Annals of "Dunarea de Jos" University of Galati
Fascicle I. Economics and Applied Informatics
Years XXXI – nº3/2025
ISSN-L 1584-0409 ISSN-Online 2344-441X
www.eia.feaa.ugal.ro

DOI https://doi.org/10.35219/eai15840409561

Hybrid Detection of Anomalies in Financial Transactions: A Rule-Based and Machine Learning Approach

Alexandra Stavrositu (Caratas)★, Cristina Barbu (Antohi)★★, Mihaela-Carmen Muntean★★★, Dragos Sebastian Cristea★★★★, Daniela Ancuta Sarpe★★★★★

ARTICLE INFO

Article history:
Received September 29, 2025
Accepted November 22, 2025
Available online December 2025

JEL Classification
G21, G28, G45

Keywords:
anomaly detection, financial transactions, rule-based modeling, machine learning, Local Outlier Factor, One-Class SVM, autoencoder, hybrid framework, fraud detection, unsupervised learning, behavioral analytics, transaction monitoring, data-driven risk management, interpretability, outlier analysis

ABSTRACT

This paper presents a hybrid methodology for detecting anomalies in financial transactions including card-initiated transactions and payments by combining rule-based logic with unsupervised machine learning techniques. Rule-based detection leverages expert-defined heuristics to flag transactions exhibiting high-risk behaviors such as card number, BIN, transaction amount, local time, date, expiry, MCC, country code, 3DSecurity Level, time interval, count, amount and location, plus excessive login attempts, abnormal transaction timing, and demographic inconsistencies. In parallel, three unsupervised models—Local Outlier Factor, One-Class SVM, and Autoencoder—are applied to extract structural and statistical anomalies without requiring labeled data. A weighted scoring mechanism aggregates model outputs to rank suspicious transactions, enhancing robustness through model complementarity. The methodology is evaluated on a synthetically enriched transactional dataset, demonstrating its ability to identify both interpretable and latent anomalies. Comparative results highlight the benefits of model diversity and reveal limited but meaningful overlap between rule-based and ML-based detections. The proposed framework offers transparency, flexibility, and practical scalability, making it well-suited for near real-time monitoring systems in the banking sector. Findings underscore the importance of multi-layered detection in modern anti-fraud card and payment management.

1. Introduction

The proliferation of digital banking and near real time financial transactions including card-initiated transactions and payments has significantly increased the complexity and volume of transactional data processed by financial institutions. With this evolution comes a parallel rise in fraudulent activities, behavioral anomalies, and system-level risks, all of which pose substantial threats to financial integrity, customer trust, and regulatory compliance. In response, anomaly detection has emerged as a critical component of modern anti-fraud card and payment management systems.

Anomaly detection refers to the identification of patterns in data that do not conform to expected behavior. In the context of financial transactions, anomalies can signify a wide range of irregularities—from unauthorized access attempts, card usage and identity theft to system failure and human error. However, the detection of such anomalies is inherently challenging due to several factors: the lack of labeled anomalies for supervised learning, the high variability of legitimate behaviors across different user segments, and the dynamic nature of financial systems that evolve over time.

Traditional rule-based financial systems have long been the cornerstone of anomaly detection in banking. These systems leverage expert knowledge to define thresholds and heuristics that trigger alerts when predefined conditions are violated. While interpretable and easy to implement, such systems often struggle with adaptability and fail to capture subtle, context-dependent patterns. On the other hand, machine learning approaches, particularly unsupervised algorithms, offer a flexible and data-driven means of detecting unusual behavior without requiring labeled training data. These models are capable of learning the underlying structure of transactional patterns and identifying outliers based on deviation from learned norms. However, their black-box nature can hinder interpretability and lead to false positives without contextual grounding.

★, ★★, ★★★, ★★★★, ★★★★★Dunarea de Jos University of Galați, Romania. E-mail addresses: alexandracaratas@yahoo.com (A. Stavrositu Caratas – Corresponding author), cristina.antoghi@ugal.ro (C. Barbu Antohi), mihaela.muntean@ugal.ro (M. C. Muntean), dragos.cristea@ugal.ro (D. S. Cristea), daniela.sarpe@ugal.ro (D. A. Sarpe)

To address these limitations, this paper proposes a hybrid methodology that integrates rule-based logic with unsupervised machine learning for anomaly detection in financial banking transactions. The framework begins with domain-informed rules that flag transactions based on expert-defined criteria such as high-value deviations, irregular login behavior, night-time activity, and temporal inconsistencies. This rule-based layer ensures interpretability and immediate operational value. Building upon this, three unsupervised models—Local Outlier Factor (LOF), One-Class Support Vector Machine (SVM), and Autoencoder—are applied to a curated feature set to capture outliers from structural, statistical, and latent behavioral perspectives. The combined system offers a multi-layered detection mechanism that leverages the precision and explainability of rule-based models and the adaptability and depth of machine learning. The results are aggregated through a weighted scoring system to synthesize model outputs and rank suspicious transactions. The framework is evaluated on a synthetically generated but behaviorally realistic financial dataset, with detailed analyses of model contributions, overlaps, and representative anomalies.

The paper demonstrates that, applying this approach, hybrid architecture not only improves anomaly detection performance but also offers operational insights that are essential for monitoring and compliance purposes. The methodology is flexible, interpretable, and extensible, and it serves as a scalable blueprint for real-world deployment in financial anomaly monitoring anti-fraud systems.

2. Literature review

In the study published by Khatri, M. R. (2024) in the International Journal of Applied Machine Learning and Computational Intelligence, the author explores the combination of rule-based systems with anomaly detection techniques to enhance fraud detection in payments. The proposed approach leverages domain knowledge and expert-defined rules alongside the ability to identify new fraud patterns through anomaly detection, aiming to improve the accuracy and adaptability of fraud detection models.

By embracing a hybrid approach, we can build a more resilient and proactive defense against fraudulent activities, fostering trust and confidence in the digital economy.[1]

In the article "An Anomaly Prediction Framework for Financial IT Systems Using Hybrid Machine Learning Methods", the authors propose a framework for anomaly prediction in financial IT systems by combining machine learning methods. The methodology includes preprocessing of data from system logs, followed by the use of a module for predicting key performance indicators, a module for anomaly detection, and a module for classifying their severity.[2]

In their work "Credit Card Fraud Detection through Parenclitic Network Analysis", Zanin et al. (2017) introduce a hybrid approach that combines complex network analysis with machine learning techniques for detecting fraud in credit card transactions. By constructing parenclitic networks, the authors extract features that reflect deviations from normal behavior, thereby improving the accuracy of anomaly detection.[3]

Hans-peter Kriegel, M. Breunig, Raymond T. Ng and Jörg Sander implemented the Local Outlier Factor (LOF) algorithm for finding abnormal data points by evaluating the local variability of a given data point in comparison to its neighbors. Local-density outliers are observed with this algorithm [4]. Locality is defined by nearest neighbors and distance is used to measure density. When matching an object's local density with its neighbors ' local densities, one can distinguish areas with similar consistency, and points that have a substantially lower density than their neighbors. The data point is called an outlier because opposed to its surroundings it has very low scale. External trends can be classified into 2 sorts: global outliers and local outliers. The entity which has a considerable distance from its k-th neighbor is called Global ooutlier while as an entity whereas a local outlier has a distance from its k-th neighbor which is large compared to its neighbors' average distance from its own k-th closest neighbors.

Diego Vallarino notes in his work “Detecting Financial Fraud with Hybrid Deep Learning: A Mix-of-Experts Approach to Sequential and Anomalous Patterns” that financial fraud detection remains a critical challenge due to the dynamic and adversarial nature of fraudulent behavior. As fraudsters evolve their tactics, detection systems must combine robustness, adaptability, and precision. In this study, the author proposes a hybrid architecture for detecting financial fraud, combining a Mixture of Experts (MoE) framework with Recurrent Neural Networks (RNNs), Transformer encoders, and Autoencoders. Each expert module contributes with specialized capabilities: RNNs capture sequential behavior, Transformers extract high-order feature interactions, and Autoencoders detect anomalies through reconstruction loss. The MoE framework dynamically assigns predictive responsibility among experts, enabling context-aware and adaptive decision-making. The hybrid model achieved an accuracy of 98.7%, a precision of 94.3%, and a recall of 91.5%, outperforming individual models and traditional machine learning baselines.[5]

In the paper "Isolation Forest and Local Outlier Factor for Credit Card Fraud Detection System", co-authors V. Vijayakumar, Nallam Sri Divya, P. Sarojini, and K. Sonika investigate the use of the Isolation Forest and Local Outlier Factor (LOF) algorithms for detecting fraud in credit card transactions. The authors analyze the performance of both methods in identifying fraudulent transactions, highlighting the advantages and limitations of each approach. The study emphasizes the importance of real-time anomaly detection in preventing financial losses.[6]

A study published in the Indonesian Journal of Computing and Cybernetics Systems applied the LOF (Local Outlier Factor) algorithm to identify suspicious transactions in credit card data. The outlier analysis method is used to build the knowledge with a local outlier factor algorithm that has high accuracy, recall, and precision results and can be used in multivariate data.

Using a dataset of 1,803 transactions from five customers, the researchers achieved an average accuracy of 96%, outperforming other methods such as INFLO (84%) and AFV (77%). This demonstrates the efficiency of the LOF algorithm in detecting anomalies in multivariate financial data.[7]

Another study, "Anomaly Detection in Networks with Application to Financial Transaction Networks", focuses on detecting anomalies in financial transaction networks, where accounts are represented as nodes and transactions as weighted edges. The authors propose a method that combines features from spectral analysis and local statistics to identify anomalies such as long paths of high-value transactions or cliques of accounts. Although it does not directly use LOF, the approach emphasizes the importance of detecting local anomalies—an aspect that is central to the LOF algorithm. [8]

In 2023, a new article on the same topic was published, with the main goal of understanding and applying an alternative approach for classifying fraudulent transactions using the Isolation Forest and Local Outlier Factor (LOF) algorithms instead of the traditional Random Forest method. The study aims to identify fraudulent transactions with higher accuracy and to compare these methods in order to find a more optimal solution.

Fraud detection involves analyzing the past behavior of credit card transactions to identify patterns associated with fraud. The model is then used to assess whether a new transaction is fraudulent or not.

The ultimate objective is to detect 100% of fraudulent transactions while simultaneously reducing the number of incorrect classifications (i.e., legitimate transactions mistakenly flagged as fraud).[9]

Another study published in 2022 proposed the use of two unsupervised machine learning algorithms—Isolation Forest (IF) and Local Outlier Factor (LOF)—for detecting fraudulent credit card transactions. These algorithms were trained and tested on a dataset containing 4,092 transactions made by European cardholders, 80% of which were fraudulent. IF works by isolating anomalies through decision trees, while LOF assesses the local density of data to identify outliers. Compared to other existing models, these methods achieved a prediction accuracy of 99%, demonstrating their efficiency in rapidly and accurately detecting fraud. This can support both banks and customers in preventing financial losses.[10]

The paper titled "Enhancing Payment Security Through AI-Driven Anomaly Detection and Predictive Analytics" provides a detailed analysis of system architectures and algorithmic approaches—such as Isolation Forest, Local Outlier Factor (LOF), and DBSCAN—for implementing anomaly detection, along with predictive analytics techniques like Autoencoders, Support Vector Machines (SVM), Random Cut Forest, and Gaussian Mixture Models (GMM), all based on artificial intelligence.

The study offers an in-depth examination of the key features, advantages, and limitations of various algorithms in each category. It also explores how these approaches handle factors such as data dimensionality, computational efficiency, and robustness to anomalies and noise. Additionally, the research discusses the use of predictive analytics techniques—including statistical models, instance-based learning, and ensemble methods—for applications such as fraud detection and forecasting.

The paper highlights the trade-offs that must be considered when selecting appropriate approaches for anomaly detection and predictive modeling, offering valuable insights for researchers developing AI-based payment security systems for a variety of infrastructural and social applications.[11]

3. Methodology

This study introduces a hybrid approach for detecting anomalies in financial transactions, such as card initiated transactions and payments, integrating rule-based logic taking in to account anti-fraud scenarios data like, card number, BIN, transaction amount, local time, date, expiry, MCC, country code, 3DSecurity Level, time interval, count, amount and location with unsupervised machine learning techniques and time-series decomposition. By leveraging real-time data and contextual analysis, such applications can proactively detect and mitigate fraudulent behavior using pre-configured or dynamically generated anti-fraud scenarios. These scenarios are built around key transaction attributes such as card number, BIN, transaction amount, local time, date, expiry date, MCC (Merchant Category Code), country code, 3D Secure level, time intervals, transaction counts, amounts, and locations.

One of the most common fraud detection logics is based on geolocation and time analysis. For instance, if a card is used in Bucharest and then, within a few minutes, another transaction is attempted in New York, the application can automatically flag this as high-risk based on the time interval and country code. A well-structured fraud engine would calculate the physical impossibility of such movement and either block the second transaction or escalate it for 3D Secure step-up authentication.

Another high-frequency scenario involves transaction velocity monitoring. Let's say the same card is used more than five times within a 60-second interval, each time for small amounts. The application, through a rule based on count, amount, and MCC, may identify this as a card testing pattern — a common technique used by fraudsters to validate stolen card credentials before executing larger fraud attempts.

The BIN and card number can also provide early signals. Some fraud engines apply rules specific to high-risk BINs — such as those from regions with high fraud rates, or recently compromised issuers. Combined with an unusual MCC, such as luxury goods or online gambling, the scenario can be configured to block or flag the transaction even before completion.

The expiry date offers another layer. If the card is close to expiry and suddenly sees a spike in usage, particularly for cross-border transactions, it may suggest that the card was stolen and is being used before expiration. When paired with a low or missing 3D Secure level, the application can treat this as a high-risk indicator.

Smart fraud logic also uses behavioral profiling over time. For example, if a card that is usually used in Romania for groceries or gas stations suddenly triggers a transaction in Hong Kong at a luxury jewelry merchant, the application will compare this against past usage patterns and evaluate whether this transaction aligns with historical location, MCC, and amount.

Behind the scenes, the application tracks these events through a combination of real-time data streams and historic behavior logs. Triggers can be defined in simple if/then rules or enhanced using machine learning models that compute dynamic thresholds. Transactions can be flagged, temporarily held, or rejected outright depending on the severity of the risk score.

In conclusion, a well-designed transaction management application serves as both a control center and a fraud shield — constantly analyzing count, amount, location, time, and security context to protect cardholders and issuers alike. Fraud detection isn't just a layer — it's an embedded intelligence at the core of modern card processing platforms.

The methodology involves main stages: (1) data preprocessing and feature engineering, (2) multi-layered anomaly detection, and (3) aggregation and interpretation of anomaly scores.

The dataset comprises timestamped transactional records with demographic, behavioral, and technical attributes. Thus, the dataset contains 2512 transaction records associated with 495 unique customer accounts. Each record includes attributes that reflect transactional, behavioral, demographic, and technical dimensions. As numeric fields we can mention TransactionAmount, which has a mean of approximately 297.59 units, and TransactionDuration, which averages 119.64 seconds. The data also captures merchant information (100 unique IDs) and access channels, with "Branch", "Online", and "Mobile" being the primary modes—Branch being the most common at 868 instances. Four customer occupation categories are present, with "Student" being the most prevalent, representing 657 transactions. Each transaction includes both a TransactionDate and a PreviousTransactionDate, enabling time-differential computations. This dual timestamp structure can be used for deriving features related to transaction frequency and behavioral insights. The descriptive statistics confirm the dataset's variability and depth, making it suitable for layered anomaly detection models that depend on both high-cardinality identifiers (e.g., DeviceID, IP. Address) and continuous behavioral metrics (e.g., amount, duration, login attempts). The data was synthetically generated to simulate realistic banking activity, preserving statistical and behavioral properties observed in real-world financial systems while avoiding exposure of sensitive or proprietary information.

Before modeling, a set of preprocessing steps was applied to ensure consistency and analytical readiness: a) All timestamp fields were parsed into a standardized POSIXct format, and transactions were sorted chronologically within each account to facilitate sequential analysis, b) Character-type variables were transformed into factors where appropriate, c) numerical fields were assessed for scale heterogeneity, and selected attributes were standardized to zero mean and unit variance to ensure compatibility with distance- and projection-based machine learning algorithms.

Any records with missing values in critical fields were removed to avoid imputation biases in anomaly scoring. In addition to the original variables, several features were engineered to enrich the dataset with temporal and behavioral insights. The time difference between successive transactions for a given account (TimeDiff) was computed in minutes. A binary flag (IsNight) was created to capture whether a transaction occurred during atypical hours, defined as between 10:00 PM and 6:00 AM. Demographic conditions were encoded to reflect contextually unusual activity, such as high transaction amounts by students or online channel use by elderly customers. Other engineered features include indicators for device and IP address change, login anomalies, and extreme transaction durations. These derived variables collectively support the rule-based anomaly detection layer and serve as inputs or filters for downstream modeling.

In the initial phase, the data was also pre-processed to standardize formats and derive additional temporal features. Timestamps were converted to a uniform format, and transaction records were sorted chronologically within each account. Based on this ordering, the time elapsed between consecutive transactions was computed

The first layer of detection used rule-based heuristics, based on domain knowledge. These rules were designed to flag potentially suspicious behavior based on statistical outliers, usage patterns, or contextual inconsistencies. For instance, anomalous transaction amounts are identified using a 3-standard-deviation threshold within each card transaction history. Device- and IP-based anomalies were flagged when changes occur relative to the account's initial usage. Temporal inconsistencies, such as extremely short or long gaps between transactions, were also treated as indicators of abnormal behavior. Demographic context is

incorporated through custom condition like unusually high transaction values for students, or the use of online banking channels by customers above the age of seventy. Behavioral anomalies are assessed based on card transaction frequency and card transaction duration, with thresholds set to capture excessive access attempts or atypical processing times. Each of these rules contributes a binary anomaly flag, and their sum constitutes a cumulative rule-based anomaly score for each transaction.

To enhance this deterministic layer, a second tier employs unsupervised learning methods aimed at capturing subtle and non-linear deviations in behavioral patterns. Three distinct models were applied: Local Outlier Factor (LOF), One-Class Support Vector Machine (SVM), and a neural autoencoder. The input feature set used for all models included transaction amount, duration, login attempts, account balance, and customer age, all of which are standardized prior to modeling. LOF detects local density deviations, marking transactions that occur in sparse neighborhoods. The One-Class SVM is trained to describe the boundary of normal transactions, flagging those that fall outside this boundary as outliers. Meanwhile, the autoencoder is trained to reconstruct input data; transactions with high reconstruction errors are interpreted as anomalous due to their poor fit to the learned data manifold. The outputs of the three models are rescaled to a common [0, 1] range to facilitate aggregation.

Two strategies were employed to synthesize a final anomaly score: a simple unweighted average and a customizable weighted sum. In the latter, relative importance is assigned to each model (default weights: 30% LOF, 30% SVM, 40% Autoencoder), allowing users to tune the scoring scheme based on operational constraints or empirical performance. Transactions were then ranked by their final anomaly score, and the top-ranked entries were highlighted for further investigation. As a final step, a macro-level anomaly detection layer was introduced using time-series decomposition. Transaction amounts were aggregated at daily granularity and decomposed into trend, seasonal, and remainder components using Seasonal-Trend decomposition via Loess (STL). Anomalies were detected in the residual component using the interquartile range (IQR) method, enabling the identification of days with unusual aggregate behavior, such as coordinated fraudulent activity or systemic events.

4. Results and discussions

The evaluated rules included unusually high login frequency, extreme inter-transaction timing, outlier transaction amounts, night-time activity, and extreme duration. Among these, only the **LoginAnomaly** and **TimeDiffAnomaly** rules yielded distinct transaction-level anomalies that met the detection threshold in isolation. No individual transactions were exclusively identified as anomalous by the AmountAnomaly, NightAnomaly, or DurationAnomaly rules under the current thresholds, though these rules contributed jointly to the total anomaly score in some cases. For example, two transactions were flagged due to excessive login attempts, a strong behavioral indicator often associated with suspicious activity such as brute-force credential testing or repeated access failures.

- **Transaction 1** involved a payment of 733.29 units with a transaction duration of 94 seconds. The account had 5 login attempts recorded before the transaction was authorized. The customer, aged 52, operated on an account with a substantial balance of 10,427.00 units. The frequency of login attempts exceeds the normal behavioral profile, suggesting either unusual user behavior or potential automated activity.
- **Transaction 2** was a smaller transaction of 331.32 units and lasted 137 seconds, also preceded by 5 login attempts. The customer, aged 61, had a relatively high account balance of 13,054.54 units. Again, the elevated number of logins implies authentication issues or a probing attempt, which may warrant further scrutiny from a cybersecurity standpoint.

In both cases, the login anomaly flag was triggered based on a threshold of more than three login attempts prior to transaction approval. The temporal irregularities were detected based on the calculated time differences between consecutive transactions from the same account. Transactions with very short or very long intervals were treated as suspect due to the implausibility of rapid succession or prolonged inactivity in certain usage contexts. For example, a transaction amounted to 212.97 units and lasted 178 seconds. The transaction was approved by a customer aged 59 with a balance of 4,180.40 units. The transaction followed an unusually short or long delay relative to the previous one on the same account, suggesting possible automated scheduling or out-of-pattern activity. Another transaction had a value of 476.99 units and a duration of 187 seconds. The customer, aged 23, had an account balance of 1,154.48 units. The time irregularity again signaled behavioral deviation from the account's typical timing sequence. Both transactions were flagged by the TimeDiffAnomaly rule, which marks transactions where the time gap is either less than one minute or exceeds 1,440 minutes (i.e., one day) between two consecutive transactions on the same account. While the dataset included rule logic for high transaction amount deviation (AmountAnomaly), night-time execution (NightAnomaly), and extreme transaction duration (DurationAnomaly), no transactions were flagged by these. This suggests that either the thresholds were appropriately conservative or that such behaviors, while potentially contributing to a compound anomaly profile, did not occur. These findings reinforce the value of rule-based systems for capturing specific, interpretable signals of deviant behavior. At the same time, the

limited detection of isolated anomalies suggests the potential need for threshold calibration or complementary detection methods, such as machine learning methods that can capture more nuanced behaviors.

The Local Outlier Factor (LOF) identified observations residing in low-density neighborhoods compared to their neighbors. The algorithm was sensitive to local patterns and was able to flag subtle structural deviations. For example: **Transaction TX000533** involved a moderate value of 120.24 units, with a transaction duration of 131 seconds, conducted by an 18-year-old user with a low account balance of 1,943.13. The LOF score reached its maximum (1.0), indicating this transaction occurred in a locally sparse region of the feature space despite having normal login attempts and channel usage. Another example was **Transaction TX002493**, worth 267.48 units, was similarly flagged (LOF = 0.915) due to low account balance (811.86), young customer age (21), and a relatively long transaction duration of 159 seconds, potentially marking it as atypical behavior in its local context.

The One-Class Support Vector Machine (SVM) defines a boundary around normal data and marks any points falling outside as outliers. It is effective in capturing global boundary violations. Still, as a method it was excessively pessimistic, identifying 2512 transactions as being suspect. As examples we have Transaction TX000062 with a high duration (227 seconds), that occurred at a branch and was performed by a 79-year-old customer. The One-Class SVM classified it as an anomaly likely due to the extended duration and customer demographic being distant from typical boundary conditions. Also, transaction TX001168 (838.57 units, duration of 273 seconds) was executed at an ATM with high account balance (14,576.47), again suggesting a boundary violation due to the combination of high transaction size and unusually long duration.

The autoencoder was trained to reconstruct normal transactions. Anomalies were identified when the reconstruction error was high, indicating that the input pattern was not well-represented by the training data's structure. Transaction like **Transaction TX000275**, with a transaction amount of 1,176.28 and 5 login attempts, was associated with the highest reconstruction error (1.0). The low account balance (323.69) and high login frequency likely contributed to poor encoding in the latent representation. Another suspect **Transaction TX001214** (1,192.20 units) also showed a high reconstruction error (0.918). Though its duration (103 seconds) and account balance (7,816.41) were not extreme in isolation, the overall pattern diverged significantly from expected latent encodings, leading to its classification as suspect.

The Weighted Anomaly Score composite model aggregated the rescaled outputs of all three models, assigning greater weight to autoencoder deviations allowing to capture consensus anomalies that are supported across dimensions. For example, transaction TX000275 (also flagged by the autoencoder) emerged as the top anomaly with a weighted score of 0.746. Its unique combination of high login attempts, large transaction value, and poor autoencoder fit resulted in a robust anomaly signal. Also, transaction TX001214, the second highest by weighted score (0.711), reinforces the autoencoder's detection while receiving marginal support from LOF and SVM models, further validating its classification as an outlier.

The following visualizations provide an intuitive overview of model performance and agreement patterns.

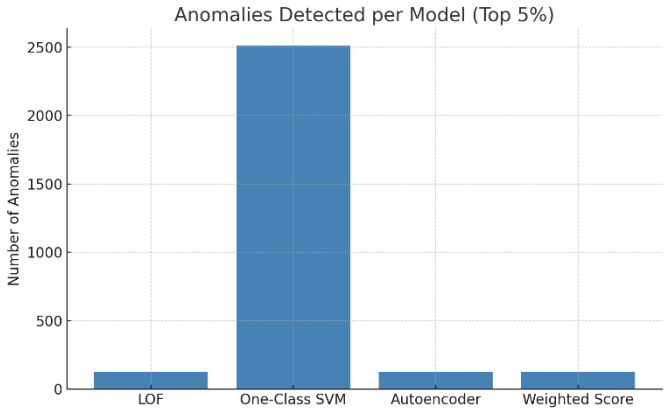

Figure 1. Number of Anomalies Detected by Each Model

Source: created by authors

This bar chart illustrates that One-Class SVM flagged many transactions as anomalies, indicating a highly sensitive configuration which, in a real anti-fraud scenario should be fine-tuned or rejected as a method. In contrast, LOF and Autoencoder models were more selective, each identifying 126 anomalies, which is consistent with a 5% threshold on the dataset size. The Venn diagram visualizes the intersection of anomaly detections across the three machine learning models.

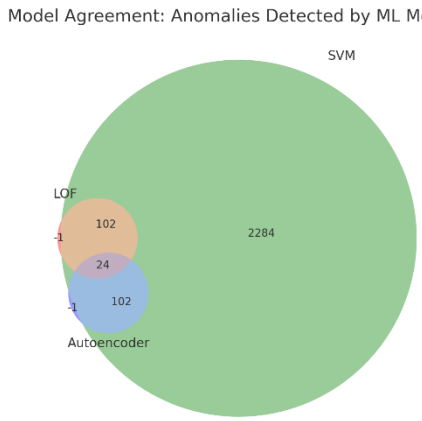

Figure 2. Overlap of Anomalies Detected by LOF, SVM, and Autoencoder
Source: created by authors

As it can be noticed, while there is some overlap between all combinations of models, only a small subset of 24 transactions was identified as anomalous by all three—highlighting the complementarity of detection strategies and the strength of triangulation.

Triangulation was used to strengthen the reliability of anomaly detection and to reduce false positives. It means a transaction was considered highly suspicious only if it was independently flagged by multiple, diverse detection mechanisms. Specifically, the transactions that were simultaneously identified as anomalies by all three unsupervised machine learning models—Local Outlier Factor (LOF), One-Class Support Vector Machine (SVM), and Autoencoder—and were also confirmed through multiple rule-based anomaly indicators. This triangulated validation approach ensures that each flagged transaction is not only statistically distinct within the data distribution but also contextually deviant according to domain-informed logic.

Thus, in the dataset there were **24 high-confidence anomalies** identified by **all three machine learning models** (LOF, One-Class SVM, Autoencoder) and confirmed through multiple rule-based flags.

1) The table below includes 24 transactions flagged as anomalies by Local Outlier Factor (LOF), One-Class SVM, and Autoencoder, and confirmed through at least one rule-based anomaly. Each entry lists the relevant transaction attributes

Table 1. Anomalies detected by all machine learning algorithms

TID	Amount	Duration	LogAttempts	Balance	Age	Group
TX000899	1531.31	62	4	859.86	18	Demographic + Login anomalies
TX001985	1512.99	50	1	9963.39	19	Demographic anomaly
TX002415	1664.33	65	1	1588.31	18	IP change + Irregular timing
TX000191	1422.55	165	1	5674.32	79	IP change + Irregular timing
TX000147	973.39	296	1	2042.22	77	IP change + Irregular timing
TX001635	1762.28	150	1	1380.34	24	IP change + Irregular timing
TX001673	1353.93	171	1	7858.41	26	IP change + Irregular timing
TX000354	432.63	137	1	13648.36	29	IP change + Irregular timing
TX001789	1612.37	173	1	3910.59	65	IP change + Irregular timing
TX000395	6.3	283	5	7697.68	80	IP change + Irregular timing
TX000376	1392.54	297	1	13347.69	41	IP change + Irregular timing
TX002216	1413.24	199	1	7272.01	58	IP change + Irregular timing
TX002273	1454.52	20	1	1189.03	27	IP change + Irregular timing
TX001852	167.92	136	2	14321.73	40	IP change + Irregular timing
TX001439	1831.02	83	1	11498.01	62	IP change + Irregular timing
TX001214	1192.2	103	5	7816.41	60	IP change + Irregular timing
TX001797	1135.8	250	1	13534.3	30	IP change + Irregular timing
TX000142	1049.92	21	1	2037.85	80	IP change + Irregular timing
TX002404	1493.0	151	1	1619.16	69	IP change + Irregular timing
TX002150	1250.94	107	2	11565.97	36	IP change + Irregular timing
TX000341	1830.0	238	1	2235.7	55	IP change + Irregular timing
TX001248	1647.74	52	1	1483.84	67	IP change + Irregular timing
TX000654	1919.11	161	1	11090.24	30	IP change + Irregular timing
TX000275	1176.28	174	5	323.69	54	IP change + Irregular timing

The majority of suspect transactions—**22 out of 24**—belong to a group marked by a consistent triad of rule-based triggers: **device change**, **IP address change**, and **abnormal timing** between consecutive

transactions. These indicators suggest suspicious access patterns, such as potential account takeover, session hijacking, or scripted automation. Despite varying customer demographics, transaction amounts, and durations, all transactions in this group show environmental instability—indicative of transactions not initiated under regular, trusted circumstances. Their simultaneous classification as anomalies by all ML models further emphasizes the statistical distinctiveness of their behavioral patterns.

One transaction stood out due to a combination of **demographic inconsistency** (e.g., high-value transaction by a user outside expected behavior norms for their profile) and **excessive login attempts**. This pairing suggests a possible scenario involving either identity misuse or an unusually persistent login behavior for a user type statistically not prone to such financial activity. While it did not involve environmental changes like IP or device shifts, the transaction's behavioral signals were strong enough to be flagged across all detection layers. The final transaction was flagged primarily due to a **demographic anomaly**—a behavioral mismatch between the customer's profile (age, occupation, channel) and the transaction's amount or method of execution. It did not exhibit anomalies in login behavior or timing, nor environmental inconsistencies, yet still stood out to all ML models. This highlights the capacity of unsupervised models to detect latent, subtle behavioral outliers even in the absence of multiple surface-level rule violations.

5. Conclusions

The proposed study and implemented a hybrid framework for detecting anomalies in financial transaction data, payments and card-initiated ones, leveraging both deterministic, rule-based logic and multiple unsupervised machine learning models. The approach aimed to balance interpretability and flexibility, enabling robust identification of suspicious behavior while maintaining transparency in logic-driven detections.

The rule-based system designed around five heuristic conditions — including transaction amount deviation, login irregularities, temporal anomalies, and night-time activity — successfully flagged a subset of financial transactions exhibiting domain-specific red flags. In particular, the LoginAnomaly and TimeDiffAnomaly rules captured behaviors consistent with common fraud indicators, such as repeated access attempts and irregular transaction pacing. However, other rules such as AmountAnomaly and NightAnomaly contributed more effectively in combination than as isolated triggers. This underscores the importance of composite scoring when relying solely on expert-defined criteria.

On the machine learning side, three unsupervised models — Local Outlier Factor (LOF), One-Class SVM, and Autoencoder — each contributed unique perspectives on behavioral outliers. LOF was sensitive to localized deviations in density, One-Class SVM captured global boundary violations, and the Autoencoder detected latent reconstruction mismatches. The weighted aggregation of these models further improved robustness by consolidating agreement across diverse detection paradigms. While each model produced partially overlapping anomaly sets, their intersection revealed a small cohort of highly suspicious transactions, many of which also aligned with rule-based findings.

Overall, the hybrid system proved effective in identifying both interpretable anomalies (e.g., flagged by clear rule violations) and emergent anomalies that would have been difficult to specify a priori. The layered architecture offers both operational transparency and analytic depth, making it suitable for real-world monitoring systems in financial services. Future enhancements may include supervised validation, drift monitoring, or reinforcement of rules through explainable AI techniques to refine the detection boundaries and maintain long-term model integrity.

While the proposed hybrid framework offers a robust approach to anomaly detection in financial transactions, several limitations should be acknowledged to contextualize the findings and guide future research and deployment efforts. First, the dataset used for experimentation, although structurally and behaviorally realistic, was synthetically generated. As such, it may not fully reflect the nuanced dynamics, noise, and irregularities found in real-world banking anti-fraud systems. The absence of labeled anomalies further limits the ability to quantitatively assess the precision and recall of the detection models, requiring the evaluation to rely on heuristic and interpretive validation rather than ground-truth benchmarking. Second, the rule-based component, while interpretable and domain-informed, is static in nature and requires manual updates to remain effective in evolving threat landscapes. Thresholds and conditions, such as the definition of "night-time" or acceptable transaction gaps, may not generalize across institutions, regions, or user populations without customization. Third, the machine learning models employed—Local Outlier Factor, One-Class SVM, and Autoencoder—were each trained on a uniform feature set and applied without model-specific tuning for segmentation (e.g., by user type, location, or channel). Additionally, the models assume stationarity in behavior over time, which may not hold in dynamic environments where concept drift is common. Finally, the ensemble anomaly score used a fixed weighting scheme, which, while practical, was not optimized or calibrated against operational metrics such as alert fatigue, fraud detection accuracy, or resource allocation in investigation workflows. This could lead to imbalanced influence among models or to under/over-representation of certain anomaly types. Despite these limitations, the study demonstrates the value of combining rule-based logic and unsupervised learning for identifying suspicious activity. Future work could

incorporate adaptive learning, feedback loops, real-time data, and labeled validation to improve detection performance and operational alignment.

References

1. *Khatri, M. R. (2024). Combining the strengths of rule-based and anomaly detection techniques for robust and comprehensive payment fraud detection. International Journal of Applied Machine Learning and Computational Intelligence, 14(4), 11–20. https://neuralslate.com/index.php/Machine-Learning-Computational-I/article/view/109*
2. *[1907.12778] An anomaly prediction framework for financial IT systems using hybrid machine learning methods, [Submitted on 30 Jul 2019 (v1), last revised 19 Dec 2019 (this version, v3), Jingwen Wang, Jingxin Liu, Juntao Pu, Qinghong Yang, Zhongchen Miao, Jian Gao, You Song]*
3. *[1706.01953] Credit card fraud detection through parenclitic network analysis, [submitted on 22 May 2017, Massimiliano Zanin, Miguel Romance, Santiago Moral, Regino Criado]*
4. *" Local outlier factor", En.wikipedia.org, 2019, Online]. https://en.wikipedia.org/wiki/Local_outlier_fact or.[Accessed: 06- May- 2019].*
5. *Diego Vallarino Dr, [2504.03750] Detecting Financial Fraud with Hybrid Deep Learning: A Mix-of-Experts Approach to Sequential and Anomalous Patterns, April 2025.*
6. *Isolation_Forest_and_Local_Outlier_Factor_for_Credit_Card_Fraud_Detection_System, V. Vijayakumar, Nallam Sri Divya, P. Sarojini, K. Sonika, Published By: Blue Eyes Intelligence Engineering & Sciences Publication, March 2020 https://www.researchgate.net/publication/341642761*
7. *Outlier Detection Credit Card Transactions Using Local Outlier Factor Algorithm (LOF), Sugidamayatno, IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*
8. *[1901.00402] Anomaly Detection in Networks with Application to Financial Transaction Networks, Andrew Elliott, Mihai Cucuringu, Milton Martinez Luaces, Paul Reidy, Gesine Reinert, [Submitted on 2 Jan 2019 (v1), last revised 24 May 2019 (this version, v2)]*
9. *Jain, A., Arora, M., Mehra, A., & Munshi, A. (2021). Anomaly Detection Algorithms in Financial Data. International Journal of Engineering and Advanced Technology, 10(5), 76–78. https://doi.org/10.35940/ijeat.e2598.0610521*
10. *Early Prediction of Credit Card Transaction Using Local Outlier Factor and Isolation Forest Tree Machine Learning Algorithms, Arjun K. P.Atlas G.[...]Arvindhan M.Lecture Notes in Mechanical Engineering (2022)*
11. *Agrawal, S. (2022). Enhancing Payment Security Through AI-Driven Anomaly Detection and Predictive Analytics. International Journal of Sustainable Infrastructure for Cities and Societies, 7(2), 1–14. Retrieved from https://vectoral.org/index.php/IJSICS/article/view/99*