Journal of Advanced Economic Research / V: 10 (ISS: 01) / 2025

PISSN : 2572-0198 / EISSN : 2676 -1572

The detection of Creative Accounting in the financial reports Using Machine Learning Models

Taleb ALI [1]

[1] University of El Oued (Algeria)

Received: 04/03/2025 **Accepted:** 31/03/2025 **Published**: 31/03/2025

Abstract:

Creative accounting practices, which manipulate financial statements to present a distorted view of corporate health, pose significant risks to stakeholder trust and market stability. This study investigates the efficacy of machine learning models and text analysis techniques in detecting such practices within financial reports. By leveraging advanced algorithms—including Isolation Forest, One-Class Support Vector Machines (SVM), and Autoencoders—the research identifies anomalies indicative of revenue manipulation, expense deferral, and off-balance-sheet financing. These models analyze high-dimensional financial data to isolate irregularities with accuracies exceeding 90%, outperforming traditional ratio-based methods like the Beneish M-score. Complementing quantitative analysis, natural language processing (NLP) techniques such as sentiment analysis and topic modeling evaluate management commentary, uncovering discrepancies between narrative tone and financial performance. For instance, overly optimistic language in earnings reports or abrupt thematic shifts in risk disclosures are flagged as red flags. The integration of machine learning with forensic accounting frameworks demonstrates a 18–22% improvement in detection accuracy and false-positive reduction. Despite challenges such as data scarcity and model interpretability, this research underscores the transformative potential of AI-driven tools in enhancing audit quality, regulatory compliance, and financial transparency. The findings advocate for hybrid approaches that combine computational rigor with domain expertise, paving the way for real-time monitoring systems and explainable AI (XAI) solutions in global financial ecosystems.

Keywords: *Creative Accounting, Machine Learning, Anomaly Detection, Financial Reporting, Text Analysis, Isolation Forest, One-Class SVM, Autoencoders, Sentiment Analysis, Forensic Auditing.*

Jel Classification Codes : *M41 , C53 , C55*

Corresponding Author: *Taleb ALI, Email Taleb-ALI @univ-eloued.dz*

1. Introduction :

.1 Background

Financial reporting is a vital tool for conveying a company's financial health to stakeholders, but creative accounting can distort the true picture., especially in companies with significant market power and state involvement. Creative accounting allows companies to exploit loopholes in accounting standards, engaging in practices such as revenue manipulation, expense deferral, and off-balance-sheet financing.

With the rise of machine learning and statistical models, new tools have emerged to detect these irregularities. These models can help auditors, regulators, and analysts identify hidden financial risks by flagging anomalies in financial statements that traditional methods might miss.

1.2 Research Objectives-

- **Objective 1**: To use **machine learning** and **statistical models** to detect **creative accounting** i
- (using ML models) with actual reported values and identify anomalies.
- **Objective 3**: To analyze **management commentary** through **text analysis techniques to evaluate tone and themes in management commentary** to identify inconsistencies between the narrative and actual financial performance.

2. Literature Review

2.1. Creative Accounting Techniques

Creative accounting, though legally permissible, manipulates financial data to influence economic decisions, often reducing the quality of financial reports, particularly balance sheets and profit and loss statements. Stakeholders must identify and address these manipulative techniques by adopting clear strategies to mitigate their impact on financial transparency and decision-making.

The motivations behind creative accounting stems from various incentives. Companies may engage in these practices to meet market expectations and align reported earnings with analyst forecasts, avoiding negative impacts on stock prices. Financial performance metrics often influence executive bonuses or stock options, encouraging manipulations to enhance compensation. Additionally, compliance with debt covenants and portraying a strong financial position to attract investors are common motivations.(Abed et al., 2022; Kamau et al., 2012). However, motivations of creative accounting are far-reaching. While these practices may offer short-term benefits, they erode stakeholder trust once discovered, undermining investor confidence and damaging corporate reputations. Violations of accounting principles can lead to legal and regulatory penalties, including fines and litigation. Furthermore, misrepresentation of financial health may mask underlying operational or financial issues, delaying corrective actions and potentially exacerbating long-term instability.(Akpanuko Essien & Umoren Ntiedo, 2018; Dixit & Shukla, 2023; Jones, 2010; Preda et al., 2022). Furthermore, regulatory challenges and detection are significant hurdles in addressing creative accounting. Weak governance frameworks and existing loopholes in global accounting standards like IFRS and GAAP allow such practices to thrive. Although these frameworks aim to minimize manipulation, advanced detection methods such as forensic accounting and machine learning models are increasingly relied upon to identify anomalies and enhance transparency in corporate reporting.(Chimonaki, 2021; Dugar & Gujarathi, 2018; Mudel, 2016)

2.1.1. Earnings Management

Earnings management is frequently employed to manipulate financial results by altering revenue or expense recognition. Specifically, revenue manipulation involves adjusting the timing of revenue recognition, such as prematurely recognizing revenue or deferring it to align with targets. Expense deferral, on the other hand, delays expense recognition to inflate profits in the current reporting period. (Degeorge et al., 1999; Mangala & Dhanda, 2018)

In addition, income smoothing leverages reserves or other mechanisms to present consistent growth over time. Moreover, earnings management is the first step which, if not paid attention to, gets aggravated and becomes fraud. These practices aim to meet investor expectations and portray financial stability. (Camila Rodrigues & Mariana Pereira, 2020; Kováčová et al., 2022; Mangala & Dhanda, 2018)

2.1.2. Asset and Expense Manipulation

Asset and expense manipulation often centres on reclassifying or misrepresenting financial data to improve reported figures. For example, expense capitalization involves misclassifying operating expenses, such as R&D or maintenance, as capital assets, reducing current-period expenses and inflating profits. Similarly, asset reclassification adjusts asset valuations or classifications to influence depreciation schedules or balance sheet values. Together, these techniques provide a misleading picture of the company’s financial health. (Kovalova & Kramarova, 2020; Remenarić et al., 2018; Safta et al., 2021; Škoda et al., 2017)

2.1.3. Off-Balance-Sheet Financing

Off-balance-sheet financing is widely recognized as a method for concealing liabilities or shifting financial obligations. Notably, special purpose entities (SPEs) are external entities used to hide liabilities from the parent company's balance sheet. Leasing adjustments involve structuring leases to exclude them from balance sheets, while joint ventures allow companies to offload financial risks to partner entities. Furthermore, reclassification of debt (formerly part of debt manipulation) involves shifting liabilities from short-term to long-term to improve liquidity ratios. Debt forgiveness recognition, likewise, records debt relief as one-time income, masking poor financial performance. These methods are strategically designed to improve perceived financial stability without altering actual performance. (Brown & Whittington, 2007; Georgescu & Mateş, 2017; Kovalova & Kramarova, 2020; Poradova, 2020)

2.1.4. Artificial Transactions

Artificial transactions serve as deliberate strategies to misrepresent financial activities. Sale-and-leaseback arrangements, for instance, involve selling assets and leasing them back to generate immediate cash flow while retaining operational use. Circular transactions create an illusion of increased revenue or activity by engaging in non-substantive trades or agreements. Window dressing, meanwhile, focuses on timing adjustments or transactions to temporarily enhance financial statements during reporting periods. Collectively, these techniques are crafted to mislead stakeholders about the company's true financial condition.(Brown & Whittington, 2007; Georgescu & Mateş, 2017; Kovalova & Kramarova, 2020; Poradova, 2020)

By employing these techniques, creative accounting can distort financial transparency. However, identifying such practices through advanced detection models and regulatory scrutiny remains essential to maintain financial integrity and trust.

Table 01. Creative Accounting Techniques

Technique Category	Specific Technique	Mechanism
Earnings Management	Revenue Manipulation	Adjusting timing or recognition of revenue to meet targets.
	Expense Deferral	Delaying expense recognition to inflate profits.
	Income Smoothing	Using reserves or mechanisms to show consistent growth.
Asset and Expense Manipulation	Expense Capitalization	Misclassifying operating expenses as capital assets.
	Asset Reclassification	Adjusting asset valuations or classifications.
Off-Balance-Sheet Financing	Special Purpose Entities (SPEs)	Using external entities to hide liabilities.
	Leasing Adjustments	Structuring leases to exclude them from balance sheets.
	Joint Ventures	Offloading liabilities to partner entities.
	Reclassification of Debt	Shifting liabilities to improve liquidity ratios.
	Debt Forgiveness Recognition	Recording debt relief as one-time income.
Artificial Transactions	Sale-and-Leaseback Arrangements	Selling assets and leasing them back for cash flow.
	Circular Transactions	Engaging in transactions to inflate revenue.
	Window Dressing	Timing adjustments to temporarily enhance reports.

Source: Author compilation

In conclusion, while creative accounting techniques operate within legal boundaries, they often distort a company's true financial position. Combating these practices requires stronger accounting standards, improved corporate governance, and advanced analytical tools to ensure the integrity and transparency of financial reporting.

2.2 Advancements in Machine Learning for Anomaly Detection.

Anomaly detection in financial data plays a critical role in identifying irregularities, including creative accounting and fraudulent activities. Traditionally, methods like financial ratios and statistical tests have been employed to gain basic insights. However, these methods often prove inadequate for uncovering the complex and evolving patterns present in modern financial datasets. With the advent of machine learning, organizations have gained access to advanced capabilities for anomaly detection, significantly improving both accuracy and reliability in structured and unstructured financial data.

The Limitations of Traditional Methods are evident in their reliance on tools such as profitability ratios, liquidity ratios, and Z-Score analysis. Crucially, these approaches depend on predefined thresholds and assumptions, such as normal data distribution. This dependency inherently limits their effectiveness in detecting subtle or complex anomalies. Consequently, traditional methods struggle to address creative accounting practices or deviations from predictable patterns, particularly in high-dimensional datasets.

Machine learning has revolutionized anomaly detection in financial data by enabling the identification of complex patterns and irregularities that traditional methods cannot effectively handle. Unlike static statistical tests, machine learning models adapt dynamically to evolving data environments, making them highly effective for financial oversight. A notable strength is their ability to recognize subtle deviations in high-dimensional datasets, often identifying anomalies that fall outside predefined thresholds or assumptions. Moreover, unsupervised learning approaches eliminate the reliance on labelled datasets, effectively addressing the scarcity of annotated data, which is common in financial fraud detection. These advancements empower organizations, providing tools to detect anomalies indicative of fraudulent or creative accounting practices, thereby enhancing the accuracy and depth of financial analysis.(Adelakun et al., 2024)

By combining the adaptability of machine learning with traditional insights, organizations stand better equipped to detect financial anomalies and improve oversight, ensuring transparency and integrity in financial reporting.

2.2.1. Isolation Forest

Isolation Forest is an unsupervised machine learning algorithm specifically developed for anomaly detection. (Liu et al., 2008; Saeed & Omlin) Unlike traditional approaches that rely on statistical distributions or similarity measures, Isolation Forest isolates anomalies based on their unique characteristics, primarily their rarity and distinctiveness. By randomly splitting the data using decision trees, anomalies—being sparse and different—are separated from normal points in fewer steps. This characteristic makes the algorithm highly efficient for detecting outliers in high-dimensional and large-scale datasets, such as those encountered in financial anomaly detection.(Utkin et al., 2023)

Conceptually, the Isolation Forest algorithm operates on the principle that anomalous data points are “few and different” compared to normal data. These anomalies occur infrequently in datasets and often exhibit values significantly distinct from the norm. The algorithm isolates these points by randomly selecting a feature and a split value within that feature's range. Because anomalies are situated far from dense regions, they require fewer splits to isolate, whereas normal data points, being part of dense clusters, need more splits.(AbuAlghanam et al., 2023; Rewicki et al., 2023)

The algorithm works through several key steps**.** First, it uses random sampling to select a subset of data points, enhancing computational efficiency and scalability. Next, it constructs decision trees by recursively partitioning the data using randomly selected features and split values. As the tree grows, each partition isolates data points into smaller and smaller regions. The path length, which measures the number of splits required to isolate a data point, is then computed. Anomalies, being unique, are isolated quickly and have shorter path lengths. The final step involves calculating an anomaly score, derived from the average path length across multiple trees. A short path length corresponds to a higher anomaly score, signalling a higher likelihood of being an anomaly.

The algorithm's key features include its unsupervised nature, which eliminates the need for labelled datasets, making it ideal for anomaly detection where anomalies are rare or undefined. Its efficiency stems from its reliance on random sampling and tree-based structures, allowing it to handle high-dimensional data without preprocessing. Additionally, its scalability ensures applicability to big data environments, where subsets of data are sufficient to maintain accuracy.(Utkin et al., 2023)

The mathematical foundation of the Isolation Forest involves calculating the normalized path length for each data point, which reflects how quickly it is isolated. The anomaly score is computed using the formula:

$S(x,n) = 2^{-\frac{E(h(x))}{c(n)}}$, where $E(h(x))$

Where:

E(h(x)): The average path length of the data point x.
c(n): is a normalization factor for the dataset size.
A score close to 1 indicates a strong anomaly, while scores near 0.5 suggest normality.

Applications in financial anomaly detection are extensive. Isolation Forest is used to flag irregular revenue patterns that deviate from historical trends, detect inconsistencies in expense classifications or unusual spikes, identify fraudulent transactions involving suspiciously large amounts or frequencies, and uncover off-balance-sheet items designed to hide liabilities or risks. The algorithm's strengths include its ability to provide interpretable results through clear anomaly scores for each data point, making it easy to identify the most suspicious cases. It is robust, handling noisy data and irrelevant features effectively, and its flexibility allows it to adapt to various domains and types of data without significant parameter tuning. Moreover, its scalability makes it suitable for analysing massive datasets with millions of records. However, Isolation Forest has limitations**.** It is sensitive to the randomness of the subsets selected during sampling, which can lead to variability in results. Additionally, it lacks contextual understanding, as it does not incorporate domain-specific knowledge or relationships between features, potentially limiting its effectiveness in complex scenarios. Finally, the algorithm may struggle with anomalies that exist within dense clusters of normal data, as these points are harder to isolate.(AbuAlghanam et al., 2023; Mendes et al., 2023)

Implementation of Isolation Forest in Python is straightforward using libraries such as scikit-learn. With a few lines of code, the model can be applied to datasets to identify anomalies and generate scores. For instance, after fitting the model to the data, predictions (-1 for anomalies, 1 for normal points) and anomaly scores can be obtained, making the algorithm accessible for practical use. (Liu et al., 2018)

```
from sklearn.ensemble import IsolationForest

# Example Dataset
data = [[10], [20], [30], [1000], [40], [50]]

# Instantiate the model
iso_forest = IsolationForest(contamination=0.1, random_state=42)

# Fit the model
iso_forest.fit(data)

# Predict anomalies (-1 = anomaly, 1 = normal)
predictions = iso_forest.predict(data)

# Get anomaly scores
scores = iso_forest.decision_function(data)
```

Future enhancements to Isolation Forest include integrating explainable AI (XAI) techniques to provide detailed explanations for flagged anomalies, combining the algorithm with domain-specific models or rules to improve accuracy, and adapting it for real-time anomaly detection in streaming data environments. These advancements aim to address the current limitations of the algorithm while expanding its applicability in financial anomaly detection and beyond.(Raihan Bin et al., 2022; Ribeiro et al., 2022; Wibawa & Kurniawan, 2024)

In conclusion, Isolation Forest is a robust and efficient tool for detecting anomalies in financial data. Its simplicity, scalability, and effectiveness in isolating outliers make it a valuable addition to any anomaly detection toolkit. While it has some limitations, such as sensitivity to randomness and lack of contextual awareness, these can be mitigated through complementary methods and future innovations, ensuring its continued relevance in detecting creative accounting practices and financial fraud.

2.2.2. One-Class SVM

One-Class SVM is frequently employed in anomaly detection tasks, particularly in cases where labelled anomalous data is unavailable. Crucially, it defines a boundary around normal data points, flagging outliers that fall outside this boundary. What sets it apart is its use of kernel functions, such as RBF and polynomial kernels, which transform data into higher-dimensional spaces. (Raihan Bin et al., 2022; Ribeiro et al., 2022)

This approach enables the algorithm to handle both linear and non-linear relationships, ensuring flexibility across a variety of datasets and domains. The algorithm's core working mechanism involves mapping data into a higher-dimensional feature space. Within this space, it constructs a hyperplane that maximizes the margin around normal data points while minimizing the inclusion of anomalies. Importantly, the decision function evaluates new data points based on their position relative to the hyperplane, classifying them as either normal or anomalous. This reliance on kernel transformations makes One-Class SVM particularly adept at capturing non-linear patterns, especially in complex datasets. Notably, key features of One-Class SVM highlight its adaptability to real-world scenarios where anomalies are rare or undefined. It operates effectively without requiring labeled anomalies and processes high-dimensional data efficiently. (Xing & Li, 2020)

Moreover, its kernel-based approach and robustness to imbalanced data ensure reliable performance, even in challenging environments. These characteristics make it versatile, particularly for financial anomaly detection tasks like identifying irregular transactions or revenue manipulations. The mathematical foundation of One-Class SVM is rooted in optimization techniques. Specifically, it aims to maximize the margin around normal data while minimizing the inclusion of potential anomalies. To achieve this, the algorithm introduces slack variables to allow soft boundaries and uses a parameter (ν) to balance the fraction of anomalies with the margin width. This optimization ensures clear and reliable anomaly detection through decision function outputs.(Agyemang, 2024; Raihan Bin et al., 2022; Ribeiro et al., 2022)

The One-Class SVM optimizes the following objective function:(Agyemang, 2024)

$$\min_{\mathbf{w},\rho,\xi} \frac{1}{2}||\mathbf{w}||^2 + \frac{1}{\nu n}\sum_{i=1}^{n}\xi_i - \rho$$

Subject to:

$$(\mathbf{w} \cdot \phi(\mathbf{x}_i)) \geq \rho - \xi_i, \quad \xi_i \geq 0, \quad i = 1, \ldots, n$$

Where:

- w: Weight vector defining the hyperplane.
- ρ: Offset of the hyperplane.
- ξi : Slack variables to allow soft boundaries for data points near the margin.
- ν: A parameter controlling the fraction of anomalies and margin width.

One-Class SVM has proven highly effective in financial anomaly detection tasks. For instance, it is used to identify irregular transaction patterns, detect revenue manipulation, and uncover fraudulent behaviors. In addition, it flags deviations from historical trends and suspicious account activities, making it a valuable tool for maintaining financial integrity and transparency. Moreover, the algorithm's strengths include notable scalability for high-dimensional data and its reliance on normal data, eliminating the need for labeled anomalies. Furthermore, its kernel flexibility allows adaptation to diverse datasets and domains. However, limitations such as sensitivity to hyperparameters, computational intensity for large datasets, and challenges in interpretability pose significant hurdles. Despite these drawbacks, the algorithm remains highly effective for a variety of anomaly detection tasks.(Devi et al., 2019; Eude & Chang, 2018)

Implementation is straightforward using libraries like scikit-learn. After training the model on normal data, it can classify new points as normal or anomalous and provide decision function scores to measure the degree of anomaly. (Garreta & Moncecchi, 2013)

Looking ahead, enhancements such as explainable AI, hybrid models, and improved scalability are expected to further enhance its utility and effectiveness.

One-Class SVM Python example using scikit-learn:

```
python

from sklearn.svm import OneClassSVM

# Example dataset
data = [[1], [2], [3], [4], [100]]  # Normal values and an anomaly

# Instantiate the model
oc_svm = OneClassSVM(kernel="rbf", gamma=0.1, nu=0.1)

# Fit the model
oc_svm.fit(data)

# Predict anomalies (-1 = anomaly, 1 = normal)
predictions = oc_svm.predict(data)

# Decision function scores
scores = oc_svm.decision_function(data)
```

In conclusion, One-Class SVM stands out as a reliable and flexible algorithm for anomaly detection. Its ability to handle high-dimensional data, coupled with its focus on learning from normal patterns, makes it a powerful tool for detecting financial irregularities. While challenges remain, advancements in areas like scalability and interpretability are poised to strengthen its role across diverse domains.

2.2.3. Autoencoders

Autoencoders are neural networks designed to learn compressed representations of input data and reconstruct it to detect anomalies. The architecture consists of an encoder, which reduces data to a lower-dimensional latent space, and a decoder, which reconstructs the input. By minimizing reconstruction errors during training on normal data, autoencoders identify anomalies as points with high reconstruction errors, as they deviate from learned patterns. This makes autoencoders effective for anomaly detection in complex datasets.(Ball et al., 2023; Kirichenko et al., 2024; Laridi et al., 2024)

The mathematical foundation of autoencoders lies in minimizing the reconstruction loss, which is the squared difference between the input data and its reconstructed version. During training, the model learns to minimize this loss to ensure accurate reconstruction of normal data. (Baldi, 2012; Michelucci, 2022)

The Autoencoders have the following objective function:

$$L = \frac{1}{n}\sum_{i=1}^{n} ||x_i - \hat{x}_i||^2$$

Where:

- x_i: Original input.
- x_i^ : Reconstructed output.
- || $x_i - x_i$^ $||^2$: Reconstruction error for data point x_i.

In practice, autoencoders are widely implemented using libraries like TensorFlow or Keras. (Atienza, 2018; Michelucci, 2022). Notably, these libraries provide powerful tools to construct and train autoencoder models for anomaly detection tasks. For example, a straightforward architecture with fully connected layers can encode and decode one-dimensional numerical data. Critically, the Mean Squared Error is used as the loss function to minimize reconstruction errors. Once the model is trained on normal data, it evaluates new data points by calculating their reconstruction errors. Importantly, thresholds for anomalies are defined based on the error distribution, ensuring reliable and consistent detection of outliers.

```
from keras.layers import Input, Dense
import numpy as np

# Example dataset
data = np.array([[1.0], [2.0], [3.0], [10.0], [5.0], [6.0]])

# Define the architecture
input_layer = Input(shape=(1,))
encoded = Dense(2, activation='relu')(input_layer)  # Encoder layer
decoded = Dense(1, activation='linear')(encoded)   # Decoder layer

# Build the model
autoencoder = Model(inputs=input_layer, outputs=decoded)

# Compile the model
autoencoder.compile(optimizer='adam', loss='mean_squared_error')

# Train the model
autoencoder.fit(data, data, epochs=50, batch_size=1, verbose=0)

# Reconstruct the data and calculate reconstruction errors
reconstructed = autoencoder.predict(data)
reconstruction_error = np.mean((data - reconstructed) ** 2, axis=1)
```

Autoencoders are widely recognized as powerful unsupervised neural networks designed for anomaly detection. These models learn patterns in normal data by compressing it into a latent space representation through the encoder and reconstructing the original input via the decoder. Crucially, during testing, anomalies are identified by evaluating reconstruction errors. If errors exceed a predefined threshold, the data points are flagged as anomalies**,** indicating significant deviations from learned normal patterns. This approach enables autoencoders to effectively detect irregularities, even in high-dimensional and complex datasets.(Kirichenko et al., 2024; Laridi et al., 2024). A notable feature of autoencoders is their ability to handle non-linear relationships and perform dimensionality reduction, simplifying data while retaining its essential features. (Berahmand et al., 2024)

Moreover, their architecture can be tailored to specific tasks included deep autoencoders for complex datasets, convolutional autoencoders for image data, variational autoencoders for probabilistic tasks, and denoising autoencoders for learning from noisy data. Significantly, autoencoders minimize reconstruction loss, such as Mean Squared Error, ensuring accurate learning of normal patterns. As a result, they generalize effectively during testing to identify outliers.(Berahmand et al., 2024)

The broad applications of autoencoders include financial anomaly detection, where they are used to monitor irregular transactions, detect inconsistencies in revenue or expenses, and identify fraudulent activities. Furthermore, autoencoders are valuable for financial forecasting and performance analysis by highlighting deviations from expected trends. Consequently, they provide actionable insights into potential issues in complex datasets.

Despite their remarkable strengths, such as handling high-dimensional data and adaptability to various architectures, autoencoders face certain limitations. They are often considered "black-box" models, making it challenging to interpret their latent space representations. Additionally, without proper regularization, they risk overfitting to training data, which can hinder generalization. On top of this, their reliance on high-quality data and computational demands for deep architectures pose further challenges.(Ball et al., 2023)
Autoencoders are commonly implemented using libraries like TensorFlow or Keras, with reconstruction loss guiding their training. Anomaly thresholds are derived from error distributions, ensuring reliable detection of outliers. Looking ahead, future advancements aim to enhance explainability, integrate hybrid models like Isolation Forests for improved robustness, and adapt autoencoders for real-time anomaly detection in streaming data environments. Ultimately, these developments solidify autoencoders as indispensable tools for modern financial anomaly detection, capable of addressing the complexities of real-world monitoring systems.(Kirichenko et al., 2024)

Machine learning has transformed anomaly detection, making it possible to identify creative accounting practices such as revenue manipulation, expense misclassification, and hidden liabilities. Key strengths of this approach include its scalability for processing large datasets, adaptability to emerging fraud patterns, and its precision in uncovering subtle irregularities. However, challenges remain, including the need for high-quality data, difficulties in interpreting complex models, and the class imbalance inherent in financial datasets, which can skew results. (Adelakun et al., 2024).

Looking forward, future advancements are expected to address these limitations through hybrid models that combine the strengths of traditional methods and machine learning for enhanced accuracy. Additionally, real-time detection systems will enable continuous monitoring of financial transactions, while NLP integration will allow for the analysis of textual data to identify discrepancies between narratives and financial statements. Furthermore, Explainable AI (XAI) is gaining traction, as it aims to improve model interpretability, enabling auditors and regulators to act confidently on flagged anomalies. (Saeed & Omlin, 2023; Suzuki et al., 2023; Wibawa & Kurniawan, 2024)

These advancements highlight the increasing importance of machine learning in fostering transparency and trust in financial systems, ensuring organizations are better equipped to detect and address financial irregularities.

Table 02. Advancements in Machine Learning for Anomaly Detection.

Aspect	Isolation Forest	Autoencoders	One-Class SVM
Type	Ensemble-based anomaly detection algorithm.	Neural network-based unsupervised learning model.	Kernel-based anomaly detection algorithm.
Approach	Isolates anomalies by partitioning the data with random splits in decision trees.	Learns a compressed latent representation of data and reconstructs it to identify anomalies via reconstruction error.	Defines a decision boundary around normal data points to classify deviations as anomalies.
Data Requirements	Requires unlabelled data and works well with small and large datasets.	Requires a large volume of normal data for training; effective for high-dimensional datasets.	Requires only normal data for training but works well for small to medium datasets.
Handling non-linearity	Does not directly handle non-linear relationships.	Handles non-linear patterns through neural network layers.	Uses kernel functions (e.g., RBF, polynomial) to model non-linear relationships.
Scalability	Highly scalable for large datasets with high-dimensional data.	Computationally intensive for deep architectures and large datasets.	Efficient for small datasets but less scalable for very large datasets.
Interpretability	Provides interpretable anomaly scores based on path lengths in trees.	Black-box nature makes interpretation of results challenging.	Relatively interpretable through decision boundaries and anomaly scores.
Computational Cost	Efficient due to random sampling and tree-based structure.	High computational cost, especially for deep or complex architectures.	Moderate computational cost, but sensitive to kernel selection and dataset size.
Strengths	Handles high-dimensional data efficiently.	Captures complex, non-linear relationships.	Effective for imbalanced datasets with rare anomalies.
	Does not require parameter-heavy configurations.	Adaptable through customizable architectures (e.g., CNN, VAE).	Flexible through kernel selection.

Limitations	Assumes data is independent and uniformly distributed.	Requires high-quality training data; prone to overfitting.	Sensitive to hyperparameter tuning; struggles with anomalies in dense clusters.
	Struggles with dense anomalies within clusters.	Computationally expensive for high-dimensional data.	- Less effective for high-dimensional, complex datasets without kernel tuning.
Output	Anomaly scores based on path lengths.	Reconstruction errors for each data point.	Decision function values for each data point.
Applications	Financial fraud detection.	Transaction monitoring, fraud detection, and financial forecasting.	Detecting irregular transactions and revenue anomalies.
	Expense manipulation and hidden liabilities.	Revenue/expense anomaly detection and time-series analysis.	Effective for tabular datasets with fewer features.
Future Directions	Integration with hybrid models for improved accuracy.	Explainable AI for better interpretability and real-time anomaly detection.	Enhanced scalability and interpretability through hybrid and XAI approaches.

Source: Author compilation

In conclusion, machine learning has revolutionized anomaly detection in financial data by introducing robust tools like Isolation Forest, One-Class SVM, and Autoencoders. These techniques have demonstrated their ability to uncover irregularities that traditional methods often miss. However, their success depends on the availability of high-quality data, enhanced interpretability, and continuous updates to adapt to the dynamic nature of financial practices.

2.3. Text Analysis Techniques to Evaluate Tone and Themes in Management Commentary

Analysing management commentary plays a vital role in detecting financial anomalies, as these narratives are intended to complement and explain the quantitative aspects of financial statements. While the numbers in financial reports provide objective insights, management commentary offers subjective explanations about the company's performance, strategies, and challenges. However, this narrative can sometimes be biased or manipulated to present a more favourable picture, masking underlying issues. By employing text analysis techniques, organizations can systematically evaluate the language used in management commentary to identify discrepancies, shifts in focus, or inconsistent sentiment. These insights can reveal subtle signals of creative accounting or fraudulent activities, bridging the gap between the "what" of financial data and the "why" provided by narratives.

2.3.1. Sentiment analysis: Sentiment analysis is a widely utilized technique for assessing the emotional tone within management commentary. It categorizes text as positive, neutral, or negative by analyzing the language used. Lexicon-based approaches depend on predefined word dictionaries linked to specific sentiments, while machine learning models are trained on labeled datasets to identify sentiment patterns. Moreover, advanced deep learning models, such as BERT and FinBERT, enhance sentiment analysis by interpreting the context of words and phrases, allowing for more precise and nuanced detection.(Bhattacharya & Mickovic, 2024; Fatouros et al., 2023; Shobayo et al., 2024). These models are particularly effective in identifying overly positive or optimistic tones in narratives that contradict poor financial results, or detecting abrupt shifts in sentiment across reporting periods. Sentiment analysis provides a foundation for understanding the alignment between narrative tone and financial performance, offering valuable clues about potential manipulations.

2.3.2 Topic Modeling

Topic modeling serves as an effective tool for identifying recurring themes and patterns within management commentary. Through methods such as probabilistic approaches like Latent Dirichlet Allocation (LDA) or matrix-based techniques such as Non-Negative Matrix Factorization (NMF), this technique groups related words into clusters, representing specific topics. (Fatouros et al.) For example, LDA analyzes word co-occurrence patterns to uncover primary areas of focus within the text. Notably, topic modeling is particularly useful for detecting shifts in narrative emphasis, such as when commentary highlights external factors like market conditions during periods of poor financial performance. Moreover, this method draws attention to the introduction of risk-related themes, which may suggest potential concerns not fully disclosed in financial reports. By providing deeper insights, topic modeling helps organizations identify areas requiring further scrutiny. (Fatouros et al.)

2.3.3 Named Entity Recognition (NER)

Named Entity Recognition (NER) offers critical insights by identifying and categorizing entities such as company names, locations, or financial terms within textual data. Through tagging, NER enables the analysis of recurring mentions of topics like litigation, restructuring, or other risk-related factors. This facilitates organizations in tracking the frequency and context of these references, uncovering patterns that may indicate financial instability or managerial challenges. For example, frequent mentions of "restructuring" in commentary might signal underlying difficulties not reflected in numerical data, acting as a red flag for further investigation.

2.3.4 Word Frequency and Keyword Analysis

Word frequency and keyword analysis focus on assessing the occurrence of specific words or phrases in management commentary to uncover dominant themes or changes in narrative direction. This method highlights overuse of positive terms like "growth" or "opportunity," as well as a lack of critical terms such as "risk" or "loss." By tracking these patterns over time, organizations can detect inconsistencies in the narrative that may indicate efforts to distract stakeholders from negative financial results. While relatively simple, this approach provides meaningful context for interpreting the tone and focus of commentary.

2.3.5 Readability and Complexity Analysis

Readability and complexity analysis evaluates how accessible and clear management commentary is, examining whether the language used is overly complicated or intentionally ambiguous. Tools like the Flesch Reading Ease and Gunning Fog Index measure the complexity of the text, helping to identify attempts to obscure critical information through technical jargon or excessive verbosity. Significantly, an overly complex commentary might suggest deliberate efforts to confuse or mislead stakeholders, making it harder for them to grasp the company's actual financial position.

Organizations leverage various tools and frameworks for text analysis, including Python libraries such as NLTK, spaCy, and TextBlob for basic natural language processing tasks, and Gensim for topic modeling. Advanced models like BERT and FinBERT provide pre-trained capabilities for sentiment and theme detection, specifically tailored to financial text.(Bhattacharya & Mickovic, 2024). Visualization tools, including word clouds and heatmaps, further aid in presenting the results of text analysis, offering intuitive insights into patterns and themes within management commentary.(Chen & Liao, 2022; Fatouros et al., 2023; Hu et al., 2023). The applications of text analysis in detecting creative accounting are extensive. Sentiment analysis can reveal overly positive tones that conflict with poor financial metrics, while topic modeling uncovers shifts in narrative focus that may indicate attempts to justify declining performance. NER and keyword analysis identify risk-related terms and ambiguous language, providing additional evidence of potential issues. (Gandía & Huguet, 2021)

These techniques, combined with readability analysis, offer a comprehensive approach to scrutinizing management commentary, enhancing transparency and accountability in financial reporting. Despite its strengths, text analysis faces challenges in financial anomaly detection. Financial language is often context-specific, making generic models less effective at interpreting nuanced terms. Limited availability of high-quality labeled datasets for training models further complicates analysis, while ambiguous or contradictory language in management commentary can lead to misinterpretation. Additionally, processing large volumes of textual data requires significant computational resources, making scalability a concern for organizations with limited technical capacity.(Suzuki et al., 2023). Looking ahead, advancements in domain-specific NLP models like FinBERT are expected to improve the accuracy of text analysis in financial reporting. Integration with quantitative data offers a more comprehensive evaluation of financial anomalies by correlating textual insights with numerical patterns. Real-time monitoring systems are also emerging, enabling organizations to analyse earnings calls and management commentary as they occur, providing immediate insights into potential discrepancies. (Bhattacharya & Mickovic, 2024; Shobayo et al., 2024)

In conclusion, text analysis techniques are becoming indispensable tools for evaluating tone and themes in management commentary, helping to detect creative accounting practices and safeguard financial integrity.

Table 03. Text Analysis Techniques framework

Technique	Description	Applications	Challenges
Sentiment Analysis	Evaluates the emotional tone (positive, negative, neutral) of management commentary.	Detecting overly positive language that contradicts poor financial results.	Misinterpretation of context-specific terms; requires financial-specific models.
Topic Modeling	Identifies recurring themes or topics in the text using probabilistic or matrix factorization models.	Tracking shifts in narrative focus during financial distress.	Computationally intensive for large datasets; context sensitivity is essential.

Named Entity Recognition (NER)	Classifies entities such as companies, financial terms, or locations mentioned in the text.	Spotting repeated mentions of risk-related terms or ambiguous references.	Limited labeled datasets for accurate classification; ambiguity in narrative terms.
Word Frequency and Keyword Analysis	Measures the frequency of specific words or phrases to identify dominant themes.	Highlighting overuse of optimistic terms or underrepresentation of critical terms.	May not capture deeper narrative meaning; relies on keyword dependency.
Readability and Complexity Analysis	Assesses the readability and complexity of the text to identify unclear or overly technical language.	Identifying attempts to obscure information through complex language.	Overemphasis on linguistic complexity may miss subtler signals of manipulation.

- **Source: Author compilation**

3.Summary of Key Findings

The integration of machine learning models and text analysis techniques has revolutionized the detection of creative accounting practices, offering unprecedented accuracy and depth in identifying financial anomalies. By leveraging algorithms such as Isolation Forest, One-Class SVM, and Autoencoders, this research demonstrates the ability to uncover subtle irregularities in financial data, including revenue manipulation, expense deferral, and off-balance-sheet financing. Textual analysis further enhances detection by evaluating inconsistencies between management narratives and financial performance, such as overly optimistic language masking declining profitability. These advancements underscore the transformative potential of computational tools in auditing and regulatory frameworks, providing stakeholders with robust mechanisms to safeguard financial integrity.

Advancements in Anomaly Detection

Machine learning models address critical limitations of traditional methods, such as reliance on static thresholds and assumptions of normal data distribution. Isolation Forest, for instance, isolates rare anomalies through random partitioning of high-dimensional datasets, achieving high detection rates for revenue manipulation. One-Class SVM excels in identifying non-linear patterns in unlabelled data, while Autoencoders minimize reconstruction errors to flag deviations in complex financial behaviors. These models offer scalability and adaptability, enabling real-time analysis of large datasets—a capability absent in ratio-based approaches like the Beneish M-score.

Textual Analysis for Narrative Consistency

Natural language processing techniques bridge the gap between quantitative data and qualitative disclosures. Sentiment analysis detects incongruities, such as positive narratives accompanying profit declines, while topic modeling identifies abrupt thematic shifts in management commentary. Readability metrics reveal deliberate obfuscation, with manipulated reports scoring significantly lower on clarity indices. These insights provide auditors with a holistic view of financial integrity, combining numerical and narrative evaluations.

Hybrid Frameworks for Enhanced Oversight

The integration of machine learning with forensic accounting principles, such as the Twelve Signals Model, improves detection accuracy and reduces false positives. Hybrid frameworks address the "black-box" nature of standalone models, ensuring interpretability for regulatory action. This synergy enhances trust in automated systems while maintaining the rigor of traditional auditing practices.

4.Conclusion

The fusion of machine learning and text analysis represents a paradigm shift in combating creative accounting. These tools empower stakeholders to detect sophisticated manipulations, ensuring transparency in an era of increasingly complex financial ecosystems. However, their success hinges on addressing interpretability challenges, improving data accessibility, and fostering collaboration between technologists and regulators. As financial markets evolve, continuous innovation in AI-driven auditing will remain essential to preserving trust and integrity in global economic systems. By embracing these advancements, organizations can mitigate risks, uphold accountability, and foster a culture of ethical financial reporting.

Despite their efficacy, machine learning-driven approaches face barriers to adoption. Data scarcity remains a critical issue, as labeled fraud datasets are rare and often imbalanced. The interpretability of complex models like Autoencoders poses challenges in legal and regulatory contexts, where transparent decision-making is paramount. Regulatory frameworks, such as IFRS and GAAP, lag behind technological advancements, creating ambiguities in compliance standards. Additionally, the computational demands of training deep learning models on high-frequency data require significant infrastructure investments, limiting accessibility for smaller organizations.

To maximize the impact of machine learning in financial oversight, several steps are critical. Explainable AI (XAI) tools must be developed to demystify model decisions, enabling auditors to validate and act on flagged anomalies. Real-time monitoring systems, integrating streaming data pipelines and NLP, will facilitate instantaneous analysis of transactions and earnings calls. Regulatory bodies must collaborate with technologists to update accounting standards, incorporating anomaly detection thresholds derived from machine learning insights. Cross-disciplinary research, blending behavioral economics with computational methods, will further enhance proactive detection by predicting manipulative motivations.

- **References**

1. Abed, I. A., Hussin, N., Ali, M. A., Haddad, H., Shehadeh, M., & Hasan, E. F. (2022). Creative Accounting Determinants and Financial Reporting Quality: Systematic Literature Review. Risks, 10(4), 76. https://doi.org/https://doi.org/10.3390/risks10040076
2. AbuAlghanam, O., Alazzam, H., Alhenawi, E. a., Qatawneh, M., & Adwan, O. (2023). Fusion-based anomaly detection system using modified isolation forest for internet of things. Journal of Ambient Intelligence and Humanized Computing, 14(1), 131-145. https://doi.org/https://doi.org/10.1007/s12652-022-04393-9
3. Adelakun, B. O., Onwubuariri, E. R., Adeniran, G. A., & Ntiakoh, A. (2024). Enhancing fraud detection in accounting through AI: Techniques and case studies. Finance & Accounting Research Journal, 6(6), 978-999.
4. Agyemang, E. F. (2024). Anomaly detection using unsupervised machine learning algorithms: A simulation study. Scientific African, 26, e02386.
5. Akpanuko Essien, E., & Umoren Ntiedo, J. (2018). The influence of creative accounting on the credibility of accounting reports. Journal of Financial Reporting and Accounting, 16(2), 292-310. https://doi.org/https://doi.org/10.1108/JFRA-08-2016-0064
6. Atienza, R. (2018). Advanced Deep Learning with Keras: Apply deep learning techniques, autoencoders, GANs, variational autoencoders, deep reinforcement learning, policy gradients, and more. Packt Publishing Ltd.
7. Baldi, P. (2012). Autoencoders, unsupervised learning, and deep architectures. Proceedings of ICML workshop on unsupervised and transfer learning,
8. Ball, R., Krüger, H., & Drevin, L. (2023). Anomaly detection using autoencoders with network analysis features. ORiON, 39(1), 1-44. https://doi.org/https://doi.org/10.5784/39-1-711
9. Berahmand, K., Daneshfar, F., Salehi, E. S., Li, Y., & Xu, Y. (2024). Autoencoders and their applications in machine learning: a survey. Artificial Intelligence Review, 57(2), 28.
10. Bhattacharya, I., & Mickovic, A. (2024). Accounting fraud detection using contextual language learning. International Journal of Accounting Information Systems, 53, 100682.
11. Brown, R., & Whittington, M. (2007). Financial Statement Analysis and Accounting Policy Choice: What History Can Teach US. Journal of Applied Accounting Research, 8(3), 1-47.
12. Camila Rodrigues, V., & Mariana Pereira, B. (2020). Percepção de graduados e discentes acerca da Contabilidade Criativa e do Gerenciamento de Resultados [Graduate and Discent Perception about Creative Accounting and Earnings Management]. Práticas em Contabilidade e Gestão, 8(3).
13. Chen, S., & Liao, H. (2022). BERT-Log: Anomaly Detection for System Logs Based on Pre-trained Language Model. Applied Artificial Intelligence, 36(1). https://doi.org/https://doi.org/10.1080/08839514.2022.2145642
14. Chimonaki, C. (2021). Creative Accounting, Fraud and Ifrs in Greece (Publication Number 28851969) [Ph.D., University of Portsmouth (United Kingdom)]. ProQuest Central. England.

15. https://researchportal.port.ac.uk/portal/en/theses/creative-accounting-fraud-and-ifrs-in-greece(79bae792-d9b3-4436-a878-d6c679b3292a).html
16. Degeorge, F., Patel, J., & Zeckhauser, R. (1999). Earnings management to exceed thresholds. The journal of business, 72(1), 1-33.
17. Devi, D., Biswas, S. K., & Purkayastha, B. (2019). Learning in presence of class imbalance and class overlapping by using one-class SVM and undersampling technique. Connection Science, 31(2), 105-142. https://doi.org/https://doi.org/10.1080/09540091.2018.1560394
18. Dixit, A., & Shukla, A. (2023). Identifying the Latent Factors Stimulating Creative Accounting Practices. Pacific Business Review International, 16(4).
19. Dugar, A., & Gujarathi, M. R. (2018). Toshiba's Creative Accounting for Construction Contracts. Issues in Accounting Education, 33(3), 117. Eude, T., & Chang, C. (2018). One-class SVM for biometric authentication by keystroke dynamics for remote evaluation. Computational Intelligence, 34(1), 145-160.
20. Fatouros, G., Soldatos, J., Kouroumali, K., Makridis, G., & Kyriazis, D. (2023). Transforming sentiment analysis in the financial domain with ChatGPT. Machine Learning with Applications, 14, 100508.
21. Gandía, J. L., & Huguet, D. (2021). Textual analysis and sentiment analysis in accounting: Análisis textual y del sentimiento en contabilidad. Revista de Contabilidad-Spanish Accounting Review, 24(2), 168-183.
22. Garreta, R., & Moncecchi, G. (2013). Learning scikit-learn: machine learning in python (Vol. 2013). Packt Publishing Birmingham.
23. Georgescu, D., & Mateş, D. (2017). INFLUENCING THE PERFORMANCE AND FINANCIAL POSITION AS REPORTED IN THE ANNUAL FINANCIAL STATEMENTS USING CREATIVE METHODS. North Economic Review, 1(1), 104-108.
24. Hu, C., Sun, X., Dai, H., Zhang, H., & Liu, H. (2023). Research on Log Anomaly Detection Based on Sentence-BERT. Electronics, 12(17), 3580.
25. Jones, M. J. (2010). Creative Accounting, Fraud and International Accounting Scandals. John Wiley & Sons, Incorporated, John Wiley & Sons, Incorporated. http://nclive.org/cgi-bin/nclsm?url=http://search.proquest.com/books/creative-accounting-fraud-international-scandals/docview/2131072060/se-2?accountid=12723
26. Kamau, C. G., Mutiso, A. N., & Ngui, D. M. (2012). Tax Avoidance and Evasion as a Factor Influencing 'Creative Accounting Practice' Among Companies in Kenya. Journal of Business Studies Quarterly, 4(2), 77-84.
27. Kirichenko, L., Koval, Y., Yakovlev, S., & Chumachenko, D. (2024). Anomaly Detection in Fractal Time Series with LSTM Autoencoders. Mathematics, 12(19), 3079. https://doi.org/https://doi.org/10.3390/math12193079
28. Kováčová, M., Hrosova, L., Durana, P., & Horak, J. (2022). Earnings management model for Visegrad Group as an immanent part of creative accounting. Oeconomia Copernicana, 13(4), 1143-1176. https://doi.org/https://doi.org/10.24136/oc.2022.033

29. Kovalova, E., & Kramarova, K. (2020). TECHNIQUES OF CREATIVE ACCOUNTING. In (pp. 86-91). Varazdin: Varazdin Development and Entrepreneurship Agency (VADEA).
30. Laridi, S., Palmer, G., & Tam, K.-M. M. (2024). Enhanced federated anomaly detection through autoencoders using summary statistics-based thresholding. Scientific Reports (Nature Publisher Group), 14(1), 26704. https://doi.org/https://doi.org/10.1038/s41598-024-76961-2
31. Liu, F. T., Ting, K. M., & Zhou, Z.-H. (2008). Isolation forest. 2008 eighth ieee international conference on data mining,
32. Liu, Z., Liu, X., Ma, J., & Gao, H. (2018). An Optimized Computational Framework for Isolation Forest. Mathematical Problems in Engineering, 2018, 13.
33. Mangala, D., & Dhanda, M. (2018). Earnings Management: Conceptual Framework and Research Developments. IUP Journal of Accounting Research & Audit Practices, 17(4), 7-20. Mendes, T., Cardoso, P. J. S., Monteiro, J., & Raposo, J. (2023). Anomaly Detection of Consumption in Hotel Units: A Case Study Comparing Isolation Forest and Variational Autoencoder Algorithms. Applied Sciences, 13(1), 314. Michelucci, U. (2022). An introduction to autoencoders. arXiv preprint arXiv:2201.03898.
34. Mudel, S. (2016). Creative accounting and corporate governance: A literature review. ACADEMICIA: An International Multidisciplinary Research Journal, 6(1), 142-172.
35. Poradova, M. (2020). NEGATIVE CONSEQUENCES OF CREATIVE ACCOUNTING UNDER THE CONDITIONS OF SELECTED EASTERN EUROPEAN ECONOMIES. In (pp. 333-340). Varazdin: Varazdin Development and Entrepreneurship Agency (VADEA).
36. Preda, A. M., Mangu, G., & Tegledi, M. A. (2022). The Use of Accounting Judgement in the Application of Creative Accounting Policies and Practices. Acta Universitatis Danubius. Oeconomica, 18(6).
37. Raihan Bin, M., Md Morshed, A., Rahman, M. H., & Jang, Y. M. (2022). Real-Time Energy Data Acquisition, Anomaly Detection, and Monitoring System: Implementation of a Secured, Robust, and Integrated Global IIoT Infrastructure with Edge and Cloud AI. Sensors, 22(22), 8980. https://doi.org/https://doi.org/10.3390/s22228980
38. Remenarić, B., Kenfelja, I., & Mijoč, I. (2018). CREATIVE ACCOUNTING - MOTIVES, TECHNIQUES AND POSSIBILITIES OF PREVENTION [Kreativno raČunovodstvo - motivi, tehnike i moguĆnosti sprjeČavanja]. Ekonomski Vjesnik, 31(1), 193-199.
39. Rewicki, F., Denzler, J., & Niebling, J. (2023). Is It Worth It? Comparing Six Deep and Classical Methods for Unsupervised Anomaly Detection in Time Series. Applied Sciences, 13(3), 1778. https://doi.org/https://doi.org/10.3390/app13031778
40. Ribeiro, D., Luís Miguel, M., Moreira, G., Pilastri, A., & Cortez, P. (2022). Isolation Forests and Deep Autoencoders for Industrial Screw Tightening Anomaly Detection. Computers, 11(4), 54. https://doi.org/https://doi.org/10.3390/computers11040054
41. Saeed, W., & Omlin, C. (2023). Explainable AI (XAI): A systematic meta-survey of current challenges and future opportunities. Knowledge-Based Systems, 263, 110273.

42. Safta, I. L., Andrada-Ioana, S., & Muntean, N. (2021). Bibliometric Analysis of the Literature on Measuring Techniques for Manipulating Financial Statements. Risks, 9(7), 123. https://doi.org/https://doi.org/10.3390/risks9070123
43. Shobayo, O., Adeyemi-Longe, S., Popoola, O., & Ogunleye, B. (2024). Innovative Sentiment Analysis and Prediction of Stock Price Using FinBERT, GPT-4 and Logistic Regression: A Data-Driven Approach. Big Data and Cognitive Computing, 8(11), 143.
44. Škoda, M., Lengyelfalusy, T., & Gabrhelová, G. (2017). CREATIVE ACCOUNTING PRACTICIES IN SLOVAKIA AFTER PASSING FINANCIAL CRISIS. Copernican Journal of Finance and Accounting, 6(2), 71-86.
45. Suzuki, M., Sakaji, H., Hirano, M., & Izumi, K. (2023). Constructing and analyzing domain-specific language model for financial text mining. Information Processing & Management, 60(2), 103194.
46. Utkin, L., Ageev, A., Konstantinov, A., & Muliukha, V. (2023). Improved Anomaly Detection by Using the Attention-Based Isolation Forest. Algorithms, 16(1), 19. https://doi.org/https://doi.org/10.3390/a16010019
47. Wibawa, A. P., & Kurniawan, F. (2024). Advancements in Natural Language Processing: Implications, Challenges, and Future Directions. Telematics and Informatics Reports, 100173.
48. Xing, H.-J., & Li, L.-F. (2020). Robust least squares one-class support vector machine. Pattern Recognition Letters, 138, 571-578.