Unsupervised Anomaly Detection in Financial Fraud: A Concise Methodological Exploration

Pulkit Chouhan

***Abstract*—Financial fraud, particularly involving credit cards, inflicts billions in annual losses and erodes consumer trust, making its detection a critical concern. This task is complicated by extreme dataset class imbalance, the adaptive nature of fraudsters, and regulatory demands for model explainability. This report offers a focused exploration of unsupervised anomaly detection techniques for identifying fraudulent financial activities. These methods are vital for detecting novel fraud without reliance on often scarce labeled data. We define anomaly detection in machine learning and its specific application to financial fraud, including types of anomalies. A comparative overview of key unsupervised methods—statistical, distance-based, density-based, isolation-based, and autoencoder-based—is presented, followed by an analysis of core challenges in this domain. Finally, salient trends from literature and industry are summarized, aiming for a concise guide to this evolving field.**

***Index Terms*—anomaly detection, unsupervised learning, financial fraud, credit card fraud, machine learning.**

I. Introduction

Credit card fraud constitutes a persistent and significant threat within the global financial ecosystem, leading to substantial annual economic losses and diminishing the crucial element of customer trust [1]. The digital transformation of finance has exponentially increased transaction data volume, creating both new avenues for fraudulent activities and advanced opportunities for their detection [2]. Traditional detection systems often falter against the dynamic and sophisticated nature of modern fraud. Unsupervised anomaly detection techniques have, therefore, become indispensable. Their capacity to identify unusual patterns indicative of fraud without prior knowledge of specific fraud signatures—that is, without needing labeled data—makes them particularly effective for uncovering novel and emerging threats [3].

The landscape of fraud detection is fraught with challenges. A primary hurdle is the *extreme class imbalance* in transaction datasets, where fraudulent instances are a minuscule fraction (often ¡0.1%) of total transactions [4]. This biases supervised models towards legitimate transactions, leading to poor fraud detection. Furthermore, fraudsters exhibit *adversarial behavior*, continuously evolving their tactics to circumvent established detection mechanisms [5]. This necessitates adaptable models capable of identifying novel fraud patterns. The need for *real-time scoring* of high-volume transactions adds computational complexity [2]. Finally, the heavily regulated financial industry mandates *model explainability*, a requirement that complex "black-box" models often struggle to meet [6]. This report focuses on how unsupervised approaches address these concerns.

II. Understanding Anomaly Detection

A. Defining Anomaly Detection

In the realm of machine learning, anomaly detection is the process of identifying data points, events, or observations that deviate markedly from the established normal behavior within a dataset [7]. This process typically involves first establishing a baseline or profile of what constitutes "normal," often derived from historical or representative sample data [8]. Any instance that does not conform to this learned normality is flagged as an anomaly, which could manifest as an outlier, an unexpected change, or an error, signaling potential issues or threats that require attention.

Within the specific domain of financial fraud, anomaly detection refers to the application of statistical and machine learning techniques to identify unusual patterns or behaviors within financial datasets that may indicate potential fraudulent activity [9]. The fundamental aim is to distinguish between normal, legitimate transactions and abnormal, potentially malicious ones [10]. For instance, in credit card fraud, this involves pinpointing irregular activities that deviate from a customer's typical spending habits or established transaction patterns, thereby serving as a proactive measure for threat management and loss mitigation.

B. Types of Anomalies in Financial Data

Anomalies encountered in financial datasets can be broadly categorized, and understanding these types helps in selecting appropriate detection methods. **Point anomalies** are individual data instances that appear anomalous with respect to the rest of the data. An example is a single, exceptionally large transaction on a card usually used for small purchases. **Contextual anomalies** (or conditional anomalies) refer to data instances that are anomalous only within a specific context. For example, high spending on winter clothing is normal in December but would be anomalous in July for the same individual in the northern hemisphere. The transaction amount relative to time of day, location, or recent activity can provide such context. **Collective anomalies** involve a collection of related data instances that, as a group, are anomalous with respect to the entire dataset, even if individual instances within the group are not particularly unusual by themselves. A series of small, rapid, geographically dispersed transactions on a single card might form a collective anomaly. Recognizing these often requires analyzing relationships and patterns among multiple data points. Financial fraud can manifest as any of these types, necessitating a diverse array of detection strategies.

III. Unsupervised Anomaly Detection Methods

Unsupervised methods are particularly relevant for financial fraud detection because they do not require pre-labeled data and are thus capable of identifying novel or evolving fraud patterns. These methods are broadly categorized based on their underlying operational principles.

A. Statistical Methods

Statistical methods are foundational, establishing a model of normal data behavior and identifying deviations. They generally assume that normal data points occur in high-probability regions of a stochastic model, while anomalies occur in low-probability regions [11]. Algorithms like *Z-Score Analysis* measure how many standard deviations a data point is from the mean ($Z = (x - \mu)/\sigma$), flagging points beyond a threshold as anomalies. For time-series data common in finance, *Autoregressive Integrated Moving Average (ARIMA)* models capture temporal patterns, with large residuals (differences between predicted and actual values) indicating anomalies. These methods are generally simple to implement and interpret, especially where data distributions are well-understood. However, their reliance on distributional assumptions (e.g., normality for Z-scores) can lead to false positives if data is skewed or complex, and they may struggle with high-dimensional data or detecting contextual anomalies [11].

B. Distance-Based Methods

Distance-based methods quantify relationships between data points based on distance or similarity measures, defining anomalies as points isolated from the majority of the data. The core principle is that normal data points reside in dense neighborhoods, while anomalous points are located far from their closest neighbors [12]. The *k-Nearest Neighbors (k-NN)* algorithm determines a point's anomaly score by its distance to its k-th nearest neighbor or the average distance to its k nearest neighbors; points with significantly larger distances are flagged [13]. *Local Outlier Factor (LOF)* offers a more nuanced view by measuring the local density deviation of a data point relative to its neighbors, identifying points in sparser regions than their surroundings [14]. These methods generally do not make strong distributional assumptions [15], and LOF is particularly effective at detecting local anomalies [16]. However, both can be computationally expensive for large datasets and may suffer performance degradation in high-dimensional spaces due to the "curse of dimensionality," where distance measures become less meaningful [13].

C. Density-Based Methods

Density-based methods define anomalies as objects in regions of low data density, assuming normal data points belong to dense neighborhoods, while anomalies lie in sparse regions or do not belong to any significant cluster [17]. *DBSCAN (Density-Based Spatial Clustering of Applications with Noise)* groups closely packed points based on an ϵ-radius and a minimum number of points (MinPts) [18]. Points not belonging to any cluster are classified as noise and thus anomalies. DBSCAN can discover clusters of arbitrary shapes and is robust to noise without requiring the number of clusters to be predefined. However, its performance is highly sensitive to parameter tuning (ϵ, MinPts) and it can struggle with datasets of varying densities or in high-dimensional spaces [18]. In finance, it can identify isolated transactions or small clusters exhibiting unusual characteristics.

D. Isolation-Based Methods

Isolation-based methods operate on the principle that anomalies are "few and different," making them more susceptible to isolation than normal points. *Isolation Forest* builds an ensemble of "isolation trees" (iTrees) where data is recursively partitioned by randomly selecting features and split values [1]. Anomalous instances, being rare and distinct, tend to have shorter average path lengths from the root to the terminating leaf node across the ensemble of trees, thus receiving higher anomaly scores [13]. This method is computationally efficient, especially for large, high-dimensional datasets, as it does not rely on distance calculations and is robust to irrelevant features [19]. A potential limitation is its tendency to misclassify normal points near distribution edges as anomalies if they can be isolated quickly and its implicit assumption of attribute independence when creating splits [19]. It is well-suited for high-volume, real-time transaction screening [20].

E. Autoencoder-Based Methods

Autoencoders are a type of artificial neural network used for unsupervised learning, primarily for learning efficient data representations (codings). They consist of an encoder, which maps input data x to a lower-dimensional latent representation z, and a decoder, which attempts to reconstruct the original input $\hat{x}$ from z [21]. The network, $z = f(x) = \sigma(Wx + b)$ and $\hat{x} = g(z) = \sigma(W'z + b')$, is trained on normal (non-anomalous) data to minimize the reconstruction error $||x - \hat{x}||^2$. Anomalies are detected when the reconstruction error for a new data point is significantly higher than that observed for normal data during training, as the model struggles to reconstruct patterns it hasn't seen before [22]. Autoencoders can learn complex, non-linear patterns and automatically perform feature learning. However, training can be computationally intensive, and performance is sensitive to network architecture and hyperparameters. They are effective for detecting sophisticated fraud involving subtle deviations across multiple transaction attributes [21].

IV. Core Challenges in Financial Fraud Detection

Detecting financial fraud is an intricate problem domain, made more complex by several inherent and evolving challenges that require sophisticated and adaptive solutions.

Class Imbalance: As previously mentioned, financial transaction datasets exhibit extreme class imbalance, with fraudulent transactions being exceptionally rare [1], [4]. This scarcity makes it difficult for machine learning models to learn the characteristics of fraud without being overwhelmed by legitimate transactions, often leading to high false negatives. Traditional accuracy metrics are misleading; precision, recall,

F1-score, and AUC-PR are more appropriate for evaluating performance [4]. Techniques to mitigate this include data-level approaches like oversampling (e.g., SMOTE) or undersampling, and algorithmic-level approaches such as cost-sensitive learning or using ensemble methods. Unsupervised methods that model normality (like autoencoders) or isolate outliers directly (like Isolation Forest) are inherently less affected by this class imbalance during their primary learning phase.

Real-Time Scoring: The nature of financial transactions, especially credit card payments, demands real-time or near real-time fraud detection [2]. Delays in identifying fraudulent activity lead directly to financial losses and erode customer trust. This imposes stringent constraints on the computational efficiency and latency of detection models. Systems must process and score millions of transactions daily with minimal delay [23]. Challenges include handling high data velocity and volume, meeting low latency requirements (often milliseconds per transaction), ensuring system scalability, and managing the complexity of real-time feature engineering. Solutions often involve efficient algorithms like Isolation Forest, scalable distributed computing architectures, and sometimes tiered systems with initial fast screening.

Adversarial Behavior: Fraudsters are not static targets; they actively adapt their strategies to bypass existing detection mechanisms [5]. This adversarial behavior means that fraud patterns constantly evolve (concept drift), making historical data less representative of future fraud. Fraudsters devise new tactics, including sophisticated phishing, synthetic identity fraud, and account takeovers. Adversarial machine learning attacks, where fraudsters craft transactions to appear legitimate or try to poison training data, are a growing concern [2]. Addressing this requires adaptive learning systems (e.g., online learning), robust model architectures, continuous model performance monitoring with feedback loops from fraud analysts [24], and ongoing research into adversarial attack understanding and defense mechanisms such as adversarial training [25].

Model Explainability: In the heavily regulated financial industry, the ability to explain why a model makes a particular decision is often a requirement, not just a preference [6]. "Black-box" models like complex neural networks, while potentially accurate, may lack transparency, hindering the ability to understand their decision-making process. Explainability is crucial for regulatory compliance (e.g., GDPR's "right to explanation"), building trust with users and regulators, model debugging and improvement, and deriving actionable insights for fraud analysts [26]. Techniques from Explainable AI (XAI) like LIME and SHAP are increasingly applied. Explaining unsupervised models is unique, as it often focuses on why a data point is anomalous relative to the learned normality, rather than classification against a ground truth [27].

V. Key Literature and Approaches in Financial Services

The field of fraud detection in financial services is dynamic, with continuous evolution driven by academic research and industry innovation, showing a clear shift from traditional rule-based systems to more sophisticated machine learning.

Academic research trends reveal a strong focus on advanced machine learning, especially deep learning techniques [3]. *Deep Learning Models* such as Autoencoders (AEs), Variational Autoencoders (VAEs), Generative Adversarial Networks (GANs) for synthetic data generation and direct anomaly detection, Recurrent Neural Networks (RNNs like LSTM/GRU) for sequential data, Graph Neural Networks (GNNs) for relational data, and even Transformers are widely explored [4], [28], [29]. Recognizing that no single model is universally optimal, *Hybrid and Ensemble Methods* combining different algorithms are increasingly common to enhance accuracy and robustness [3]. Significant effort is also dedicated to tackling *Class Imbalance and Explainability*, integrating techniques like SMOTE and XAI methods (SHAP, LIME). Moreover, research addresses practical issues like data privacy through *Federated Learning* [6].

Industry best practices reflect these academic trends. Financial institutions are adopting multi-layered fraud detection strategies, often starting with simpler rule-based systems followed by sophisticated ML models [1]. *Hybrid and Ensemble Systems*, combining rule-based logic with supervised and unsupervised ML, are common. *Real-Time Monitoring and Analysis* of numerous transaction features are standard, enabling immediate detection [1]. Given fraud's adversarial nature, *Continuous Model Monitoring and Updating*, with feedback loops from investigators, is crucial [24]. While high detection is key, minimizing false positives is also critical for operational efficiency and customer experience. *Unsupervised Learning for Novel Threats* (using Isolation Forest, One-Class SVM, Autoencoders) is valued for identifying new, unseen fraud patterns [1]. Robust fraud detection also hinges on *Data Quality and Feature Engineering*. Finally, ensuring *Scalable Infrastructure* and meeting *Regulatory Compliance and Explainability* requirements, often through XAI tools, are paramount [30].

VI. Conclusion

Unsupervised anomaly detection provides a critical suite of techniques for addressing the persistent and complex challenge of financial fraud. Methodologies ranging from statistical models to advanced neural networks like autoencoders offer diverse tools for identifying transactions that deviate from established normal behavior, thereby signaling potential fraudulent activity without relying on predefined fraud labels [9]. This capability is invaluable for detecting novel threats in a rapidly evolving fraud landscape.

Despite the sophisticated toolkit, significant challenges persist. Extreme class imbalance in datasets, the stringent need for real-time scoring, the adaptive adversarial behavior of fraudsters, and the crucial requirement for model explainability in a regulated environment continue to test the limits of detection systems [2], [4]–[6]. Current literature and industry practices clearly indicate a trend towards advanced machine learning,

particularly deep learning and hybrid/ensemble approaches, to tackle these multifaceted complexities [3].

For financial institutions, this underscores the necessity of adopting robust, adaptable, and increasingly sophisticated multi-layered anomaly detection strategies where unsupervised learning plays a vital role in identifying novel threats. The choice of specific techniques must be carefully guided by data characteristics, operational constraints, and the types of fraud targeted. Addressing the inherent challenges through appropriate data handling, algorithmic design, and investment in explainable AI solutions is critical for effective and compliant fraud management. As fraud tactics continue to evolve, ongoing research and development in areas such as adversarial robustness, enhanced explainability for unsupervised models, effective concept drift handling, and scalable, privacy-preserving systems will be pivotal in strengthening the financial industry's defenses.

REFERENCES

[1] Number Analytics, "5 Key Anomaly Detection Techniques in Finance & Banking," Number Analytics Blog, Mar. 2025. [Online]. Available: https://www.numberanalytics.com/blog/5-key-anomaly-detection-techniques-finance-banking

[2] Prove, "Financial Fraud Detection Challenges and How to Solve Them," Prove Blog, Feb. 2025. [Online]. Available: https://www.prove.com/blog/financial-fraud-detection-challenges

[3] Y. Chen, et al., "Year-over-Year Developments in Financial Fraud Detection via Deep Learning: A Systematic Literature Review," arXiv:2502.00201, Feb. 2025.

[4] ResearchGate, "Addressing the Challenges of Imbalanced Datasets in AI-Driven Fraud Detection," Apr. 2025. (Placeholder for a specific relevant paper)

[5] MDPI Electronics, "Financial Fraud Detection Using Voted Perceptron Model," vol. 14, no. 9, p. 1875, May 2025.

[6] arXiv, "Financial Fraud Detection Using Explainable AI and Stacking Ensemble Methods," arXiv:2505.10050, May 2025.

[7] AWS, "What is Anomaly Detection?" [Online]. Available: https://aws.amazon.com/what-is/anomaly-detection/

[8] Elastic, "What is anomaly detection?" [Online]. Available: https://www.elastic.co/what-is/anomaly-detection

[9] Fraud.com, "Anomaly detection for fraud prevention - Advanced strategies." [Online]. Available: https://www.fraud.com/post/anomaly-detection

[10] Fraud.Net, "What is Anomaly Detection?" Fraud.Net Glossary. [Online]. Available: https://www.fraud.net/glossary/anomaly-detection

[11] MindBridge, "Anomaly Detection Techniques: How to Uncover Risks, Identify Patterns, and Strengthen Data Integrity," MindBridge Blog, Feb. 2025.

[12] GlassFlow, "What Is Anomaly Detection? - A Comprehensive Guide," GlassFlow Blog.

[13] K. Praxitelis, "Anomaly Detection Techniques Summary," Kaggle. [Online].

[14] Number Analytics, "Nonparametric Outlier Detection: A Comprehensive Guide," Number Analytics Blog, May 2025.

[15] Pickl.AI, "Unlocking the Power of KNN Algorithm in Machine Learning," Pickl.AI Blog, Mar. 2024.

[16] Activeloop AI, "Local Outlier Factor (LOF)," Activeloop AI Glossary.

[17] Number Analytics, "DBSCAN: 5 Essential Facts for Data Scientists," Number Analytics Blog, Mar. 2025.

[18] Number Analytics, "DBSCAN in Data Science: Clustering Techniques and Applications," Number Analytics Blog.

[19] Dremio, "Isolation Forest," Dremio Wiki.

[20] SciSpace, "Using Isolation Forest in Anomaly Detection: The Case of Credit Card Fraud Detection," 2018.

[21] Number Analytics, "Leveraging Autoencoder Techniques for Anomaly Detection," Number Analytics Blog, Mar. 2025.

[22] Fraud Detection Handbook, "Autoencoders for Fraud Detection."

[23] ResearchGate, "Assessing the Challenges of Implementing Real-Time Fraud Detection Solutions," Jan. 2025. (Placeholder)

[24] Number Analytics, "Anomaly Detection in Finance: 8 Key Strategies for Fraud Prevention," Number Analytics Blog.

[25] Journal of Engineering Research and Reports, "Adversarial Attacks and Defenses in AI-Driven Systems," Feb. 2025.

[26] M. Veen, "Integrating Causal Discovery with XAI for Trustworthy Fraud Detection in Finance," Univ. of Twente, Feb. 2025.

[27] PMC NCBI, "Explainable unsupervised anomaly detection for healthcare insurance data."

[28] arXiv, "Comparative Analysis of Contemporary Variational Autoencoder (VAE) Architectures for Anomaly Detection," arXiv:2408.13561, Aug. 2024.

[29] ResearchGate, "A Survey on Credit Card Fraud Detection using Deep Learning Model." (Placeholder)

[30] Florida Atlantic University News Desk, "FAU Develops Novel Machine Learning Method for Fraud Detection," Apr. 2024.