2023) 1521-1542
ISSN 2548-1401 (Print) ISSN 2548-4346 (Online)

Jabbar, M. A., and Suharjito. 2020. Fraud detection call detail record using machine learning in telecommunications company. *Advances in Science, Technology and Engineering Systems.* Vol. 5 No. 4. Pp.63–69.

Jans, M., Lybaert, N., and Vanhoof, K. 2010. Internal fraud risk reduction: Results of a data mining case study. *International Journal of Accounting Information Systems.* Vol.11 No.1.Pp. 17–41.

Kachigan, S. K. 1991. Multivariate Statistical Analysis: a Conceptual Introduction. In *Radius Press*. Radius Press.

Kitchenham, B. A. and Charters, S. 2007. Guidelines for performing Systematic Literature Reviews in Software Engineering. *Technical Report EBSE 2007-001. Keele University and Durham University Joint Report*.

Kock, F., Georgi, M., and Schöndube, J. 2017. Fraud detection using data analytics in the banking industry. *Journal of Money Laundering Control.* Vol. 20 No. 4. Pp. 441–456.

Koerniawati, D. 2021. the Remote and Agile Auditing: a Fraud Prevention Effort To Navigate the Audit Process in the Covid-19 Pandemic. *Jurnal Riset Akuntansi Dan Bisnis Airlangga.* Vol. 6 No. 2. Pp. 1131–1149.

Kshetri, N. 2018. 5G in E-Commerce Activities. *IT Professional*. Vol. 20 No. 4. Pp. 73–77.

Kumari, N., Kannan, S., and Muthukumaravel, A. 2014. Credit card fraud detection using Hidden Markov Model-A survey. *Middle - East Journal of Scientific Research.* Vol. 20 No. 6.Pp. 697–699.

Li, M., Kashef, R., and Ibrahim, A. 2020. Multi-level clustering-based outlier's detection (MCOD) using self-organizing maps. *Big Data and Cognitive Computing.* Vol. 4 No.4.Pp. 1–17.

Lokanan, M., Tran, V., and Vuong, N. H. 2019. Detecting anomalies in financial statements using machine learning algorithm: The case of Vietnamese listed firms. *Asian Journal of Accounting Research.* Vol. 4 No.2. Pp.181–201.

Mangala, D. and Soni, L. 2023. A systematic literature review on frauds in banking sector. *Journal of Financial Crime.* Vol. 30 No.1. Pp.285–301.

Mitchell, T. 1997. *Machine learning*. New York: McGraw Hill.

Mqadi, N., Naicker, N., and Adeliyi, T. 2021. A SMOTe based oversampling data-point approach to solving the credit card data imbalance problem in financial fraud detection. *International Journal of Computing and Digital Systems.* Vol. 10 No.1. Pp.277–286.

Mustika, N. I., Nenda, B., and Ramadhan, D. 2021. Machine Learning Algorithms in Fraud Detection: Case Study on Retail Consumer Financing Company. *Asia Pacific Fraud Journal.* Vol. 6 No. 2.Pp.213–221.

Mytnyk, B., Tkachyk, O., Shakhovska, N., Fedushko, S., and Syerov, Y. 2023. Application of Artificial Intelligence for Fraudulent Banking Operations Recognition. *Big Data and Cognitive Computing.* Vol. 7 No.93.Pp.1-19.

Ngwenduna, K. S., and Mbuvha, R. 2021. Alleviating class imbalance in actuarial applications using generative adversarial networks. *Risks.* Vol. 9 No.3.Pp.1–33.

Jurnal Riset Akuntansi dan Bisnis Airlangga Vol. 8. No. 2 (2023) 1521-1542
ISSN 2548-1401 (Print) ISSN 2548-4346 (Online)

Nian, K., Zhang, H., Tayal, A., Coleman, T., and Li, Y. 2016. Auto insurance fraud detection using unsupervised spectral ranking for anomaly. *Journal of Finance and Data Science.* Vol. 2 No. 1. Pp. 58–75.

Niranjan, A., Akshobhya, K. M., Deepa Shenoy, P., and Venugopal, K. R. 2019. EKMC: Ensemble of kNN using MetaCost for Efficient Anomaly Detection. *Advances in Science, Technology and Engineering Systems.* Vol. 4 No. 5. Pp. 401–408.

Normasari, E., and Sekar Mayangsari. 2022. Influence of Fraud Star and Digital Banking on Ffr in Banking Sector and the Moderating Role of Foreign Ownership. *Jurnal Riset Akuntansi Dan Bisnis Airlangga.* Vol. 7 No. 2. Pp. 1319–1342.

Orche, A. E. and Bahaj, M. 2020. Approach to combine an ontology-based on payment system with neural network for transaction fraud detection. *Advances in Science, Technology and Engineering Systems.* Vol. 5 No. 2. Pp. 551–560.

Pourhabibi, T., Ong, K. L., Kam, B. H., and Boo, Y. L. 2020. Fraud detection: A systematic literature review of graph-based anomaly detection approaches. *Decision Support Systems.* Vol. 133. No.2. Pp.113-136.

Pugliese, R., Regondi, S., and Marini, R. 2021. Machine learning-based approach: Global trends, research directions, and regulatory standpoints. *Data Science and Management.* Vol. 4 No.2. Pp. 19–29.

Raghavan, P. and Gayar, N. El. 2019. Fraud Detection using Machine Learning and Deep Learning. *Proceedings of 2019 International Conference on Computational Intelligence and Knowledge Economy.* Pp. 334–339.

Saeidi S.P. 2020. Detecting Financial Statement Fraud Using Machine Learning Algorithms: A Comprehensive Review of the Literature. *Journal of Accounting, Auditing & Finance.* Vol. 35 No. 1. Pp. 115–139.

Sánchez-Aguayo, M., Urquiza-Aguiar, L., and Estrada-Jiménez, J. 2021. Fraud detection using the fraud triangle theory and data mining techniques: A literature review. *Computers.* Vol.10 No.1. Pp. 1–23.

Sarker, I. H. 2021. Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. *SN Computer Science*, Vol.2 No. 6. Pp. 1–20.

Sengupta, S., Basak, S., Saikia, P., Paul, S., Tsalavoutis, V., Atiah, F., Ravi, V., and Peters, A. 2020. A review of deep learning with special emphasis on architectures, applications and recent trends. *Knowledge-Based Systems.* Vol. 194 No. 3. Pp. 1-33.

Singleton, T. W. and Singleton, A. J. 2010. *Fraud Auditing and Forensic Accounting* (4th ed.). John Wiley and Sons, Inc.

Slabber, E., Verster, T., and de Jongh, R. 2023. Some Insights about the Applicability of Logistic Factorisation Machines in Banking. *Risks.* Vol. 11 No. 48. Pp.1-21.

Jurnal Riset Akuntansi dan Bisnis Airlangga Vol. 8. No. 2 (2023) 1521-1542
ISSN 2548-1401 (Print) ISSN 2548-4346 (Online)

Thiprungsri, S., and Vasarhelyi, M. A. 2011. Cluster analysis for anomaly detection in accounting data: An audit approach. *International Journal of Digital Accounting Research.* Vol. 11 No.2. Pp. 69–84.

Tiwari, M., Gepp, A., and Kumar, K. 2020. A review of money laundering literature: the state of research in key areas. *Pacific Accounting Review.* Vol. 32 No. 2. Pp. 271–303.

Vadakara, J. M., and Kumar, D. V. 2019. Aggrandized random forest to detect the credit card frauds. *Advances in Science, Technology and Engineering Systems.* Vol. 4 No. 4. Pp. 121–127.

Vanini, P., Rossi, S., Zvizdic, E., and Domenig, T. 2023. Online payment fraud: from anomaly detection to risk management. *Financial Innovation.* Vol. 9 No. 66. Pp. 1-25.

Vičič, J. and Tošić, A. 2022. Application of Benford's Law on Cryptocurrencies. *Journal of Theoretical and Applied Electronic Commerce Research.* Vol. 17 No.1. Pp. 313–326.

Vosseler, A. 2022. Unsupervised Insurance Fraud Prediction Based on Anomaly Detector Ensembles. *Risks.* Vol. 10 No.132. Pp. 1-20.

Widhiastuti, N. L. P., and Kumalasari, P. D. 2020. Kemampuan Mahasiswa Dalam Mendeteksi Fraud. *Jurnal Riset Akuntansi Dan Bisnis Airlangga.* Vol. 5 No. 1. Pp.762-783.

Wihartiko, F. D., and Wihartika, D. 2019. Fraud Detection of Bus Ticket Sales by Using Spatio Temporal Data Mining. *International Journal of Recent Technology and Engineering (IJRTE).* Vol. 8 No. 2. Pp. 17-21.

Zhai, J., Cao, Y., Yao, Y., Ding, X., and Li, Y. 2017. Computational intelligent hybrid model for detecting disruptive trading activity. *Decision Support Systems.* Vol. 93 No.3. Pp. 26–41.

Zhang, Q., Zhou, Y., and Fang, Y. 2021. A new approach to detecting financial statement fraud using machine learning algorithms. *Journal of Intelligent & Fuzzy Systems.* Vol. 40 No. 2. Pp. 1897–1908.