AI-Driven Fraud Detection in Financial Statements: A Comparative Study of Machine Learning Models

Arman Daryan[1*],Niloufar Kamyabi[2],Shirin Mehraban[3],Tara Khosravi[4]

Master of Accounting,Master of Accounting,Master of Accounting,Master of Accounting

Abstract

The increasing complexity of financial transactions and the growing volume of corporate data have significantly intensified the risk of financial statement fraud, challenging the effectiveness of traditional detection methods. This study examines the role of artificial intelligence in enhancing fraud detection by conducting a comparative analysis of machine learning and deep learning models applied to financial statement data. Drawing on an extensive review of prior literature and empirical findings, the research evaluates the performance of widely used algorithms, including logistic regression, decision trees, support vector machines, random forest, gradient boosting, and deep learning architectures. The results indicate that ensemble and deep learning models consistently outperform conventional statistical and rule-based approaches in terms of accuracy, precision, and overall predictive capability. However, the findings also highlight critical challenges related to data imbalance, model interpretability, and computational complexity, which may limit the practical adoption of advanced models in real-world auditing and regulatory contexts. The study contributes to the existing literature by providing a structured comparison of AI-driven fraud detection techniques and offers practical implications for auditors, regulators, and policymakers seeking to improve the reliability and transparency of financial reporting. Future research directions include the development of explainable and hybrid models, the integration of non-financial and textual data, and the expansion of cross-country datasets to enhance model generalizability and support sustainable capital market development.

KeyWords:

Artificial Intelligence; Financial Statement Fraud; Machine Learning; Deep Learning; Fraud Detection; Data Mining; Ensemble Learning; Financial Reporting

Introduction:

The primary objective of this research is to contribute to financial auditing processes by utilizing advanced machine learning techniques for anomaly and fraud detection in accounting records. Financial reporting and auditing environments are increasingly challenged by factors such as growing data volume, complexity, and the need for timely detection of anomalies. Traditional auditing methods often rely on limited approaches, such as sampling and red flag indicators, and these methods are insufficient for detecting rare, sophisticated, or hidden fraudulent activities. These limitations create significant audit risks, reduce the reliability of financial reporting, and erode stakeholder trust (Mehmet & Ganji, 2021; Ayboğa & Ganii, 2022).

In recent decades, AI has had a profound impact on various industries, including banking. By providing advanced tools for data analysis and automating processes, AI has helped banks achieve significant improvements in productivity, cost reduction, and customer experience enhancement. This article explores the diverse applications of AI in banking.

1. Fraud Detection and Prevention:

The increasing volume of digital transactions, such as bill payments, withdrawals, and check deposits through apps or online accounts, has intensified the need for enhanced fraud detection efforts. Fraud can take various forms, such as using fake identities to obtain loans or making unauthorized transactions. AI plays a critical role in identifying and preventing such illegal activities.

Banks use machine learning models to detect fraud patterns. These models analyze transaction data and identify abnormal behaviors that may indicate fraudulent activity. For instance, if a customer suddenly starts making large transactions in locations far from their usual area, AI models can flag this as suspicious activity. Non-supervised models are also used to identify unknown frauds by detecting unusual patterns in data.

An example of a bank using AI for fraud detection is Danske Bank, a Danish bank, which has increased its fraud detection capabilities by 50% and reduced false positives by 60% by implementing AI-driven fraud detection algorithms.

From a theoretical perspective, this study aims to fill a gap in the accounting and auditing literature. While anomaly detection has been widely studied in areas such as cybersecurity and network breach detection, it appears that such applications in accounting information systems have not been adequately addressed. The aim of this study is to develop a novel approach to anomaly detection by adapting deep learning methodologies to the unique characteristics of financial transactions. Autoencoder neural networks can be particularly effective in detecting subtle deviations in complex data relationships (Ganji, 2024).

From a practical perspective, the research aims to contribute to audit practice by providing a powerful AI-powered tool for auditors, regulators, and organizations to provide full-population testing. Unlike traditional sample-based approaches, the proposed model offers anomaly detection and tracking capabilities across all ledger entries, thereby enhancing the timely and effective functioning of fraud detection and internal control systems (Ganji & Ganji, 2025). Furthermore, the integration of the anomaly detection system into accounting information systems can support proactive risk management, reduce the likelihood of fraudulent activities going undetected, and ultimately strengthen investor confidence and corporate governance (Apak & Ganji, 2025).

2. Chatbots and Virtual Assistants:

Banks use chatbots and virtual assistants to provide 24/7 customer service. These tools are capable of answering frequently asked questions, resolving common issues, and even providing financial advice. By incorporating chatbots into banking apps, banks ensure that they are always available to customers, regardless of the time.

AI-driven chatbots can personalize customer support by analyzing customer behavior, reducing the workload of emails and other communication channels, and suggesting suitable financial products. A prime example is Erica, a virtual assistant from Bank of America, which effectively manages credit card debt reduction and security updates, handling over 50 million customer requests annually in 2019.

3. Customer Experience:

Customers are consistently seeking better and more convenient experiences. For example, ATMs owe their success to the fact that customers can perform basic transactions such as deposits and withdrawals outside of bank hours. This convenience has led to further innovation. Customers can now open bank accounts via their smartphones. Integrating AI into banking services enhances the customer experience by improving convenience and reducing errors.

AI also reduces the time required for customer identification (KYC) and eliminates mistakes. It automates processes for credit applications, such as personal loans, making them faster and less burdensome. AI software also shortens the approval time for loans, improving overall customer experience by ensuring accurate data entry and streamlining customer interactions.

4. Credit Risk Assessment:

A crucial application of AI in banking is credit risk assessment. Banks need to evaluate the creditworthiness of customers to provide loans and other financial services. AI models use historical transaction data, debt levels, and default records to predict the likelihood of a customer defaulting on a loan. This helps banks make better lending decisions and identify high-risk customers.

Traditional methods, such as credit scorecards, involve evaluating a set of customer features like income, debt-to-income ratio, and credit history. These scores help banks make more informed decisions regarding lending.

5. Data Collection and Analysis:

Banks handle millions of transactions daily. Collecting and recording this vast amount of data accurately can be challenging for employees. AI-driven systems help collect and analyze data efficiently, which, in turn, improves the overall user experience. This data is also used for fraud detection and decision-making processes related to credit risk assessment.

6. Regulatory Compliance:

The banking sector is one of the most regulated industries globally. Governments use their regulatory power to ensure that customers do not use banks for illicit financial activities and that banks maintain acceptable risk profiles to prevent widespread defaults. Banks typically have internal teams responsible for compliance with these regulations, but manual processes can be time-consuming and costly.This research focuses on two main objectives: (i) to deepen the academic understanding of anomaly detection in accounting by applying deep autoencoder neural networks to transactional-level data, and (ii) to provide viable insights and technological solutions for transforming audit methodologies in line with the data-intensive demands of modern financial systems (Ganji, 2024).

AI, particularly deep learning and natural language processing, is used in banking to study new compliance requirements and improve decision-making processes. While AI cannot replace compliance analysts, it can accelerate and streamline their operations, helping banks stay updated on constantly changing regulations.

7. Cybersecurity Improvement:

AI can assist banks in enhancing their cybersecurity by identifying potential threats and preventing cyber-attacks. AI algorithms analyze network data and detect cyber threats by identifying abnormal behavior. These systems can take automated actions to mitigate risks, and AI also allows banks to predict potential attacks and take proactive measures to protect their systems.

8. Global Examples of AI in Banking:

AI is transforming banking globally, helping banks improve security, offer better services, and optimize operations. For example:

- ❖ JPMorgan Chase uses AI to develop an early warning system for detecting malware, trojans, and phishing attacks. This system provides alerts before attacks occur, enabling the cybersecurity team to intervene.
- ❖ Capital One uses an AI-driven virtual assistant, Eno, to help customers manage their accounts, detect suspicious activities, and provide personalized support.

Challenges of AI in Banking

Despite its wide-ranging benefits, the use of AI in banking comes with challenges, including:

- ✓ Data Security: As banks handle vast amounts of sensitive customer data, ensuring the security of this data is critical. Banks need to work with technology partners that specialize in both AI and data security.
- ✓ Lack of Quality Data: AI models require structured, high-quality data for training. If the data is poor, the AI system may produce inaccurate results. Banks must update their data policies and ensure compliance with privacy regulations while gathering and using data.
- ✓ Regulatory Complexity: The ever-changing regulatory landscape can complicate the adoption of new technologies. Banks must ensure their AI systems comply with all applicable regulations and stay updated as these regulations evolve.

- ✓ High Costs and Investments: Implementing AI systems requires significant investment in infrastructure, data management, and skilled human resources. Banks need to evaluate the costs and benefits of AI adoption carefully.
- ✓ Cultural Resistance: AI adoption may face resistance from employees due to concerns about job displacement. Banks need to foster a culture of innovation and provide training programs to help employees adapt to new technologies.

The future of AI in banking looks promising. With continued advancements in AI technology, new innovations will emerge, bringing banks to new levels of efficiency and service delivery. However, the impact on the workforce should also be considered, as some jobs may change or become obsolete.

Banks must develop strategies for AI integration, invest in talent, and create robust policies to ensure the ethical use of AI. Furthermore, AI will play a critical role in addressing evolving customer needs and improving decision-making capabilities.

AI has become a vital tool in banking, enhancing operational efficiency, improving security, and delivering personalized services. From fraud detection and credit risk evaluation to customer service and regulatory compliance, AI is transforming the banking industry. However, challenges such as data security, regulatory compliance, and cost management must be addressed for successful AI implementation. The future of AI in banking promises even greater innovations, making it essential for banks to stay ahead by embracing and adapting to this evolving technology.

2.Literature Review:

The overall objective of this article is to present a comprehensive and practical perspective on the importance and applications of algorithms and data structures, the challenges involved in learning them, and effective strategies for improving these skills. This article provides an in-depth and systematic examination of the role of algorithms and data structures in programming and software development. It begins by highlighting their critical role in improving efficiency and scalability of applications, enhancing problem-solving and coding skills, and emphasizing their importance in mastering advanced computer science concepts such as machine learning and data science. The article also discusses the role of algorithms in everyday life and their impact on various aspects of human activity.

Subsequently, the challenges associated with learning and implementing algorithms and data structures are examined. Issues such as poor design, inappropriate selection of data structures, and implementation problems are analyzed, and corresponding solutions are proposed. Common misconceptions regarding algorithmic complexity and the importance of consistency in learning are also addressed. Finally, one of the key sections of this article focuses on algorithmic thinking and its significance in solving programming problems. Algorithmic thinking represents a mental shift that enables programmers to approach problems systematically and logically, rather than relying on shortcuts or unwritten rules, and to design optimal solutions.

Machine learning enables computers to learn from data and improve their performance without explicit programming. It has become a fundamental component of modern data-driven systems and is generally classified into three main categories: supervised learning, unsupervised learning, and reinforcement learning. Each of these paradigms addresses different types of problems and data structures and plays a distinct role in applied artificial intelligence.

Supervised learning is the most widely used type of machine learning and relies on labeled datasets in which both input features and correct output values are known. The primary objective of supervised learning is to learn a mapping function that accurately transforms new, unseen inputs into correct outputs. This learning paradigm is commonly divided into regression and classification tasks. Regression focuses on predicting continuous numerical values, such as estimating house prices based on structural and locational attributes.

Classification, on the other hand, aims to assign observations to predefined categories, such as distinguishing spam emails from non-spam messages or diagnosing diseases based on clinical indicators.

In contrast, unsupervised learning operates on unlabeled data, where no explicit output labels are provided. The main goal of this approach is to uncover hidden patterns, structures, or relationships within the data. Two prominent unsupervised learning techniques are clustering and dimensionality reduction. Clustering algorithms group data points into clusters based on similarity measures, which is particularly useful in customer segmentation and behavioral analysis. Dimensionality reduction techniques aim to reduce the number of features while preserving essential information, thereby improving computational efficiency and facilitating data visualization and compression.Financial statement manipulation scandals, such as the Luckin Coffee case in China that falsified revenues exceeding US$300 million (J. Li et al., 2024), or the PT Asabri case in Indonesia that caused state losses of up to Rp22 trillion, show that financial statement fraud is not only a local phenomenon, but a global threat capable of shaking capital markets and eroding public trust (Ritonga & Budhiawan, 2024). Both cases illustrate how financial statement fraud can result in systemic losses, foster public distrust, and weaken the integrity of capital markets and internal control systems. Financial statements are a fundamental component in communicating a company's financial information to stakeholders (Wahyu Fikri Darmawan & Umaimah Umaimah, 2025). These documents reflect the financial condition and operational performance of a business entity in a certain accounting period, including the income statement, balance sheet, statement of changes in equity, and cash flow statement (Agustan & Sari, 2022; Daeli et al., 2024). The information presented is the basis for decision making for investors, creditors, regulators, and internal management. However, increasing market pressure and weak internal controls can lead to manipulative practices in the preparation of financial statements. In an increasingly competitive market landscape, financial reports have strategic implications. Reliable financial information enables management to make data-driven decisions and improve the efficiency of resource management (Barman, 2023; Indawatika, 2017). However, the quality of these reports can be compromised by fraud, which not only damages the credibility of the company but also causes significant economic losses and market instability (Iskandar et al., 2022; Kootanaee et al., 2021). The prevalence of fraud in financial reporting has become an increasingly worrying issue. According to a report from (ACFE, 2022), organisational fraud is classified into three main categories: asset misappropriation, financial statement fraud, and corruption. Asset misappropriation is the most common form of fraud, accounting for approximately 86% of all reported cases. It involves the theft or misuse of organisational resources, such as cash embezzlement, inventory theft, use of company assets for personal gain, and cost manipulation. Although highly prevalent, this category has a relatively low average loss of USD 100,000 per incident. Corruption, on the other hand, occurs in 50% of fraud cases and involves the abuse of authority or position for personal gain. This can include bribery, conflict of interest, extortion, or collusion - either between individuals or between the organisation and external parties. The average loss caused by corruption is reported to be USD 150,000 per case. However, financial statement fraud is the category with the highest financial impact. While it only accounts for around 9% of all fraud cases, the average loss is USD 593,000 per incident. These practices include actions such as inflating revenue, understating liabilities, delaying expense recognition, or concealing financial information to mislead investors, creditors, or regulators. Key motivations include meeting earnings targets, inflating share prices, gaining managerial incentives, or avoiding legal and tax penalties. These data suggest that although financial statement fraud is less common than other types, its financial and systemic impacts are much more severe and damaging (Alfian & Triani, 2019; Hari et al., 2025). These findings highlight that, despite lower incidence rates, financial statement fraud poses more systemic risks and significantly undermines investor and creditor confidence. Therefore, the authors argue that early detection of this type of fraud should be a top priority in the financial supervision system. Financial statement fraud refers to manipulative actions taken to present a financial picture that does not reflect the actual condition of the company (Prayoga & Sudaryati, 2020; Supriadi & Aryati, 2022). The objectives of main such practices usually include portraying better company performance, increasing stock prices, avoiding taxes, or obtaining funding from external parties such as investors or creditors (Ashtiani & Raahemi, 2022). One of the most influential conceptual frameworks in explaining fraudulent behaviour is the Fraud Triangle Theory, introduced by (Cressey, 2018). A criminologist, Cressey developed this theory based on interviews with prisoners convicted of fraud in the United States. He found that individuals who commit fraud are typically driven by three core elements: pressure to meet financial demands or organisational targets, opportunity resulting from weak oversight or internal control systems, and rationalisation, where perpetrators morally justify their actions as "reasonable" or "temporary" (Prasetyo & Dewayanto, 2024; WS Albrecht, 2019). Although developed more than half a century ago, this theory is still very relevant in the context of modern financial fraud detection, including in the era of artificial intelligence. The three components of the Fraud Triangle represent patterns of behaviour that can be traced through digital footprints and data trends, such as abnormal financial ratios, reporting frequency, or narrative patterns in annual reports. This is where machine learning and data mining become crucial: these models can detect patterns that reflect financial pressures (e.g., unusual profitability ratios), opportunities (e.g., irregularities in internal auditing), and even forms of rationalisation in MD&A or management reports. Thus, AI-based approaches are not only able to identify statistical outliers but also conceptually model the behaviour of fraudsters. This highlights the potential for integrating theoretical frameworks with intelligent technology to improve the effectiveness of financial fraud detection systems. Conventional fraud detection methods, such as manual audits, have several limitations. Manual processes are often expensive, as they require considerable human resources and time (West & Bhattacharya, 2016). In addition, these methods tend to be less accurate due to their reliance on the subjective judgement of the individuals involved. As a result, human error is a major factor that can delay fraud identification or increase the risk of poor decision-making (Prasetyo & Dewayanto, 2024). The continued use of traditional methods, especially in many developing countries, reflects gaps in technology adoption. It also indicates resistance to the transition to a technology-based audit system. Amidst these limitations, technological advances offer more effective and efficient solutions for fraud detection (Ashtiani & Raahemi, 2022). Artificial intelligence (AI), particularly machine learning technology, has emerged as one of the most successful approaches to fraud detection, while data mining, as a core component of AI, plays an important role in identifying patterns and detecting fraudulent behaviour quickly (Prasetyo & Dewayanto, 2024). These techniques enable the analysis of millions of financial statements to uncover suspicious trends and flag potentially fraudulent disclosures. With the ability to operate in real time, such technologies not only reduce operational costs but also provide faster and more accurate responses, thereby strengthening the company's internal control system (Massi et al., 2020). The authors argue that utilising these technologies can be a logical alternative for companies looking to improve the effectiveness of their supervisory systems, especially in the face of the increasing complexity of financial data. A growing literature shows that intelligent approaches can be applied in various forms: from supervised learning for binary classification to unsupervised learning for anomaly detection, especially in cases where labelled data is very limited (Ashtiani & Raahemi, 2022). Some recent literature reviews have focused more narrowly on specific areas within the financial sector, such as fraud detection in credit card transactions, insurance, and fraud prediction in banking credit administration (Al-Hashedi & Magalingam, 2021), motor vehicle insurance fraud detection (Schrijver et al., 2024), and a comparative review between machine learning-based fraud detection and traditional detection methods (Gupta & Mehta, 2024). These studies have a different focus to the current research. The Systematic Literature Review (SLR) conducted in this study is more specifically targeted at financial statement fraud detection and offers a different analytical approach. There are several empirical gaps in the existing literature, including a lack of studies that systematically integrate financial data, non-financial data, narrative textual data, social media data, and accounting-based detection models (such as FScore and M-Score) to build comprehensive fraud detection systems. In addition, there are some limited studies that combine advanced algorithms with traditional accounting models. From a methodological perspective, existing gaps include a dearth of truly comprehensive and comparative literature reviews, especially in terms of the variety of algorithms, data sources and quality assessment protocols used. The few existing SLRs still lack methodological diversity and are often not explicit enough in evaluating the effectiveness of the reviewed techniques. This research aims to address these gaps by presenting a Systematic Literature

Review (SLR) that specifically evaluates the application of artificial intelligence (AI) and data mining (DM) technologies in detecting corporate financial statement fraud. The review covers the period from 2014 to 2024 and adopts the Kitchenham methodology, which provides a systematic framework for literature identification, selection, quality assessment and synthesis (Kitchenham & Brereton, 2007). Unlike narrative reviews, this SLR utilised clear inclusion and exclusion criteria, a structured search protocol, and a rigorous study quality evaluation approach. Unlike previous reviews that tend to be limited to one dimension, this study offers several important contributions. First, it presents a comprehensive SLR that systematically evaluates AI/DM methods and data types (structured and unstructured) used in financial statement fraud detection. Second, this study integrates geographical perspectives and temporal trends, thereby enhancing the understanding of the context and evolution of technology adoption in financial reporting across different countries. Third, this research explores the integration of traditional accounting-based models (such as F-Score and M-Score) with modern machine learning approaches, a dimension that has not been a major focus in previous SLRs. Research on this topic is still relatively scarce in terms of systematic investigations, especially those that explicitly explore the relationship between dataset types, AI/DM methods, and geographical dimensions in the context of corporate financial reporting. Therefore, this study is expected to provide a comprehensive synthesis and serve as a theoretical and practical foundation for the development of more adaptive and intelligent financial audit and supervision systems. This study aims to systematically review the literature on financial statement fraud detection using artificial intelligence (AI) and data mining (DM) approaches. As an SLR study, this research does not propose formal hypotheses, but instead focuses on mapping and evaluating previous empirical findings. Therefore, this study is organised around the following research questions, which also reflect analytical expectations regarding the development of this field: 1) What techniques have been used in the literature to detect financial statement fraud? 2) What data sets have been used in the literature to detect financial statement fraud? 3) How effective are the techniques used in the literature to detect financial statement fraud? 4) How effective are the datasets used in the literature to detect financial statement fraud? The practical implications of this research include strengthening AI-based internal audit systems that are more adaptive to diverse data types, as well as providing strategic insights for regulators and policymakers in designing financial oversight frameworks that are more responsive to fraud risks.

Reinforcement learning represents a different paradigm in which an autonomous agent learns optimal actions through continuous interaction with an environment. The agent receives feedback in the form of rewards or penalties and iteratively adjusts its strategy to maximize cumulative rewards. This learning framework is especially effective for sequential decision-making problems and has been widely applied in areas such as robotics, game playing, autonomous systems, and recommendation engines.

Scikit-learn (sklearn) is a powerful and open-source machine learning library in Python that is designed to simplify the implementation of machine learning models. Built on top of scientific computing libraries such as NumPy and SciPy, scikit-learn provides a consistent and user-friendly interface for a wide range of supervised and unsupervised learning algorithms. It supports various data formats, including NumPy arrays and Pandas DataFrames, making it particularly suitable for applied data science and academic research.

The machine learning workflow in scikit-learn typically begins with data loading and exploration. The library offers built-in datasets that eliminate the need for external data acquisition. Once the data are loaded, they can be converted into structured DataFrames for exploratory analysis using descriptive statistics and summary information. This exploration phase helps researchers understand the structure, scale, and characteristics of the dataset before proceeding to modeling.

Data preprocessing is a critical stage in machine learning projects and includes tasks such as handling missing values, removing duplicate records, splitting datasets into training and testing subsets, and feature scaling. Standardization techniques such as StandardScaler transform features to have zero mean and unit variance, ensuring that variables with different scales contribute equally to model training. Alternatively, MinMaxScaler rescales features into a fixed range, typically between zero and one, which can accelerate convergence and improve model performance for certain algorithms.

Following preprocessing, appropriate machine learning models are selected based on the nature of the task. Scikit-learn provides a broad collection of regression, classification, and clustering models. Regression models such as linear regression, Ridge, Lasso, and ElasticNet are widely used for predicting continuous outcomes, while decision trees, random forests, support vector machines, and k-nearest neighbors are employed across both regression and classification problems. Model performance is evaluated using suitable metrics such as mean squared error and R-squared for regression tasks, and accuracy, precision, recall, and confusion matrices for classification tasks.

Hyperparameter optimization plays a vital role in improving model performance. GridSearchCV is a commonly used technique that systematically evaluates different combinations of hyperparameters using cross-validation. By selecting the parameter set that yields the best validation performance, researchers can significantly enhance model accuracy and generalizability. Empirical results demonstrate that optimized models often outperform baseline configurations, highlighting the importance of proper parameter tuning.

For classification problems, evaluation metrics provide complementary perspectives on model behavior. Accuracy reflects the overall correctness of predictions, while precision measures the reliability of positive predictions, and recall assesses the model's ability to identify all relevant positive cases. These metrics are particularly important in imbalanced datasets, where accuracy alone may be misleading. Confusion matrices further enable a detailed analysis of classification errors by visualizing true positives, false positives, true negatives, and false negatives.

Unsupervised learning techniques such as clustering are also supported extensively in scikit-learn. Algorithms like K-means, DBSCAN, and agglomerative clustering are used to group data points without predefined labels. Their performance is commonly evaluated using internal validation metrics such as the Silhouette Score, Davies–Bouldin Index, and Calinski–Harabasz Index. These metrics assess cluster cohesion and separation, providing quantitative insights into clustering quality.

Dimensionality reduction methods such as Principal Component Analysis (PCA) are often applied prior to visualization. By projecting high-dimensional data into lower-dimensional spaces, PCA enables intuitive graphical representations of clustering results. Visual inspection, combined with quantitative evaluation metrics, provides a comprehensive understanding of model behavior and data structure.this study presents a comprehensive overview of machine learning workflows using the scikit-learn library, covering supervised, unsupervised, and reinforcement learning concepts. From data preprocessing and model selection to evaluation and optimization, each stage plays a crucial role in achieving robust and reliable results. Owing to its simplicity, flexibility, and extensive documentation, scikit-learn serves as an indispensable tool for researchers, data scientists, and practitioners aiming to develop effective machine learning solutions in both academic and real-world contexts.

3.Methodology:

Programming languages are fundamental tools for software development. They provide specific rules and commands that allow programmers to express ideas and logic in a form that computers can execute. Familiarity with programming languages enables developers to write correct code, handle compilation and runtime errors, and produce reliable software.

Data structures and algorithms are core concepts in computer science and play a crucial role in programming and software development. Data structures are designed to organize and store data in a way that allows operations to be performed efficiently. Algorithms, on the other hand, are step-by-step procedures or formulas for solving specific problems. Together, these concepts contribute to improved performance, optimal resource utilization, scalability, systematic problem-solving, and success in technical job interviews.

Reasons for Learning Algorithms and Data Structures:

The key reasons for learning Data Structures and Algorithms (DSA) are as follows:

Improving Problem-Solving Skills:

Learning DSA enhances problem-solving skills, including problem definition, analysis, selection of solutions, implementation, evaluation, and revision. It enables the development of a systematic thinking approach that can be applied across a wide range of computer science problems. When faced with complex problems, logical thinking and precise analysis are essential for identifying appropriate solutions.

Improving Coding Efficiency:

Understanding how to use appropriate data structures and algorithms allows developers to write efficient, high-performance code more quickly. For example, different sorting algorithms can significantly reduce execution time and optimize data organization.

Facilitating Learning of Advanced Computer Science Concepts:

Many advanced computer science topics are built upon the principles of data structures and algorithms. A strong foundation in DSA enables easier mastery of complex fields such as machine learning, data science, and system design.

Career Advancement:

Many employers place high value on DSA skills. A strong understanding of these concepts demonstrates the ability to design intelligent and effective solutions to complex problems.

Enhancing Flexibility and Scalability:

Knowledge of data structures and algorithms enables the design of flexible and scalable systems that can adapt to future changes and new requirements.

Improving System Reliability and Stability:

Using appropriate data structures and algorithms improves system reliability by preventing errors caused by inefficient data management and ensuring consistent performance under various conditions.

Role of Algorithms in Everyday Life:

Beyond their importance in computing, algorithms are widely present in daily life. Algorithmic logic can be observed in natural patterns and everyday activities such as sorting, planning, counting, and optimization. This logic often helps us perform tasks more effectively.

In the book *"The Formula: How Algorithms Solve All Our Problems"* by Luke Dormehl, the role and influence of algorithms in modern life is extensively discussed. Dormehl defines a "formula" as a set of actions through which resources are used to create value. He argues that data-driven processes shape our identities and that even personal relationships are increasingly influenced by algorithms. He raises concerns about automation's impact on the rule of law and the future of art and human creativity. According to Dormehl, algorithms shape human identity and decision-making, creating challenges related to individual autonomy and social control. He encourages reflection on what it means to be human in the age of algorithms and highlights the emerging crisis of individual identity.

Major Challenges in Learning Algorithms and Data Structures:

One of the most significant causes of weak implementation of data structures and algorithms is poor design, which can result in high time and space complexity or failure to consider appropriate input data. Selecting unsuitable data structures for specific operations can significantly degrade system performance. Implementation issues such as programming errors and improper memory management further contribute to ineffective execution.

Other contributing factors include ignoring real-world constraints, insufficient testing with realistic data, and inadequate documentation, which makes algorithms difficult to understand and maintain. Theoretical misunderstandings, incorrect problem modeling, and poor comprehension of foundational principles also play a major role in failed implementations.

By understanding these issues and addressing the human factors involved, more efficient and effective algorithmic solutions can be developed.

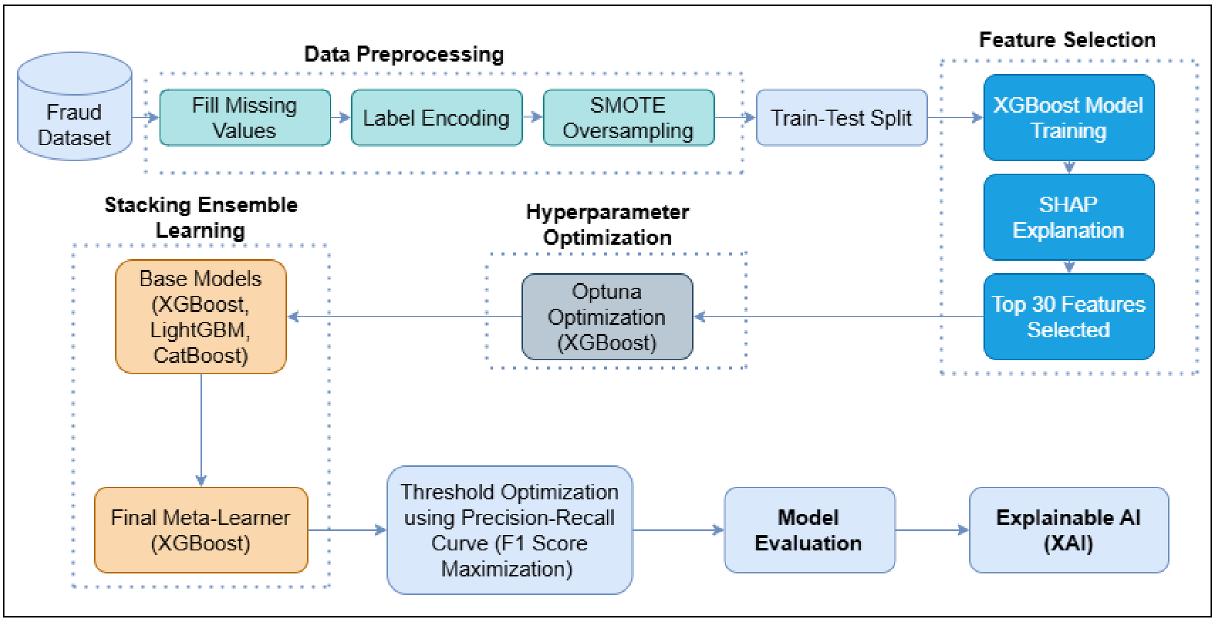

Figure 1:The Proposed Fraud Detection Framework.

This study introduces a robust framework for detecting financial fraud using XAI techniques. The project focuses on two primary objectives: developing a highly accurate model to identify fraudulent transactions and ensuring that its decisions are interpretable. The overall framework consists of several phases:

- ✓ Data collection
- ✓ Data preparation
- ✓ Feature selection
- ✓ Hyperparameter optimization
- ✓ Building a stacking ensemble model
- ✓ Evaluation
- ✓ XAI tools

Algorithmic Thinking:

Algorithmic thinking is a relatively new concept in this domain and is often considered a core component of computational thinking. While learning common algorithms individually is useful, developing algorithmic thinking as a habit is far more beneficial. Training the mind to follow algorithmic logic makes writing algorithms more intuitive.

Algorithmic thinking represents a mental shift in how humans approach problem-solving. It emphasizes systematic, logical thinking similar to how computers execute instructions. However, this approach can be challenging because humans naturally rely on assumptions and shortcuts in everyday problem-solving.

For example, when sorting ten numbers, humans can quickly determine the correct order intuitively. Computers, however, require explicit instructions detailing where to start, how to compare elements, when to stop, and how to proceed.

Algorithmic thinking is essential for two core programming skills:

1. Designing an efficient step-by-step solution
2. Translating that solution into correct code

In technical interviews, employers often value the candidate's problem-solving approach more than the final implementation. Demonstrating strong algorithmic thinking can therefore significantly improve employability.

How to Think Algorithmically:

Algorithmic thinking, like any skill, can be learned through practice. It involves solving problems systematically through logical and iterative processes. The key steps include:

1. Clearly defining the problem
2. Breaking it down into smaller components
3. Designing solutions for each component
4. Implementing the solution
5. Reviewing and optimizing the result

Algorithms are simply methods for accomplishing tasks, whether searching for a word in a dictionary, sorting a list, generating Fibonacci numbers, cooking a meal, or doing laundry.

Practicing Algorithmic Thinking Through Problem Solving:

To solve problems effectively, it is essential to follow structured steps. This practice saves time, helps identify recurring patterns, and enables efficient solutions even for complex problems.

The main stages include:

- ✓ Understanding the problem statement
- ✓ Selecting appropriate strategies
- ✓ Designing a solution using pseudocode
- ✓ Implementing the solution
- ✓ Testing, analyzing complexity, and optimizing

Practicing on platforms such as LeetCode, HackerRank, and participating in online courses (e.g., Udemy, Coursera, Codecademy) significantly enhances algorithmic thinking skills. programming and software development. As foundational pillars of computer science, they play a vital role in improving efficiency, optimizing resource utilization, and ensuring scalability. They significantly enhance problem-solving abilities and are essential for mastering advanced topics such as machine learning and data science.

Despite the challenges involved in learning DSA, including poor design choices and misconceptions about complexity, adopting structured learning strategies and developing algorithmic thinking can overcome these obstacles. Algorithmic thinking enables programmers to approach problems logically and systematically, leading to more efficient and robust solutions.

Ultimately, mastery of algorithms and data structures empowers programmers to advance their skills, improve career prospects, and design high-quality software systems.

4.Results:

When evaluating the model, we aimed to achieve the highest accuracy while also ensuring its decisions can be clearly understood and trusted.

4.1Model Evaluation Metrics:

Using the top 30 SHAP-selected features, we trained our stacking ensemble. Training took approximately 278 seconds, reflecting the model's efficiency despite the large dataset. The performance metrics on the test set are summarized in the classification report:The identification and prevention of financial fraud represent one of the most critical challenges facing organizations and financial institutions, with significant implications for the health and stability of economic systems. The present study investigates the application of advanced machine learning methods in detecting financial fraud, aiming to evaluate and compare the performance of several machine learning models in this domain. In this research, decision tree, artificial neural networks, support vector machines, gradient boosting, and random forest methods were applied to financial data from companies listed on the Tehran Stock Exchange. After conducting the necessary preprocessing steps, the models were trained. The results indicate that the gradient boosting model, with an accuracy of 95%, a sensitivity of 92%, and an F1-score of 0.94, delivered the best performance in detecting financial fraud. Additionally, artificial neural networks and decision tree models demonstrated acceptable performance with accuracies of 90% and 80%, respectively. Unlike previous studies that typically focused on a limited set of algorithms, this study adopts a comprehensive and multi-model approach, enabling a precise and practical comparison of various algorithms' performance. The use of real financial data and rigorous preprocessing enhanced the validity and generalizability of the findings. The results of this study suggest that employing advanced machine learning methods—particularly the gradient boosting algorithm—can significantly enhance the accuracy and sensitivity of financial fraud detection compared to traditional methods, and can serve as an effective and practical solution in fraud detection systems. This multi-model and application-oriented approach highlights the study's main innovation in improving the precision and efficiency of financial fraud detection using state-of-the-art machine learning technologies, and constitutes a meaningful step toward strengthening security and trust in financial systems.

Table 1: Data Mining Products Evaluated

Company	Product	Version
Integral Solutions, Ltd. (ISL) [5]	Clementine	4.0
Thinking Machines (TMC) [6]	Darwin	3.0.1
SAS Institute [7]	Enterprise Miner (EM)	Beta
IBM [8]	Intelligent Miner for Data (IM)	2.0
Unica Technologies Inc. [9]	Pattern Recognition Workbench (PRW)	2.5

Table 2: Software and hardware Supported

Product	Server	Client	Oracle Connect
Clementine	Solaris 2.X	X Windows	Server side ODBC
Darwin	Solaris 2.X	Windows NT	Server side ODBC
Enterprise Miner	Solaris 2.X[1]	Windows NT	SAS Connect>for Oracle
Intelligent Miner	IBM AIX	Windows NT	IBM Data Joiner>
PRW	Data only	Windows NT	Client side ODBC

1 most testing performed on a standalone Windows NT version.

Table 3: Algorithms Implemented

Algorithm	IBM	ISL	SAS	TMC	Unica
Decision Trees	✓	✓	✓	✓	
Neural Networks	✓	✓	✓	✓	✓
Regression	[1]	✓	✓		✓
Radial Basis Functions	✓[2]				✓
Nearest Neighbor			✓		
Nearest Mean Kohonen Self-Organizing Maps		✓	✓		
Clustering	✓	✓			✓
Association Rules	✓	✓			

1 accessed in data analysis only 2 estimation only (not for classification)

Table 4: Options for Decision Trees

Algorithm	IBM	ISL	SAS	TMC
Handles Real-Valued Data	✓	✓	✓	✓
Costs for Misclassification		✓	✓	✓
Assign Priors to Classes			✓	✓
Costs for Fields	✓			
Multiple Splits			✓	
Advanced Pruning Options		✓		
Graphical Display of Trees	✓	✓	✓	

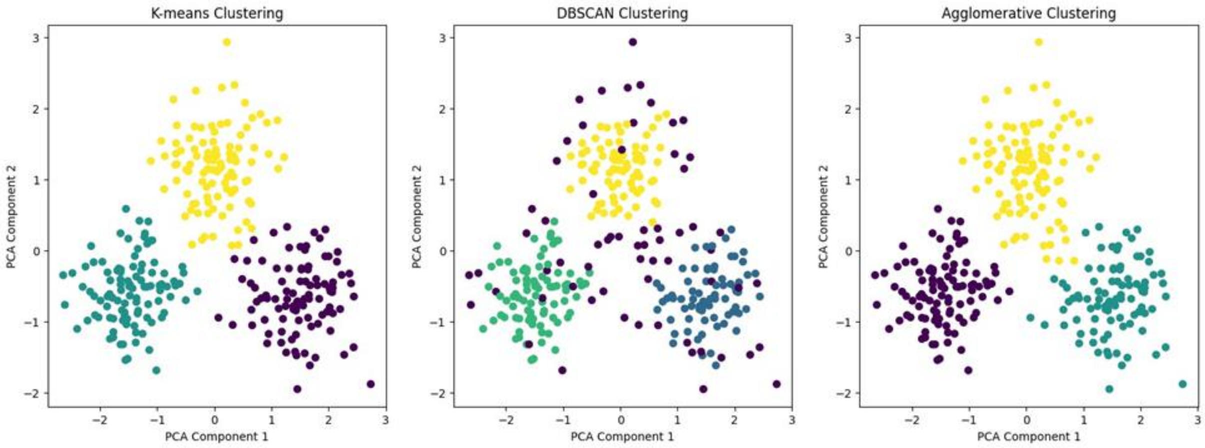

Figure 2:Precision-Recall (PR) Curve of Stacking Classifier.

The PR curve (Figure 2) highlights the model's ability to detect fraud in the context of class imbalance. The curve shows that the model maintains very high precision even as recall increases, meaning it catches the majority of fraud cases with very few false alarms. This is especially important because fraudulent transactions are rare relative to normal ones; a high precision ensures that investigators' attention is not wasted on too many false fraud alerts.The primary aim of this research is to present a thorough and all-encompassing examination of artificial intelligence (AI) methodologies employed in the detection of financial fraud. The present study employs a systematic literature review (SLR) that was conducted utilizing the PRISMA approach. A comprehensive search was undertaken on reputable academic databases including ScienceDirect, Scopus, Springer, and Emerald, yielding a total of 24 papers published throughout the timeframe of 2014 to 2023. These articles will, thereafter, undergo further analysis. The findings of this study demonstrate that the implementation of artificial intelligence (AI) techniques for detecting financial fraud yields favorable outcomes. Specifically, the AI approach proves to be effective in enhancing the precision and efficiency of fraud pattern identification, thereby making a substantial contribution in this domain. In contrast, the prevailing methodology employed in the realm of financial fraud detection is frequently centered around machine learning. Furthermore, a majority of the research encompassed a diverse range of industries, with particular emphasis on the financial industry as the primary domain for the implementation of artificial intelligence (AI) in the detection of financial fraud.

5.Discussion and Conclusion:

Purpose - This study aims to compare machine learning models, datasets and splitting training-testing using data mining methods to detect financial statement fraud. Design/methodology/approach - This study uses a quantitative approach from secondary data on the financial reports of companies listed on the Indonesia Stock Exchange in the last ten years, from 2010 to 2019. Research variables use financial and non-financial variables. Indicators of financial statement fraud are determined based on notes or sanctions from regulators and financial statement restatements with special supervision. Findings - The findings show that the Extremely Randomized Trees (ERT) model performs better than other machine learning models. The best original-sampling dataset compared to other dataset treatments. Training testing splitting 80:10 is the best compared to other training-testing splitting treatments. So the ERT model with an original-sampling dataset and 80:10 training-testing splitting are the most appropriate for detecting future financial statement fraud. Practical implications - This study can be used by regulators, investors, stakeholders and financial crime experts to add insight into better methods of detecting financial statement fraud. Originality/value - This study proposes a machine learning model that has not been discussed in previous studies and performs comparisons to obtain the best financial statement fraud detection results. Practitioners and academics can use findings for further research development.This study provides a comprehensive examination of artificial intelligence–driven approaches for detecting fraud in financial statements by comparatively evaluating a wide range of machine learning and deep learning models. Drawing on an extensive body of prior literature and empirical evidence, the findings confirm that AI-based models significantly outperform traditional rule-based and ratio-based methods in identifying fraudulent financial reporting. In particular, ensemble learning techniques and deep learning architectures demonstrate superior predictive accuracy and robustness when handling complex, high-dimensional financial data.

The comparative analysis reveals that tree-based ensemble models such as Random Forest and XGBoost consistently achieve high classification performance due to their ability to capture nonlinear relationships and interactions among financial indicators. Deep learning models, including neural networks and hybrid architectures, further enhance detection capability by automatically learning latent patterns from large-scale datasets. However, these models often require substantial computational resources and large, high-quality datasets, which may limit their practical applicability in smaller firms or data-constrained environments.

Despite their strong predictive power, AI-driven fraud detection models are not without limitations. Issues related to data imbalance, model interpretability, and generalizability across industries and countries remain significant challenges. The lack of transparency in complex models, particularly deep learning approaches, may hinder their adoption by auditors and regulators who require explainable and legally defensible decision-making tools. Therefore, integrating explainable AI techniques and domain knowledge into fraud detection systems is essential to enhance trust and usability.

Overall, this study underscores the transformative role of artificial intelligence in improving the effectiveness and efficiency of financial statement fraud detection. By systematically comparing existing models, the research contributes to both academic literature and practical applications, offering valuable insights for researchers, auditors, regulators, and policymakers. Future research should focus on developing hybrid and explainable models, expanding cross-country datasets, and incorporating non-financial and textual data to further strengthen fraud detection frameworks and support sustainable capital market development.

Referense :

Alfian, F., & Triani, N. N. A. (2019). Fraudulent financial reporting detection using Beneish M-Score model in public companies in 2012–2016. *Asia Pacific Fraud Journal, 4*(1), 27–42. https://doi.org/10.21532/apfj.001.19.04.01.03

Ali, A. A., 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, 13*(4), 2272. https://doi.org/10.3390/app13042272

Altman, E. I. (1974). Corporate financial distress. *The Journal of Finance, 29*(1), 312. https://doi.org/10.1111/j.1540-6261.1974.tb00057.x

Ashtiani, M. N., & Raahemi, B. (2022). Intelligent fraud detection in financial statements using machine learning and data mining: A systematic literature review. *IEEE Access, 10*, 72504–72525. https://doi.org/10.1109/ACCESS.2021.3096799

Barman, R. D. (2023). Financial statements: A tool to evaluate business performance. *Business Management and Economics Engineering, 21*, 819–835.

Beneish, M. D. (1999). The detection of earnings manipulation. *Financial Analysts Journal, 55*(5), 24–36. https://doi.org/10.2469/faj.v55.n5.2296

Chen, S. (2016). Detection of fraudulent financial statements using the hybrid data mining approach. *SpringerPlus, 5*(1), 1–16. https://doi.org/10.1186/s40064-016-1707-6

Chen, S., Goo, Y.-J. J., & Shen, Z.-D. (2014). A hybrid approach of stepwise regression, logistic regression, support vector machine, and decision tree for forecasting fraudulent financial statements. *The Scientific World Journal, 2014*, 1–11. https://doi.org/10.1155/2014/968712

Craja, P., Kim, A., & Lessmann, S. (2020). Deep learning for detecting financial statement fraud. *Decision Support Systems, 139*, 113421. https://doi.org/10.1016/j.dss.2020.113421

Mehmet, H., & Ganji, F. (2021). Detecting fraud in insurance companies and solutions to fight it using coverage data in the COVID-19 pandemic. PalArch's Journal of Archaeology of Egypt / Egyptology, 18(15), 392–407. https://scholar.google.com/citations?view_op=view_citation&hl=tr&user=_RyCeTEAAAAJ&citation_for_view=_RyCeTEAAAAJ:Y0pCki6q_DkC

Ayboğa, M. H., & Ganii, F. (2022). The Covid 19 Crisis and The Future of Bitcoin in E-Commerce. Journal of Organizational Behavior Research, 7(2), 203-213. https://doi.org/10.51847/hta7Jg55of

GANJI, F. (2024). LEVERAGING BIO-INSPIRED ALGORITHMS TO ENHANCE EFFICIENCY IN COVID-19 VACCINE DISTRIBUTION. TMP Universal Journal of Research and Review Archives, 3(4). https://doi.org/10.69557/ujrra.v3i4.103.

Ganji, F. (2024). Incorporating emotional intelligence in shark algorithms: Boosting trading success with affective AI. TMP Universal Journal of Research and Review Archives, 3(4). https://doi.org/10.69557/ujrra.v3i4.105

Ganji, F., & Ganji, F. (2025). The Role of Sports Sponsorships in Shaping Financial Strategy and Accounting Practices. International Journal of Business Management and Entrepreneurship, 4(2), 86–99. Retrieved from https://www.mbajournal.ir/index.php/IJBME/article/view/79

Ganji, F., & Ganji, F. (2025). The Impact of Financial Reporting Standards on Sports Franchise Valuation. International Journal of Business Management and Entrepreneurship, 4(1), 46–60. Retrieved from https://mbajournal.ir/index.php/IJBME/article/view/64

GANJI, F. (2024). ASSESSING ELECTRIC VEHICLE VIABILITY: A COMPARATIVE ANALYSIS OF URBAN VERSUS LONG-DISTANCE USE WITH FINANCIAL AND AUDITING INSIGHTS. TMP Universal Journal of Research and Review Archives, 3(4). https://doi.org/10.69557/ujrra.v3i4.107

GANJI, F. (2025). Exploring the Integration of Quantum Computing and Shark Algorithms in Stock Market Trading: Implications for Accounting, Finance and Auditing. International Journal of Business Management and Entrepreneurship, 4(2), 40–56. Retrieved from https://mbajournal.ir/index.php/IJBME/article/view/76

ACFE. (2022). *Occupational fraud 2022: A report to the nations* (pp. 1–96). Association of Certified Fraud Examiners.

Agustan, T. J., & Sari, U. P. (2022). Analisis laporan keuangan guna alat ukur kinerja keuangan pada PT Global Imoo Telekomunikasi. *Worksheet: Jurnal Akuntansi, 1*(2), 94–103. https://doi.org/10.46576/wjs.v1i2.2116

Al-Hashedi, K. G., & Magalingam, P. (2021). Financial fraud detection applying data mining techniques: A comprehensive review from 2009 to 2019. *Computer Science Review, 40*, 100402. https://doi.org/10.1016/j.cosrev.2021.100402

GANJI, F. (2025). Biomimetic Shark Algorithms: Leveraging Natural Predator Strategies for Superior Market Performance and Advanced Accounting Techniques. International Journal of Business Management and Entrepreneurship, 4(2), 57–70. Retrieved from https://mbajournal.ir/index.php/IJBME/article/view/77

APAK, . R., & GANJI, F. . (2025). Using Decision Tree Algorithms and Artificial Intelligence to Increase Audit Quality: A Data-Based Approach to Predicting Financial Risks. International Journal of Business Management and Entrepreneurship, 4(1), 87–99. Retrieved from https://mbajournal.ir/index.php/IJBME/article/view/65

GANJI, F. (2025). Investigating Neuroscience-Inspired Shark Algorithms: Mimicking Human Decision-Making in Trading Systems and Their Implications for Accounting, Risk Management, Technical Analysis, and the Stock Market. International Journal of Business Management and Entrepreneurship, 4(2), 71–85. Retrieved from https://www.mbajournal.ir/index.php/IJBME/article/view/78

Ganji, F. (2021). Knowledge and acceptance and use of technology in accounting students. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 12(13), 7466–7479. https://scholar.google.com/citations?view_op=view_citation&hl=tr&user=_RyCeTEAAAAJ&citation_for_view=_RyCeTEAAAAJ:4DMP91E08xMC

Cressey, D. R. (2018). The Cressey hypothesis (1953) and the investigation of corporate fraud occurrence: An empirical analysis in Brazilian banking institutions. *Revista Contabilidade & Finanças, 29*(76), 60–81. https://doi.org/10.1590/1808-057x201803270

Dechow, P. M., Ge, W., Larson, C. R., & Sloan, R. G. (2011). Predicting material accounting misstatements. *Contemporary Accounting Research, 28*(1), 17–82. https://doi.org/10.1111/j.1911-3846.2010.01041.x

Dong, W., Liao, S., & Zhang, Z. (2018). Leveraging financial social media data for corporate fraud detection. *Journal of Management Information Systems, 35*(2), 461–487. https://doi.org/10.1080/07421222.2018.1451954

Gupta, S., & Mehta, S. K. (2024). Data mining-based financial statement fraud detection: A systematic literature review and meta-analysis. *Global Business Review, 25*(5), 1290–1313. https://doi.org/10.1177/0972150920984857

GANJI, F. . (2025). Advanced Accounting Techniques for Pandemic-Era Financial Reporting. International Journal of Business Management and Entrepreneurship, 4(1), 36–45. Retrieved from https://mbajournal.ir/index.php/IJBME/article/view/63