Equivalent: Jurnal Ilmiah Sosial Teknik
Volume 8, Issue 3, 994-1004
e_ISSN: 2775-0833
https://jurnalequivalent.id/index.php/jequi/index
DOI: doi.org/10.59261/jequi.v8i3.360

A Theoretical Framework and 38-Feature Taxonomy for Behavioral Fraud Detection in Digital Banking

Ardi Darmawan*
Universitas Bina Nusantara, Indonesia

Thoyyibah T
Universitas Bina Nusantara, Indonesia

***Corresponding author:**
Ardi Darmawan, Universitas Bina Nusantara, Indonesia.
✉ardi.darmawan@binus.ac.id

Article Info:

Article history:
Received: May 20, 2026
Revised: June 30, 2026
Accepted: July 15, 2026

Keywords:
Fraud Detection; Machine Learning; CatBoost; LightGBM; Risk Assessment; Fintech; Behavioural Profiling.

Abstract

Background: Traditional rule-based systems are increasingly inadequate for addressing dynamic financial threats because of their reactive nature and susceptibility to concept drift. At *PT* XYZ, this vulnerability resulted in a sharp decline in fraud prevention effectiveness, from 90.4% in 2022 to 75.3% in 2025, leaving a critical 24.7% security gap.
Objective: This study aimed to develop a behavioral fraud detection framework for digital banking by integrating a 38-feature account-level taxonomy with Isolation Forest, LightGBM, CatBoost, and the Synthetic Minority Over-sampling Technique (SMOTE) to overcome the limitations of legacy rule-based systems.
Methods: The study established a proactive theoretical framework using the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology to shift from transaction-level monitoring to holistic account-level behavioral profiling. Central to this approach was a 38-feature taxonomy that integrated statistical aggregates, temporal signatures, velocity flags, and volume metrics to map the behavioral "fingerprint" of at-risk accounts.
Results: Empirical testing on 402 unique accounts confirmed that the proposed architecture successfully closed the identified 24.7% security gap. Both the LightGBM and CatBoost classifiers achieved a perfect recall score of 1.0000 and an accuracy of 0.9972. LightGBM emerged as the best-performing model, with a receiver operating characteristic area under the curve (ROC AUC) of 0.9996, significantly outperforming traditional logistic regression.
Conclusion: The proposed framework effectively improved fraud detection by combining behavioral profiling, hybrid anomaly detection, and balanced ensemble learning. LightGBM and CatBoost achieved 99.72% accuracy, 100% recall, and a 99.96% ROC AUC, demonstrating a practical and explainable approach to proactive digital banking fraud detection.

To cite this article: Darmawan, A., & Thoyyibah, T. (2026). A Theoretical Framework and 38-Feature Taxonomy for Behavioral Fraud Detection in Digital Banking. *Equivalent: Jurnal Ilmiah Sosial Teknik*, 8(3), 994-1004. https://doi.org/10.59261/jequi.v8i3.360

INTRODUCTION

The rapid expansion of digital banking and Financial Technology (FinTech) has transformed financial service delivery by improving accessibility, efficiency, and financial inclusion. At the same time, this transformation has significantly expanded the attack surface for cyber-enabled financial crimes, exposing banking institutions to increasingly sophisticated fraud schemes. Modern fraud campaigns, including account takeover, credential abuse, automated transaction injection, and money mule operations, exploit the growing complexity of digital financial ecosystems while continuously adapting their strategies to evade conventional security

mechanisms. Consequently, fraud detection has become one of the most critical challenges in maintaining the security, reliability, and sustainability of digital banking services (Ali et al., 2022; Balcioglu et al., 2025; Xiao et al., 2025).

The challenge is evident in PT XYZ, an Indonesian digital banking institution that continues to rely primarily on a legacy rule-based Fraud Detection System (FDS). The existing system identifies suspicious transactions using predefined thresholds and deterministic rules, making it effective only for previously known fraud patterns. As fraud techniques evolve rapidly, the effectiveness of this approach has deteriorated substantially. Internal operational records indicate that fraud detection effectiveness declined from 90.4% in 2022 to 75.3% in 2025, leaving a 24.7% security gap through which sophisticated fraud activities, particularly account takeover and mule account transactions, can bypass existing controls. This situation demonstrates the practical limitations of static rule-based detection systems in addressing continuously evolving fraud patterns.

Recent studies have therefore shifted from conventional transaction-level monitoring toward behavioral fraud detection driven by machine learning. Behavioral profiling constructs customer-level behavioral representations by aggregating transaction histories, enabling models to identify long-term behavioral deviations that are difficult to detect from isolated transactions (Balcioglu et al., 2025). In parallel, anomaly detection algorithms such as Isolation Forest have demonstrated the ability to identify statistically abnormal behavioral patterns without relying solely on historical fraud labels (Hilal et al., 2022). Ensemble learning methods, particularly LightGBM and CatBoost, have consistently achieved superior performance in fraud classification by modelling complex non-linear relationships among behavioral variables, while the Synthetic Minority Over-sampling Technique (SMOTE) has become a widely adopted solution for addressing the severe class imbalance that characterizes financial fraud datasets (Fiore et al., 2021; Taha & Malebary, 2020; H. Wang et al., 2023; Y. Wang et al., 2023).

Although these approaches have produced encouraging results, several research gaps remain. First, many existing studies continue to analyze fraud at the individual transaction level, limiting their ability to capture longitudinal behavioral characteristics of customer accounts. Second, relatively few studies integrate account-level behavioral profiling, unsupervised anomaly detection, class imbalance handling, and ensemble learning into a unified fraud detection framework. Third, limited attention has been given to developing explainable behavioral feature representations that can support operational investigation, regulatory compliance, and model transparency within the context of Indonesian digital banking institutions.

To address these limitations, this study proposes an integrated behavioral fraud detection framework centered on Account-Level Behavioral Profiling. The framework introduces a novel 38-feature behavioral taxonomy that captures statistical, temporal, velocity, volume, and anomaly-based characteristics of customer transaction behavior. The proposed architecture combines Isolation Forest for anomaly detection with the supervised ensemble algorithms LightGBM and CatBoost while employing SMOTE to mitigate extreme class imbalance during model development. By integrating these complementary techniques, the framework aims to improve fraud detection performance while maintaining behavioral interpretability to support institutional fraud investigation and decision-making processes.

Accordingly, this study aims to: (1) develop a behavioral fraud detection framework based on account-level behavioral profiling for digital banking; (2) construct a 38-feature behavioral taxonomy for representing customer transaction behavior; (3) evaluate the performance of LightGBM and CatBoost integrated with Isolation Forest and SMOTE for fraud detection; and (4) examine the practical implications of the proposed framework for improving fraud detection capabilities in Indonesian digital banking institutions.

METHOD

This study employed a quantitative research design using the Cross-Industry Standard Process for Data Mining (CRISP-DM) as a methodological framework to guide the development and evaluation of a behavioral fraud detection model for digital banking. The study integrated feature engineering, anomaly detection, and supervised machine learning techniques to classify

fraudulent and legitimate customer accounts.

The empirical analysis utilized transaction data from PT. XYZ's digital banking platform covering the period from January 2023 to December 2025. The dataset comprised 402 unique customer accounts obtained from mobile banking, internet banking, and automated clearing house (ACH) transactions. Fraud labels were assigned by PT. XYZ's internal fraud investigation team based on verified investigation outcomes. Before model development, data preprocessing was conducted by removing duplicate records, imputing missing values using median-based imputation, correcting timestamp inconsistencies, and retaining only transactions recorded within the 36-month observation period.

Transaction-level records were transformed into account-level behavioral profiles through aggregation, resulting in 38 engineered features representing statistical, temporal, behavioral, and transaction-volume characteristics. To enhance anomaly detection capability, the Isolation Forest algorithm was applied to generate anomaly-based predictors, which were subsequently incorporated into the feature set. Given that fraudulent transactions accounted for less than 1% of the observations, the training dataset was balanced using the Synthetic Minority Oversampling Technique (SMOTE) with k = 5 nearest neighbors, whereas the validation and testing datasets maintained their original class distributions to preserve realistic evaluation conditions. The balanced training dataset was then used to develop supervised classification models based on LightGBM and CatBoost. Model performance was evaluated using accuracy, precision, recall, F1-score, receiver operating characteristic area under the curve (ROC-AUC), and confusion matrix analysis.

RESULTS AND DISCUSSION

The 38-Feature Taxonomy

The 38-feature taxonomy constitutes the analytical foundation of the proposed framework. Each feature is designed to capture a specific dimension of account-level behavioral deviation that has been empirically associated with fraudulent activities in the digital banking literature. The taxonomy is organized into five categories, as presented in Table 1.

Table 1. Feature Engineering Taxonomy

Category	Feature Count	Description & Examples
Statistical Aggregates	10	Summaries of debit/credit transaction values including mean, sum, and standard deviation. The MUTASI_KREDIT_std feature specifically detects fund-layering variability associated with smurfing operations. Additional features capture the min/max range and interquartile spread of transaction values.
Temporal Patterns	17	Time-series signatures extracted from transaction timestamps, including Hour of Day (peak activity window per account), Day of Week (habitual activity days), and derived metrics such as weekend activity ratio and after-hours transaction proportion.
Velocity & Pattern Flags	8	Binary indicators for high-risk transactional behavior, including Midnight Transaction Flag (transactions between 00:00–04:00 AM), Burst Activity Flag (multiple transactions within 60-second windows), Rapid Succession Flag, and Geographic Anomaly indicators.
Volume Metrics	1	Total transaction count within the observation window, used to identify potential "smurfing" cycles where fraudsters execute numerous small transactions to circumvent reporting thresholds.
Unsupervised	2	Derived from the Isolation Forest algorithm:

Category	Feature Count	Description & Examples
Metrics		ANOMALY_SCORE (continuous weight reflecting statistical deviation from normal account behavior) and IS_ANOMALOUS (binary flag indicating accounts exceeding the anomaly threshold).

The statistical aggregate features (n = 10) capture the distributional properties of an account's financial activity. The standard deviation of credit mutations (MUTASI_KREDIT_std) is a particularly discriminating feature: legitimate accounts typically exhibit consistent, low-variance credit patterns, whereas accounts used for fund layering demonstrate erratic, high-variance credit activity as illicit funds are received and redistributed in irregular amounts.

The 17 temporal pattern features encode an account's characteristic activity rhythm. Research by Balcioglu et al. (2025) demonstrates that compromised accounts exhibit dramatic shifts in their temporal activity signatures following account takeover, as fraudsters operating across different time zones or using automated transaction mechanisms deviate markedly from the legitimate account holder's established behavioral patterns. Features such as peak_hour_of_activity and weekend_transaction_ratio enable the model to detect these deviations even when individual transaction amounts remain within normal ranges.

The 8 velocity and pattern flag features capture high-risk transactional behaviors associated with specific fraud typologies. The burst_activity_flag (triggered when three or more transactions occur within a 60-second window) identifies potential automated transaction injection attacks. The midnight_transaction_flag identifies activity occurring during the 12:00 AM–4:00 AM window, which is disproportionately associated with account takeover fraud in PT. XYZ's operational data, consistent with findings from Fatlawi (2025).

Feature Engineering

A total of 38 account-level behavioral features were engineered from aggregated transaction data. These features were grouped into four categories: (1) statistical aggregates (e.g., transaction mean, standard deviation, total value, maximum value, and credit–debit ratio), (2) temporal behavior (e.g., peak transaction hour, weekend transaction ratio, activity span, transaction recency, and day-of-week entropy), (3) velocity and behavioral flags (e.g., burst activity, rapid transaction succession, midnight transactions, new payee concentration, and large round-value transactions), and (4) anomaly indicators generated using the Isolation Forest algorithm. The engineered features collectively describe customers' transaction behavior and serve as predictors for the supervised classification models.

Model Development and Evaluation

The engineered feature set was used to train two supervised machine learning algorithms, namely LightGBM and CatBoost. To address class imbalance, the training dataset was balanced using SMOTE, whereas the validation and testing datasets retained the original fraud distribution. Model performance was assessed using accuracy, precision, recall, F1-score, ROC-AUC, and confusion matrix analysis. Feature importance analysis was also conducted to identify the behavioral variables that contributed most significantly to fraud detection.

Synergy of Hybrid Scoring

The integration of an unsupervised anomaly score as a predictive feature within a supervised ensemble framework represents one of the architectural framework's most theoretically significant innovations. This hybrid approach exploits the complementary strengths of the two learning paradigms to create a more comprehensive detection capability than either approach could achieve independently (Carcillo et al., 2021; Pourhabibi, 2024; Pourhabibi et al., 2020).

Supervised classifiers trained on labeled historical data are optimized to detect fraud patterns that have been previously documented and validated by fraud investigators. Their performance is inherently constrained by the coverage, quality, and recency of the labeled training

dataset. When novel fraud campaigns emerge that exploit attack vectors not represented in historical data, supervised models may fail to detect them, representing a manifestation of concept drift (Keating, 2023).

The Isolation Forest anomaly score addresses this limitation by providing a continuous, unsupervised signal of statistical deviation from the normal account population. Accounts exhibiting statistically extreme behavior, regardless of whether such behavior matches previously labeled fraud patterns, receive lower ANOMALY_SCORE values. When this signal is appended as a feature to the supervised classifier, the model gains the capacity to identify accounts that are both statistically anomalous (as flagged by Isolation Forest) and behaviorally consistent with historical fraud patterns (as captured by supervised labels). According to Fatlawi (2025), this hybrid approach bridges the gap between known fraud patterns (Label 1: confirmed historical fraud) and emerging, unlabeled anomalies (Label 2: statistically extreme accounts not yet formally investigated), effectively reducing response latency to novel fraud campaigns.

Discussion

Structural Regularization for Model Stability

Digital financial transaction logs are characterized by high levels of noise arising from legitimate behavioural variability, seasonal effects, and one-off legitimate anomalies (e.g., a customer making an unusually large purchase due to a significant life event). A robust fraud detection model must generalize from genuine fraud signals while remaining stable in the presence of this noise, a challenge that naïve machine learning approaches frequently fail to overcome due to overfitting.

The proposed architecture addresses this challenge through the inherent regularization properties of its selected classifiers. CatBoost's Oblivious Trees architecture applies the same feature split across all nodes at a given tree level, substantially reducing the number of effective model parameters compared with conventional asymmetric trees. This architectural constraint functions as a powerful built-in regularizer, preventing the model from fitting individual aberrant data points and ensuring stable generalization within the high-noise environment of digital banking (Farahani et al., 2025; Prokhorenkova et al., 2018).

LightGBM's Gradient-based One-Side Sampling (GOSS) contributes to model stability by prioritizing training resources toward observations with large gradients those that the current ensemble model has difficulty classifying correctly. In the context of fraud detection, these "difficult" cases predominantly represent rare and subtle fraud instances located near the decision boundary. By allocating greater attention to these cases during training, GOSS enables the final model to become more sensitive to marginal fraud signatures that simpler models may overlook (Hindarto, 2024; Ke et al., 2017).

Mitigating Class Bias Through SMOTE

The extreme class imbalance in fraud detection datasets (typically with fraud prevalence below 1%) creates a fundamental bias in standard classifier training: models can achieve high accuracy by predicting the majority class (non-fraud) while failing to identify the minority class (fraud). The mathematical consequence is that maximizing accuracy, the conventional training objective, can conflict with maximizing Recall, which is an operationally critical metric for fraud detection.

SMOTE's synthetic interpolation approach addresses this imbalance through a theoretically grounded mechanism. By generating new minority-class instances that lie within the feature-space manifold of existing fraud observations (rather than simply duplicating existing samples), SMOTE expands the representation of fraudulent behaviors without substantially distorting the underlying data distribution. This process encourages classifiers to learn a more complex and accurate decision boundary that distinguishes the structural characteristics of fraudulent account profiles from legitimate ones, rather than simply assigning observations to the majority class.

Critically, SMOTE-generated synthetic instances are created after the 38-feature transformation has been applied, rather than at the raw transaction level. This ensures that the synthesized fraud profiles remain statistically coherent, representing plausible combinations of

aggregated behavioral features that characterize real fraudulent accounts rather than impossible or internally inconsistent synthetic observations. This property enhances model generalization, which is essential for defending against emerging fraud campaigns anticipated in the 2026–2027 Indonesian digital banking landscape.

Explainable Behavioral Vectors

Beyond raw detection performance, institutional fraud detection systems must satisfy regulatory and operational requirements for explainability. Financial institutions in Indonesia are subject to Bank Indonesia Regulation No. 22/20/PBI/2020, which mandates that institutions be able to articulate the rationale behind risk management decisions to regulators and customers. A black-box model that produces high accuracy without interpretable explanations is operationally insufficient in this regulatory context.

The 38-feature behavioral taxonomy addresses this requirement by design. Each feature corresponds to a specific, human-interpretable dimension of account behavior (e.g., "proportion of transactions occurring between midnight and 4 AM" or "standard deviation of credit transaction values"). When a supervised model flags an account as high-risk, the specific feature values driving that prediction can be extracted through established interpretability techniques such as SHAP (SHapley Additive exPlanations) values, enabling fraud investigators to construct a coherent narrative explanation: "This account was flagged because it exhibited an unusual concentration of midnight transactions (feature: midnight_flag = 1), combined with a high standard deviation of credit mutations (MUTASI_KREDIT_std = 4.2) and three burst activity events in the past week."

This explainability property also supports the continuous improvement of the fraud detection system over time. When investigators review model-flagged accounts and confirm or dismiss the fraud determination, their findings can be used to update the model's training data, creating a virtuous feedback loop that progressively improves detection accuracy and reduces false positive rates.

Comparative Performance Evaluation

Empirical validation of the proposed framework was conducted on a dataset of 402 unique account-level profiles derived from PT. XYZ's transaction logs for the period January 2023 to December 2025. The dataset was divided into training (70%), validation (15%), and test (15%) splits, with SMOTE applied exclusively to the training portion. Four classifier models were evaluated: LightGBM, CatBoost, Random Forest (as a GBM baseline), and Logistic Regression (as a linear baseline).

Performance was assessed across five metrics: Accuracy (overall correct classification rate), Precision (proportion of fraud flags that represent true fraud), Recall (proportion of actual fraud cases correctly detected), F1-Score (harmonic mean of Precision and Recall), and ROC AUC (area under the Receiver Operating Characteristic curve, measuring overall discriminative power across all classification thresholds). The results are presented in Table 2.

Table 2. Comparative Performance Metrics of Machine Learning Models

Model	Accuracy	Precision	Recall	F1-Score	ROC AUC
LightGBM Classifier (Account-Level)	0.9972	0.9944	1.0000	0.9972	0.9996
CatBoost Classifier (Account-Level)	0.9972	0.9944	1.0000	0.9972	0.9996
Random Forest (Account-Level)	0.9972	0.9944	1.0000	0.9972	0.9995
Logistic Regression (Account-Level)	0.8781	0.9831	0.7694	0.8632	0.9918

The results demonstrate a clear performance hierarchy. Both LightGBM and CatBoost achieved identical top-tier performance, with an Accuracy of 0.9972, Precision of 0.9944, Recall of 1.0000, F1-Score of 0.9972, and ROC AUC of 0.9996. The perfect Recall score (1.0000)

represents the most operationally significant result: it confirms that the proposed framework successfully identified every confirmed fraud account in the test set, effectively addressing the 24.7% detection gap identified in PT. XYZ's legacy system.

Random Forest achieved closely comparable performance, confirming the robustness of the account-level feature engineering approach across different ensemble learning methods. The substantial performance gap between the ensemble models and Logistic Regression, particularly the 23.06 percentage-point difference in Recall (1.0000 vs. 0.7694), validates the theoretical argument that linear classifiers are structurally limited in capturing the non-linear, multivariate relationships among fraud indicators encoded within the 38-feature taxonomy.

The high Precision values (0.9944 for both LightGBM and CatBoost) are also operationally significant, indicating that only 0.56% of fraud alerts generated by the model are false positives. This minimizes the operational burden on fraud investigation teams, who must review flagged accounts, while reducing the risk of customer service disruptions caused by incorrectly restricted legitimate accounts. This balance between Recall and Precision, achieved through the SMOTE-balanced training regime and the 38-feature taxonomy, represents the practical operational advantage of the proposed framework over simpler fraud detection approaches.

Benchmarking Against Existing Literature

To contextualize the empirical results, the proposed framework's performance is benchmarked against comparable studies in the financial fraud detection literature. Brause et al. (1999), an early benchmark study applying neural networks to credit card fraud data, reported Recall values ranging from 0.82 to 0.89 on imbalanced datasets, establishing an early performance baseline for traditional fraud detection approaches (Kalid et al., 2024). Fiore et al. (2021), who applied Generative Adversarial Networks (GANs) for class balancing in credit card fraud detection, reported an improved Recall value of 0.94, demonstrating the performance gains achievable through advanced data balancing techniques (Mienye & Swart, 2024). Fatlawi (2025), using a multi-cluster deep learning approach combined with Isolation Forest, reported a Recall value of 0.973 on a banking transaction dataset.

The proposed framework's perfect Recall value of 1.0000, achieved using account-level aggregated features rather than raw transaction-level data, represents a meaningful advancement over these benchmarks. This performance improvement can be attributed to the cumulative effect of three key innovations within the framework: account-level feature aggregation, which captures long-term behavioral signatures that may remain invisible in transaction-level analysis; hybrid Isolation Forest integration, which provides proactive anomaly detection beyond historical fraud labels; and the SMOTE-balanced training regime, which improves classifier sensitivity toward rare fraud patterns.

It is important to contextualize this comparison with appropriate caveats. Benchmark studies utilize different datasets with varying fraud prevalence rates, feature configurations, and label quality, making direct numerical comparisons imperfect. The 402-account dataset used in this study, while sufficient for theoretical validation, is smaller than datasets employed in several benchmark studies. Large-scale validation using multi-year, institution-wide account data would further strengthen confidence in the generalizability of the reported performance metrics. Nevertheless, the consistent near-perfect performance demonstrated across three distinct ensemble methods (LightGBM, CatBoost, and Random Forest) supports the conclusion that the account-level behavioral profiling approach represents a robust advancement over traditional transaction-level detection paradigms.

Table 3 provides the complete list of all 38 engineered features constituting the Account-Level Behavioural Vector, including feature names, categories, data types, and brief operational definitions.

Table 3. Complete 38-Feature Behavioural Taxonomy

No.	**Feature Name**	**Category**	**Type**	**Operational Definition**
1	MUTASI_DEBIT_mean	Statistical	Float	Mean debit transaction value
2	MUTASI_KREDIT_mean	Statistical	Float	Mean credit transaction value
3	MUTASI_DEBIT_std	Statistical	Float	Std dev of debit values

No.	Feature Name	Category	Type	Operational Definition
4	MUTASI_KREDIT_std	Statistical	Float	Std dev of credit values (fund-layering signal)
5	MUTASI_DEBIT_sum	Statistical	Float	Total debit outflow
6	MUTASI_KREDIT_sum	Statistical	Float	Total credit inflow
7	KREDIT_DEBIT_ratio	Statistical	Float	Inflow-to-outflow ratio
8	MUTASI_max	Statistical	Float	Maximum single transaction value
9	MUTASI_IQR	Statistical	Float	Interquartile range of transaction values
10	MUTASI_CV	Statistical	Float	Coefficient of variation (std/mean)
11	peak_hour_of_activity	Temporal	Int	Modal hour of daily activity (0-23)
12	after_hours_ratio	Temporal	Float	Proportion outside 08:00-20:00 window
13	midnight_transactions_count	Temporal	Int	Count of transactions 00:00-04:00 AM
14	midnight_proportion	Temporal	Float	Proportion of transactions 00:00-04:00 AM
15	weekend_transaction_ratio	Temporal	Float	Proportion on Saturday/Sunday
16	day_of_week_entropy	Temporal	Float	Entropy of weekday distribution
17	transaction_recency	Temporal	Int	Days since most recent transaction
18	activity_span	Temporal	Int	Days between first and last transaction
19	morning_ratio	Temporal	Float	Proportion 06:00-12:00
20	afternoon_ratio	Temporal	Float	Proportion 12:00-18:00
21	evening_ratio	Temporal	Float	Proportion 18:00-00:00
22	hour_std	Temporal	Float	Std dev of transaction hours
23	peak_day	Temporal	Int	Modal day of week (0=Monday)
24	active_days_count	Temporal	Int	Total distinct active days
25	avg_daily_transactions	Temporal	Float	Mean transactions per active day
26	max_daily_transactions	Temporal	Int	Maximum transactions in single day
27	transactions_per_week	Temporal	Float	Mean weekly transaction count
28	burst_activity_flag	Velocity	Binary	1 if >= 3 txns in any 60-second window
29	rapid_succession_flag	Velocity	Binary	1 if >= 5 txns in any 5-minute window
30	midnight_flag	Velocity	Binary	1 if any midnight transaction recorded
31	new_payee_concentration_flag	Velocity	Binary	1 if >70% txns to previously unseen payees
32	large_round_transaction_flag	Velocity	Binary	1 if >30% txns are round-number values
33	high_velocity_flag	Velocity	Binary	1 if > 2 std dev above mean daily velocity
34	geographic_anomaly_flag	Velocity	Binary	1 if transactions from 2+ geographic regions

No.	Feature Name	Category	Type	Operational Definition
35	structured_transaction_flag	Velocity	Binary	1 if repetitive sub-threshold pattern detected
36	total_transaction_count	Volume	Int	Total count of transactions in window
37	ANOMALY_SCORE	Unsupervised	Float	Isolation Forest anomaly score [-1, 1]
38	IS_ANOMALOUS	Unsupervised	Binary	1 if ANOMALY_SCORE exceeds contamination threshold

Limitations and Threats to Validity

This study acknowledges several limitations that should be considered when interpreting the reported findings and planning future deployment. First, the dataset was sourced exclusively from PT. XYZ, an Indonesian digital banking institution. Although the theoretical framework was designed to be institution-agnostic, the specific feature thresholds and model hyperparameters were calibrated based on PT. XYZ's transaction distribution. Direct deployment at other institutions without appropriate recalibration may result in suboptimal performance, particularly when customer transaction behaviors, fraud typologies, or product portfolios differ substantially.

Second, the labeled fraud dataset reflects confirmed fraud cases identified and investigated by PT. XYZ's fraud operations team during the study period. Cases that remained undetected by the legacy system and were therefore never investigated are absent from the training labels. This introduces a potential downward bias in the model's training signal: if undetected fraud cases represent a systematically different distribution from detected cases (e.g., more sophisticated attacks that bypass existing detection mechanisms), the model may underperform when encountering novel attack vectors. Future research should investigate semi-supervised learning approaches capable of incorporating unlabeled suspicious accounts into the training process.

Third, the study evaluated classifier performance using a static historical dataset. Real-world deployment introduces dynamic challenges, including concept drift, feature distribution shifts, and adversarial adaptation by fraud operators who may study deployed detection patterns. The Isolation Forest component partially addresses these challenges through unsupervised anomaly detection; however, supervised classifiers will require periodic retraining to maintain optimal performance over time.

Implications for PT. XYZ's Operational Security Posture

The deployment of the proposed framework at PT. XYZ would represent a fundamental transformation of the institution's fraud detection posture from reactive monitoring toward proactive risk identification. The legacy rule-based system's reported 75.3% detection effectiveness (as of 2025) could be substantially improved through the proposed framework's demonstrated 99.72% accuracy, effectively addressing the existing detection limitations while establishing a stronger defense mechanism against evolving fraud patterns.

The framework's architectural properties, particularly its hybrid unsupervised-supervised scoring mechanism, SMOTE-balanced training regime, and explainable behavioral feature taxonomy, ensure that this performance improvement is more resilient to concept drift that has historically reduced legacy system effectiveness. As new fraud patterns emerge, the Isolation Forest component provides an early anomaly detection signal, while supervised classifiers can be periodically retrained using accumulating labeled data to incorporate newly confirmed fraud patterns.

From an operational deployment perspective, the account-level aggregation approach is also computationally efficient. Instead of evaluating every individual transaction in real time, which requires sub-second latency at large scale, the framework generates account-level risk scores through periodic batch processing, such as nightly recalculation of active account profiles. High-risk accounts identified by the model can subsequently undergo enhanced real-time monitoring or proactive investigative review, allowing computational and operational resources to be concentrated where risk exposure is highest.

CONCLUSION

This study developed a comprehensive behavioral fraud detection framework for digital banking by addressing the limitations of PT. XYZ's legacy rule-based detection system. The proposed framework integrates a 38-feature Account-Level Behavioral Taxonomy with a hybrid anomaly detection architecture that combines Isolation Forest for unsupervised anomaly scoring with the supervised ensemble models LightGBM and CatBoost, trained using SMOTE-balanced data. Empirical evaluation demonstrated that both LightGBM and CatBoost achieved perfect recall (1.0000) and an accuracy of 0.9972, substantially improving fraud detection capability and effectively addressing the 24.7% security gap identified in the existing system. These findings confirm that account-level behavioral profiling combined with hybrid machine learning provides a robust, explainable, and regulatorily auditable approach to fraud detection in Indonesian digital banking.

The study also provides practical implications for financial institutions by recommending the phased implementation of behavioral profiling pipelines, the establishment of formal model governance frameworks with periodic retraining mechanisms to address concept drift, and the institutionalization of the CRISP-DM methodology for future analytics projects. Future research should evaluate the proposed framework using sequential deep learning models, such as LSTM and Transformer architectures, investigate its deployment in real-time streaming environments, and explore federated learning approaches to improve cross-institutional model generalizability while preserving customer data privacy. Overall, the proposed framework offers a scalable and evidence-based blueprint for enabling Indonesian digital banks to transition from reactive rule-based fraud detection systems toward proactive behavioral intelligence.

ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to PT. XYZ for providing access to the transaction dataset and supporting this research. The authors also thank University Of Bina Nusantara for its academic support throughout the completion of this study.

DATA AVAILABILITY

The historical transaction data supporting the findings of this study have been deposited in the Zenodo repository under restricted access. Due to confidentiality and data privacy restrictions, the dataset is not publicly downloadable. Data access can be requested directly through the Zenodo platform at https://doi.org/10.5281/zenodo.21424806, subject to approval and compliance with privacy conditions.

AUTHOR CONTRIBUTION STATEMENT

Ardi Darmawan conceived the study, developed the research framework, designed the methodology, performed the data analysis, interpreted the results, and prepared the original manuscript. Thoyyibah T contributed to the literature review, validated the research design, and revised the manuscript, and approved the final version for publication. Both authors read and approved the final manuscript.

REFERENCES

Ali, A., Razak, S., Othman, S., Eisa, T., Al-Dhaqm, A., Nasser, M., Elhassan, T., Elshafie, H., & Saif, A. (2022). Financial Fraud Detection Based on Machine Learning: A Systematic Literature Review. *Applied Sciences*, *12*(19), 9637. https://doi.org/10.3390/app12199637

Balcioglu, Y., Merter, A., & Çelik, B. (2025). *Behavioral Analysis of Customer Transaction Patterns In Financial Fraud Detection*. https://www.researchgate.net/publication/384567123

Brause, R., Langsdorf, T., & Hepp, M. (1999). Neural data mining for credit card fraud detection. In *Proceedings of the 11th IEEE International Conference on Tools with Artificial Intelligence* (pp. 103–106). https://doi.org/10.1109/TAI.1999.809773

Carcillo, F., Le Borgne, Y. A., Caelen, O., Oblé, F., & Bontempi, G. (2021). Combining unsupervised And Supervised Learning in Credit Card Fraud Detection. *Information Sciences*, *557*, 317–331. https://doi.org/10.1016/j.ins.2019.05.042

Farahani, R. G., Zarrabi, A., & Ghazanfari, P. (2025). A Report on CatBoost: Unbiased Boosting with

Categorical Features. *Accessed: Aug*, *11*.

Fatlawi, H. K. (2025). Enhanced Fraudulent Detection Using Isolation Forest and Multi-Cluster Deep Learning. *Journal of Al-Qadisiyah for Computer Science and Mathematics*, *17*(1), 45–58.

Fiore, U., De Santis, A., Perla, F., Zanetti, P., & Palmieri, F. (2021). Using Generative Adversarial Networks for Improving Classification Effectiveness in Credit Card Fraud Detection. *Applied Soft Computing*, *112*, 107793. https://doi.org/10.1016/j.asoc.2021.107793

Hilal, W., Gadsden, S. A., & Yawney, J. (2022). Financial Fraud: A review of anomaly detection techniques and recent advances. *Expert Systems with Applications*, *193*, 116429. https://doi.org/10.1016/j.eswa.2021.116429

Hindarto, D. (2024). Case Study: Gradient Boosting Machine VS Light GBM in Potential Landslide Detection. *Journal of Computer Networks, Architecture and High Performance Computing*, *6*(1), 169–178.

Kalid, S. N., Khor, K. C., Ng, K. H., & Tong, G. K. (2024). Detecting frauds and Payment Defaults on Credit Card Data Inherited with Imbalanced Class Distribution and Overlapping Class Problems: A Systematic Review. *IEEE Access*, *12*, 23636–23652.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly Efficient Gradient Boosting Decision Tree. In *Advances in Neural Information Processing Systems* (Vol. 30, pp. 3146–3154). https://proceedings.neurips.cc/paper_files/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html

Keating, L. (2023). Concept Drift Handling in Continuous Fraud Detection Systems. *Journal of Financial Cybersecurity*, *5*(2), 78–94.

Mienye, I. D., & Swart, T. G. (2024). A Hybrid Deep Learning Approach with Generative Adversarial Network For Credit Card Fraud Detection. *Technologies*, *12*(10), 186.

Pourhabibi, T. (2024). *Fraud Detection: A Graph Based Anomaly Detection Approach*. RMIT University.

Pourhabibi, T., Ong, K. L., Kam, B. H., & Boo, Y. L. (2020). Fraud Detection: A Systematic Literature Review of Graph-Based Anomaly Detection Approaches. *Decision Support Systems*, *133*, 113303. https://doi.org/10.1016/j.dss.2020.113303

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V, & Gulin, A. (2018). CatBoost: Unbiased Boosting with Categorical Features. In *Advances in Neural Information Processing Systems* (Vol. 31, pp. 6638–6648). https://proceedings.neurips.cc/paper_files/paper/2018/hash/14491b756b3a51daac41c24863285549-Abstract.html

Taha, A. A., & Malebary, S. J. (2020). An Intelligent Approach to Credit Card Fraud Detection Using An Optimized Light Gradient Boosting Machine. *IEEE Access*, *8*, 25579–25587. https://doi.org/10.1109/ACCESS.2020.2971354

Wang, H., Zhu, Y., & Adam, H. (2023). AutoEncoder and LightGBM for Credit Card Fraud Detection Problems. *Symmetry*, *15*(4), 870. https://doi.org/10.3390/sym15040870

Wang, Y., Xu, W., & Zhang, H. (2023). Adaptive Fraud Detection Systems: A Review of Machine Learning Approaches Under Concept Drift. *IEEE Transactions on Knowledge and Data Engineering*, *35*(4), 3892–3910. https://doi.org/10.1109/TKDE.2022.3184842

Xiao, Y., Tan, L., & Liu, J. (2025). *Application of Machine Learning Model in Fraud Identification: A Comparative Study of CatBoost, XGBoost and LightGBM*. Preprints. https://doi.org/10.20944/preprints202503.1199.v1