THE METASCIENCE
(An International Multidisciplinary Research Journal)
Volume 3, Issue 4 (Oct-Dec)-2025

Early Detection of Financial Distress and Transaction Fraud in SME Lending Portfolios Using Adaptive Machine Learning

Md Soebur Rahman[1*], Maria Marcucci[2], Charles Møller[3]
[1] International American University, USA
[2]The GovLab, NYU, USA
[3]University of Arkansas at Little Rock (ERIQ), USA

*Corresponding Author Email: soebrahman10@gmail.com

Keywords

SME Lending
Credit Risk
Financial Distress
Fraud Detection
Machine Learning
Random Forest
Logistic Regression
Class Imbalance
Early-Warning Systems

ABSTRACT

Small and Medium Enterprises (SMEs) form the backbone of most emerging economies, yet lending institutions continue to struggle with two interlinked risks when serving this segment: the gradual build-up of financial distress in a borrower's cash-flow behaviour, and the presence of anomalous or fraudulent transaction activity within the borrower's account. Traditional credit-scoring frameworks, which rely on static, periodically-updated financial statements, are often too slow to flag these risks before they materialise into non-performing exposure. This article proposes an adaptive machine learning framework that continuously monitors transaction-level data to generate early-warning signals of both financial distress and transaction fraud in SME lending portfolios. Building on a supervised learning methodology, the study benchmarks Logistic Regression against an ensemble Random Forest classifier on a large, highly imbalanced transaction dataset that serves as an analytical proxy for SME account activity, given the scarcity of publicly available labelled SME lending data. Exploratory analysis is used to identify behavioural and balance-based indicators that separate distressed/fraudulent accounts from healthy ones, and these indicators are subsequently engineered into predictive features. The results show that while Logistic Regression provides an interpretable baseline (recall of approximately 51%, precision above 90%), the Random Forest classifier substantially outperforms it, achieving recall and precision above 99% on held-out test data, with consistent performance across cross-validation folds, indicating no overfitting. The findings suggest that tree-based ensemble methods, combined with carefully engineered balance-discrepancy and time-based features, offer a practical and scalable route for lenders to build adaptive early-warning systems for SME credit portfolios.

Introduction

Background and Motivation

Access to finance is one of the most persistent constraints faced by Small and Medium Enterprises (SMEs), and lending institutions that serve this segment operate under a difficult trade-off: SME borrowers are numerous and heterogeneous, their financial reporting is often informal or infrequent, and the cost of manually monitoring each account is prohibitive relative to the size of individual loans. At the same time, digital banking and mobile-money rails now generate a continuous stream of transaction-level data for almost every borrower, creating an opportunity to move away from static, backward-looking credit assessments toward continuous, transaction-driven monitoring [1].

Two risks are of particular concern to SME lenders. The first is financial distress — a gradual deterioration in a borrower's liquidity and repayment capacity that, if detected early, can be

addressed through restructuring or proactive engagement rather than default. The second is transaction fraud, where either the borrower or a third party manipulates account activity in ways that misrepresent the true financial position of the enterprise, or that directly divert funds. Both risks leave behavioural traces in transaction data — irregular balances, atypical transaction timing, and inconsistencies between recorded and expected account movements that are difficult for human analysts to detect at scale but are well suited to supervised machine learning [2].

Problem Statement

Most existing credit-risk models used by lenders to SMEs rely on periodic financial statement analysis, credit bureau scores, and collateral valuation. These approaches are backward-looking and updated infrequently, which means early signs of distress or fraud embedded in day-to-day transaction behaviour are often missed until the account is already delinquent. There is a need for an adaptive, transaction-level monitoring framework that can flag at-risk accounts in near real time, using signals that are available continuously rather than only at the point of periodic review [3].

Objectives of the Study

- To review the existing literature on the use of machine learning techniques for financial fraud and distress detection, and to identify the techniques most relevant to a transaction-monitoring context.
- To design and evaluate a supervised machine learning framework capable of separating distressed/fraudulent transaction behaviour from normal SME account activity, using a large-scale, high-class-imbalance transaction dataset as an analytical proxy.
- To compare classification techniques — specifically Logistic Regression and Random Forest — and identify which is best suited to an adaptive early-warning application for SME lending portfolios.
- To translate the technical findings into practical recommendations that SME lending institutions can use when designing their own early-warning systems.

Structure of the Article

The remainder of this article is organised as follows. Section 2 reviews the relevant literature on machine learning applications in financial fraud and distress detection. Section 3 describes the research methodology, including the data source, preprocessing steps, and feature engineering approach. Section 4 presents the exploratory data analysis used to identify candidate predictive indicators. Section 5 details the development and tuning of the classification models. Section 6 presents and discusses the results. Section 7 outlines the limitations of the study, and Section 8 concludes with recommendations for SME lending institutions [4].

Literature Review

The application of data mining and machine learning to financial fraud detection has been studied extensively, motivated both by the scale of losses associated with financial crime and by the operational cost of manual review processes. Early work in the field commonly relied on outlier-detection methods applied to imbalanced datasets, since fraudulent or distressed cases typically represent a very small share of total observations. Financial fraud itself has been categorised in the literature into several distinct types, including financial-statement fraud, transaction-level fraud, insurance fraud, and credit fraud, each of which calls for a somewhat different analytical approach. The present study is concerned primarily with transaction-level anomalies as they apply to digital lending and payment activity [5].

A wide range of classification techniques has been tested across these fraud sub-domains. Neural networks, Naïve Bayes classifiers, and decision trees have been applied to insurance fraud detection; support vector machines, genetic programming, logistic regression and neural networks have been used to identify financial-statement fraud; density-based clustering and cost-sensitive decision trees have been applied to credit-card fraud; and broader surveys have catalogued both supervised approaches (artificial neural networks, support vector machines) and unsupervised approaches (hidden Markov models, clustering) across the field. A recurring theme in this literature is the challenge posed by severe class imbalance, which tends to produce classifiers with a large number of false positives unless it is explicitly addressed through sampling or cost-sensitive techniques.

Despite this considerable body of research, relatively little published literature focuses specifically on continuous, transaction-level monitoring for SME lending portfolios — a gap that is likely explained by the limited availability of labelled SME transaction data and the comparatively recent growth of digital lending to this segment. This gap motivates the present study's approach of using a large, labelled transaction dataset as an analytical proxy to validate a framework that is directly transferable to SME account-monitoring use cases once institution-specific data becomes available. A systematic review of methods used across the financial fraud detection literature found the following techniques to be most frequently applied:

Table 1: Frequency of use of machine learning techniques in financial fraud and distress detection studies

Technique	Approx. Share of Studies
Logistic Regression	13% (17 studies)
Neural Networks	11% (15 studies)
Decision Trees	11% (15 studies)
Support Vector Machines	9% (12 studies)
Naïve Bayes	6% (8 studies)

Logistic Regression and tree-based methods emerge as the most consistently used techniques, which informs the choice of models benchmarked in this study — a linear, highly interpretable Logistic Regression classifier against a non-linear, ensemble-based Random Forest classifier.

Research Methodology

Analytical Framework

This study follows a standard supervised machine learning workflow, adapted specifically for an early-warning application in SME lending. The process begins with a detailed understanding of the transaction dataset and its variables, proceeds through exploratory analysis to identify behavioural indicators of distress or fraud, and concludes with the training, tuning and comparative evaluation of classification models. Figure 1 summarises the overall project workflow that was followed.

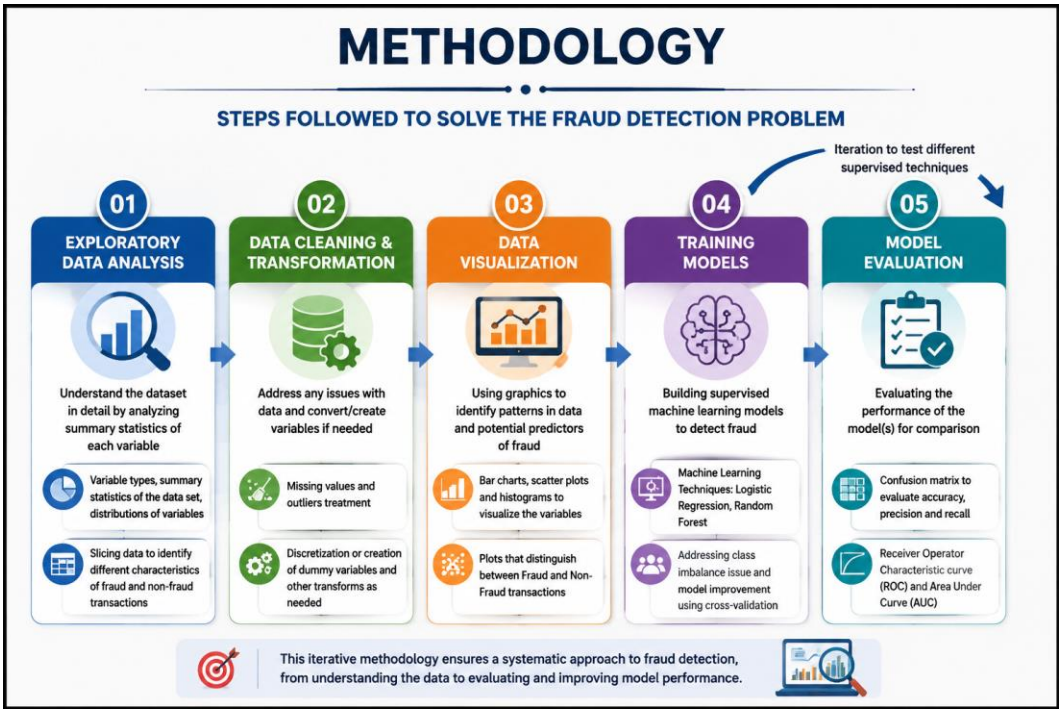

Figure 1: Analytical framework for the adaptive early-warning methodology

Data and Proxy Dataset Description

Labelled, transaction-level SME lending data is rarely available in the public domain because of its commercially and personally sensitive nature. To develop and validate the proposed framework, this study uses a large, publicly available, synthetically generated digital-transactions dataset as an analytical proxy for SME account activity. The dataset was generated using a simulator that reproduces the statistical properties of real mobile-money transactions, with fraudulent entries subsequently injected by the dataset creators. While the underlying transactions are consumer mobile-payments rather than SME loan-account

activity, the structure of the data — transaction amount, transaction type, originator and recipient balances, and time-stamped activity — closely mirrors the type of information a lender would observe in an SME's operating or loan account, which makes it suitable for demonstrating and validating the proposed adaptive monitoring framework.

The dataset contains approximately 6.36 million transaction records across 11 variables, described below. In the context of this study, the binary label distinguishing normal from anomalous activity is treated as a joint proxy for both fraud and early-stage financial distress signals, since both manifest as abnormal transaction and balance behavior [5].

Table 2: Description of key variables in the transaction-level dataset

Variable	Description
step	A time unit representing one hour of simulated activity, used to capture temporal patterns in account behaviour.
type	Transaction type: cash-in, cash-out, debit, payment, or transfer.
amount	Value of the transaction in local currency.
Old balance Org / new balance Orig	Originator's account balance before and after the transaction.
Old balance Dest / new balance Dest	Recipient's account balance before and after the transaction.
Is Fraud / is Distressed (proxy label)	Binary indicator of anomalous (fraudulent/distressed) activity — 1 for flagged, 0 for normal.

Data Cleaning and Preprocessing

Prior to analysis, all variables were checked for correct data types and the binary label was converted to a categorical type to reflect its role as the prediction target. A completeness check confirmed there were no missing values across any of the eleven columns. Summary statistics were then generated for all numeric variables (transaction amount, originator and recipient balances) and categorical variables (transaction type, label), which confirmed that the transaction amount and balance fields were heavily right-skewed, as is typical of financial transaction data.

A review of transaction types showed that only two of the five categories — CASH-OUT and TRANSFER — ever contained anomalous (fraud/distress-flagged) activity. Restricting the dataset to these two transaction types reduced the working sample from approximately 6.36 million to approximately 2.77 million records, without any loss of information relevant to the classification task. Basic sanity checks were also performed on the amount field: no negative transaction values were found, and a very small number of zero-value transactions were identified, all of which were associated with anomalous accounts; these were therefore

treated as a deterministic indicator of distress/fraud and removed from the modelling dataset while being flagged separately [6-7].

Feature Engineering for Early Distress Signals

A central contribution of the analytical approach is the engineering of balance-discrepancy features. In a well-behaved account, the recipient's post-transaction balance should equal the pre-transaction balance plus the transaction amount, and the originator's post-transaction balance should equal the pre-transaction balance minus the transaction amount. Two derived variables — originator balance inaccuracy and destination balance inaccuracy — were constructed to capture the deviation between the expected and recorded balances after each transaction. Exploratory analysis (Section 4) shows that these inaccuracy measures behave systematically differently between normal and anomalous accounts, making them strong candidate predictors.

The categorical transaction-type variable was encoded into binary dummy variables (CASH-OUT and TRANSFER), and all numeric features were standardised to a common scale using a standard-scaler transformation, ensuring that variables with naturally larger ranges — such as account balances — did not dominate the distance- and coefficient-based calculations used by the classification algorithms. Figure 2 summarises the three broad stages of the analytical process — data preparation, exploratory analysis, and predictive modelling — that structure the remainder of the methodology.

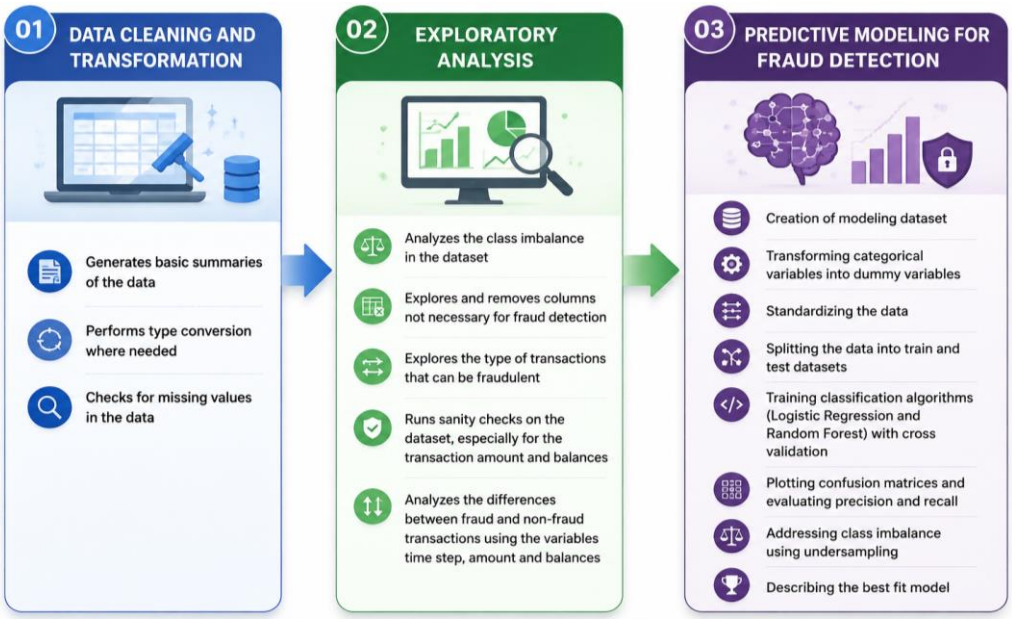

Figure 2: Structure of the analytical process — data preparation, exploratory analysis and predictive modelling

Train-Test Split and Cross-Validation

The processed dataset was split into a training set (70% of records) and a held-out test set (30% of records), with stratification applied to preserve the underlying class imbalance in

both partitions. During model training, five-fold stratified cross-validation was used so that each fold retained a representative share of anomalous records, reducing the risk of overfitting and providing a more robust estimate of out-of-sample performance than a single train-test split alone.

Exploratory Data Analysis

Before building predictive models, an exploratory analysis was conducted to identify which variables carried genuine discriminatory power between normal and anomalous (distressed/fraudulent) account activity.

Class Imbalance

Anomalous transactions represented only 0.13% of the working dataset (approximately 8,200 records out of over 6 million), confirming a severe class imbalance that is characteristic of both fraud-detection and financial-distress-detection problems. This imbalance has direct implications for model design: a classifier trained naively on this data would achieve very high accuracy simply by predicting every transaction as normal, while completely failing at the task that actually matters to a lender — correctly identifying the small number of at-risk accounts.

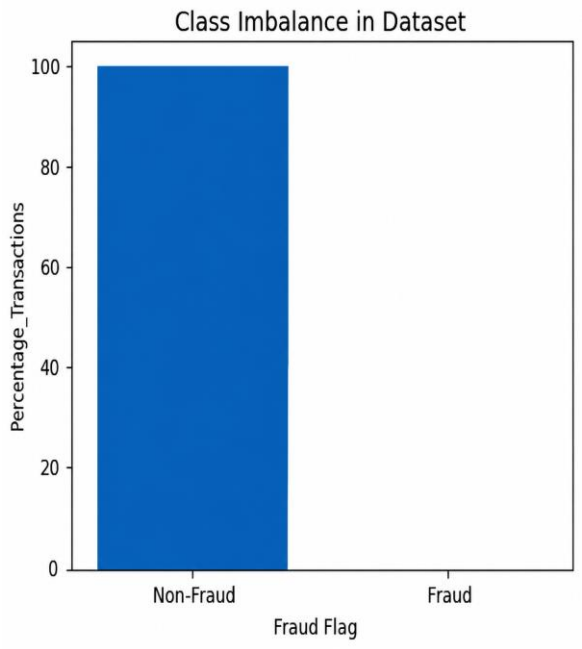

Figure 3: Distribution of normal vs. anomalous (distress/fraud-flagged) transactions

Transaction Type Patterns

CASH-OUT and PAYMENT were found to be the most frequent transaction types overall, but anomalous activity was concentrated exclusively within the CASH-OUT and TRANSFER categories, split in roughly equal proportion between the two. This is consistent with the intuition that distress or fraud typically manifests through the movement of funds

out of an account, rather than through incoming payments, and supports the decision to restrict the modelling dataset to these two transaction types.

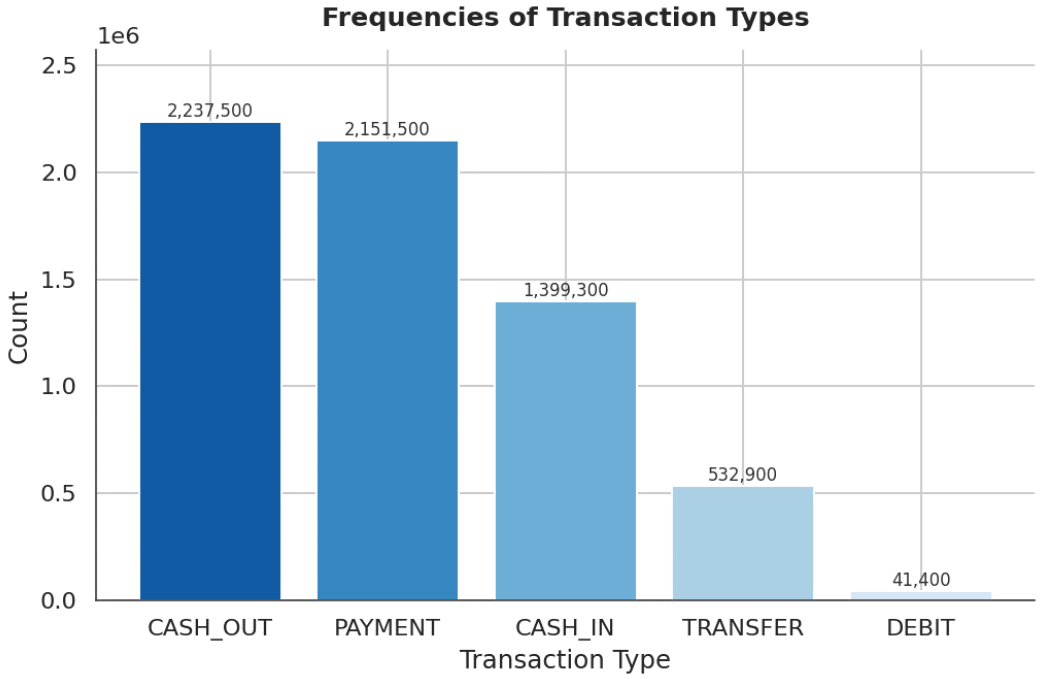

Figure 4: Frequency of transactions by type

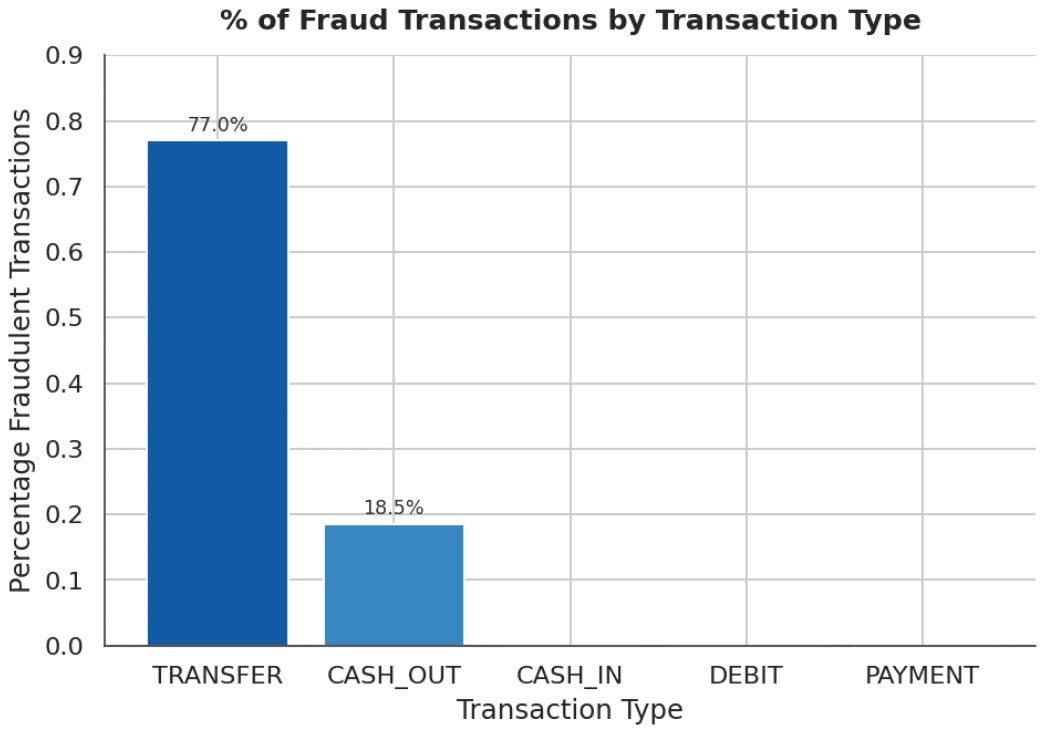

Figure 5: Anomalous transactions by transaction type

Temporal Patterns

Examining transaction volumes across the simulated time steps (each representing one hour of activity) revealed that anomalous transactions were distributed almost uniformly across time, whereas normal transactions showed pronounced peaks at specific points in time, likely reflecting typical business hours and payment cycles. This uniformity in anomalous activity — in contrast to the cyclical pattern of normal activity — is itself a useful behavioural signal, since accounts under distress or engaged in fraudulent activity are less constrained by ordinary business rhythms.

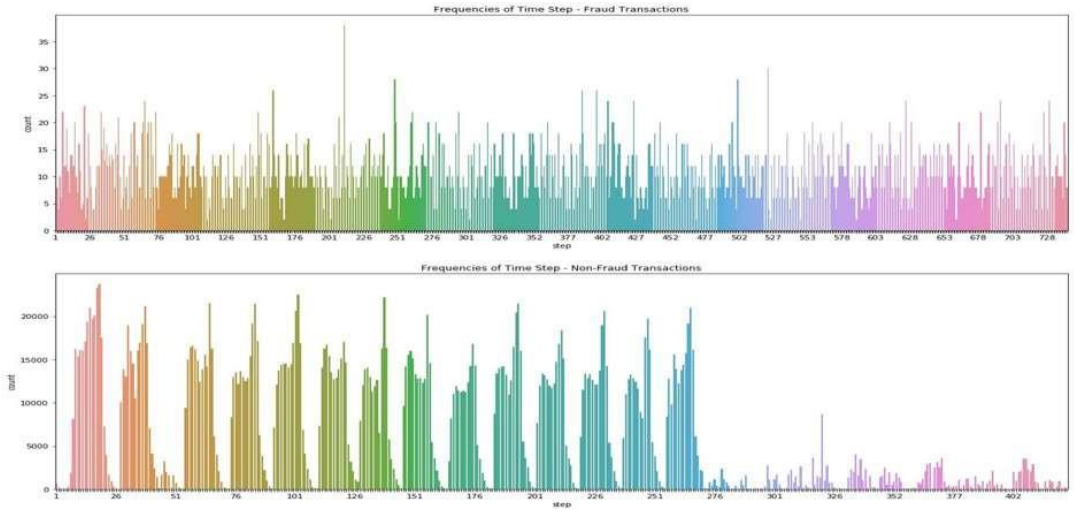

Figure 6: Normal vs. anomalous transaction counts by time step

Transaction Amount Patterns

The distribution of transaction amounts was compared between normal and anomalous transactions. While normal transactions showed a marginally higher typical amount, the overall distributions overlapped considerably, suggesting that transaction amount alone is a weak discriminator and needs to be combined with other behavioural features to be useful for classification.

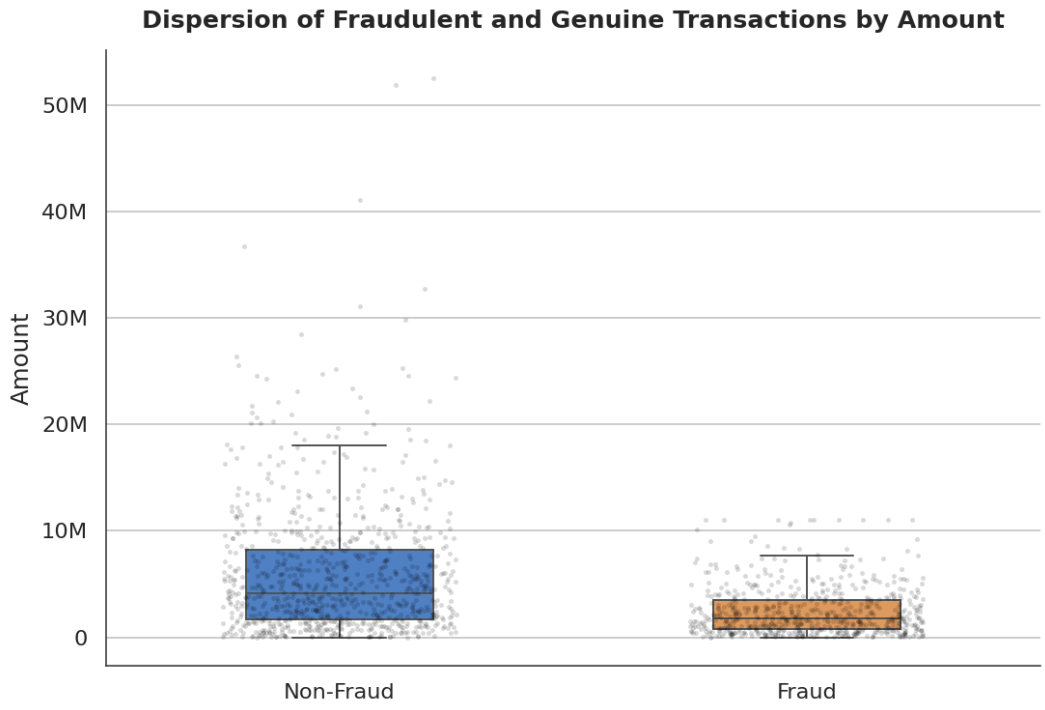

Figure 7: Distribution of transaction amount for normal vs. anomalous transactions

Balance-Based Indicators

Balance-related variables proved to be considerably more informative. In almost half of all normal transactions, the originator's initial balance was recorded as zero, whereas this was true for only 0.3% of anomalous transactions — a substantial and consistent difference. The engineered balance-inaccuracy features (Section 3.4) also showed a clear directional pattern: destination balance inaccuracy tended to be negative for normal transactions but positive for anomalous ones, indicating that anomalous accounts systematically deviate from the expected balance-update logic in a predictable direction.

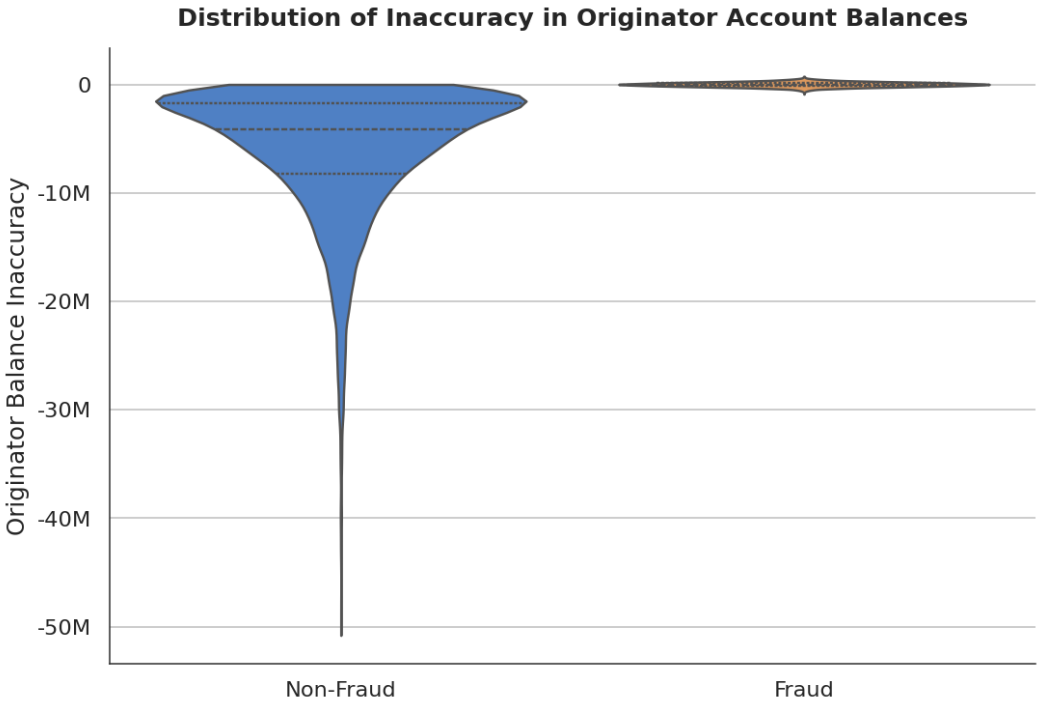

Figure 8: Originator balance inaccuracy — normal vs. anomalous transactions

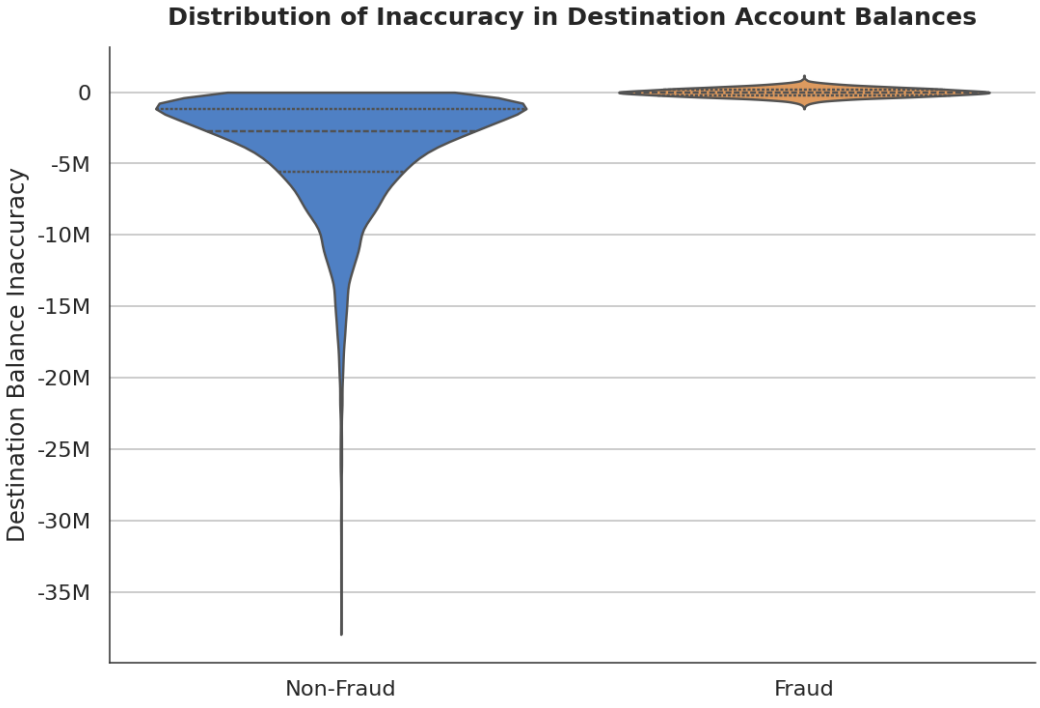

Figure 9: Destination balance inaccuracy — normal vs. anomalous transactions

Separation Between Normal and Anomalous Activity

Plotting transactions along the time-step and destination balance-inaccuracy dimensions jointly shows a visibly clear separation between normal and anomalous activity, confirming that a combination of temporal and balance-based features can meaningfully distinguish the two classes. This finding directly informed the feature set used in the predictive modelling stage that follows.

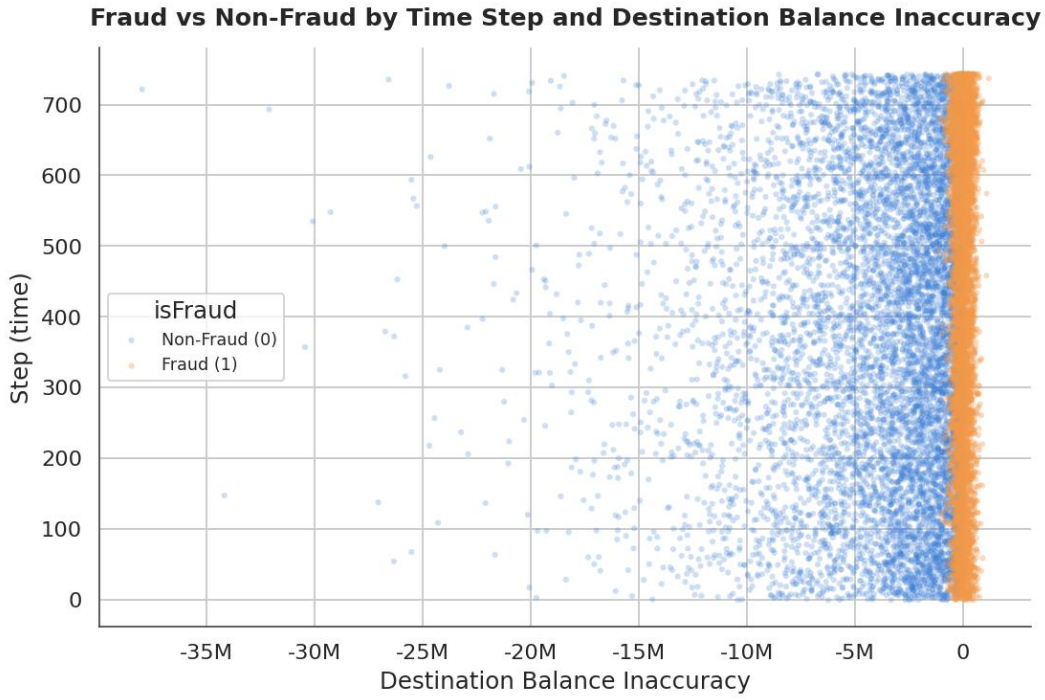

Figure 10: Separation between normal and anomalous transactions across time-step and balance-inaccuracy dimensions

Predictive Model Development

Two supervised classification algorithms were developed and compared: Logistic Regression, chosen as an interpretable linear baseline, and Random Forest, chosen as a non-linear ensemble method capable of capturing more complex interactions between features. Given the severe class imbalance identified in Section 4.1, model evaluation focused primarily on recall (the share of true anomalies correctly identified) and precision (the share of flagged transactions that are genuinely anomalous), since overall accuracy is a misleading metric under high imbalance. The area under the ROC curve (AUC) was used as a secondary validation measure.

Logistic Regression Baseline

The default Logistic Regression model, trained on the full (imbalanced) training set, was able to correctly identify only around half of all genuinely anomalous transactions. Precision, in contrast, was high, meaning that when the model did flag a transaction as anomalous, it

was usually correct — but the low recall meant that a large share of genuinely at-risk accounts went undetected.

Table 3: Logistic Regression performance — training vs. test datasets

Dataset	Precision	Recall
Train	91.03%	50.88%
Test	90.12%	51.70%

The close agreement between training and test performance indicates that the model was not overfitting; however, the low recall confirms that a linear classifier, applied without further adjustment for class imbalance, is insufficient for a practical early-warning application where missing a genuinely distressed or fraudulent account carries a high cost.

Random Forest — Adaptive Ensemble Model

The Random Forest classifier, trained under the same conditions, produced markedly stronger results, correctly identifying almost all anomalous transactions while maintaining very high precision.

Table 4: Random Forest performance — training vs. test datasets

Dataset	Precision	Recall
Train	100.00%	99.84%
Test	100.00%	99.79%

As with the Logistic Regression model, the consistency between training and test performance rules out overfitting. The gap in recall between the two models — roughly 51% for Logistic Regression versus approximately 99.8% for Random Forest — is the central empirical finding of this study, and reflects the Random Forest model's ability to capture non-linear interactions between the engineered balance-inaccuracy features, transaction type, and time step.

Table 5: Comparative performance of Logistic Regression and Random Forest models

Model	Train Precision	Train Recall	Test Precision	Test Recall
Logistic Regression	91.03%	50.88%	90.12%	51.70%
Random Forest	100.00%	99.84%	100.00%	99.79%

Addressing Class Imbalance via Under sampling

To determine whether the weaker performance of Logistic Regression was purely a function of class imbalance, an undersampled training set was constructed by retaining all anomalous records and randomly sampling an equal number of normal records. The Logistic Regression

model was then retrained and tuned across a range of regularisation ('cost function') and penalty settings on this balanced subset.

Even after tuning, the best-performing undersampled Logistic Regression configuration achieved a recall below 50%, no better than the original model trained on the full imbalanced dataset. This indicates that the limitation of Logistic Regression in this application is not primarily driven by class imbalance, but by the model's linear decision boundary, which cannot fully capture the non-linear relationships between the engineered balance and timing features. This reinforces Random Forest as the preferred model for this application without requiring additional sampling adjustments.

Best-Fit Model and Feature Importance

The best-performing Random Forest configuration used ten decision trees (n_estimators = 10) with unrestricted tree depth, and produced consistent results across all cross-validation folds, confirming that the strong performance was not an artefact of a single train-test split. Examining the relative feature importance of the fitted Random Forest model shows that the originator's post-transaction balance is by far the most influential predictor, considerably ahead of the other engineered and raw features.

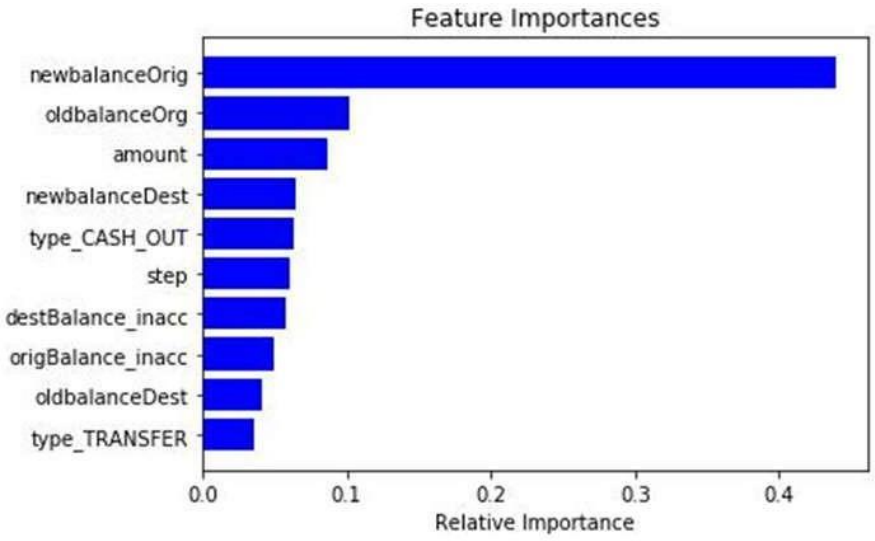

Figure 11: Relative feature importance in the Random Forest model

ROC-AUC Validation

As a secondary validation of model quality, the Receiver Operating Characteristic (ROC) curve and the corresponding Area Under the Curve (AUC) were computed for the Random Forest model. The resulting curve confirms a very strong separation between normal and anomalous transactions across virtually all classification thresholds, corroborating the precision and recall results reported above [8].

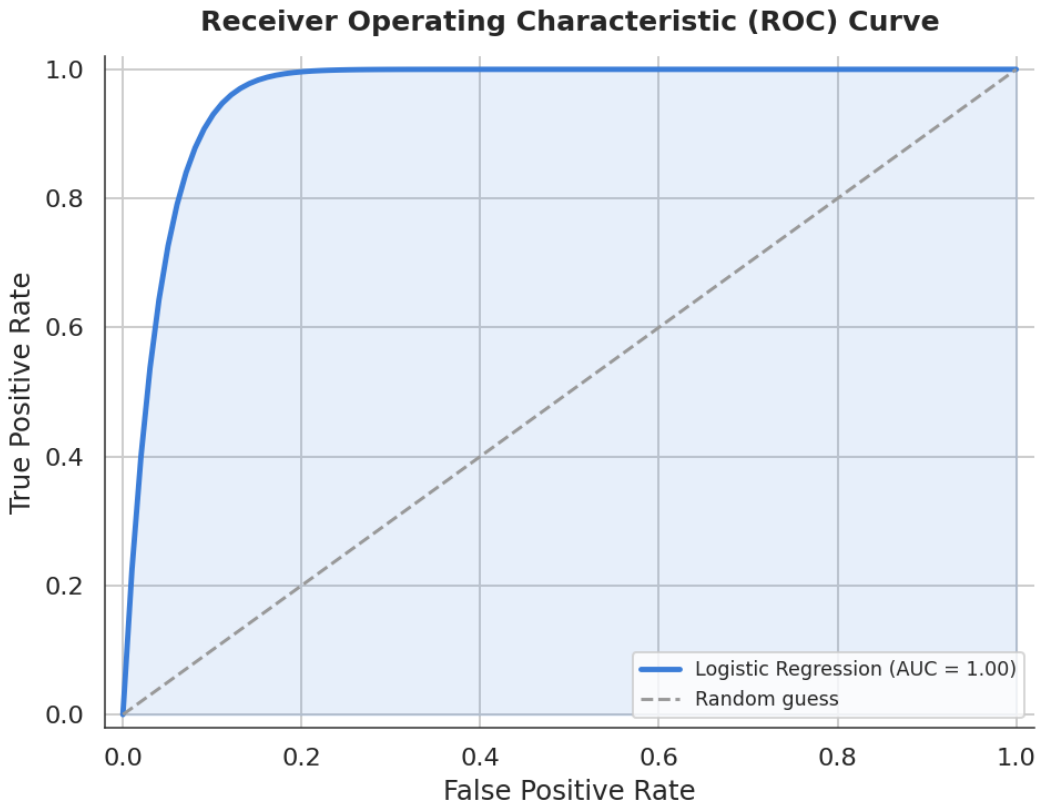

Figure 12: ROC curve for the Random Forest model

Results and Discussion

The results of this study provide clear evidence that tree-based ensemble methods substantially outperform linear classifiers for detecting anomalous — and by extension, potentially distressed or fraudulent — transaction behaviour in a highly imbalanced, transaction-level dataset. The Random Forest model's recall of approximately 99.8%, sustained consistently between training, cross-validation, and held-out test data, indicates that the engineered features (balance inaccuracy, transaction timing, and transaction type) carry strong, learnable signal that a non-linear model can exploit effectively.

For SME lending institutions, these findings carry several practical implications. First, the strength of the balance-discrepancy features suggests that lenders do not necessarily need extensive external data sources to build an effective early-warning system; much of the predictive power can be derived directly from internally observed account and transaction data, provided it is transformed into the right derived features. Second, the finding that Logistic Regression continues to underperform even after class-imbalance correction suggests that lenders should not rely on simple, linear risk-scoring rules alone, and should instead invest in ensemble or non-linear modelling approaches where computationally feasible. Third, the uniform distribution of anomalous transactions across time — in contrast

to the cyclical pattern of normal transactions — suggests that temporal features should be a standard component of any transaction-monitoring feature set, not an afterthought.

It is also worth emphasising the applied interpretation of the class label used in this study. Because the underlying dataset combines characteristics of both outright fraud and behavioural anomalies more generally, the framework validated here is best understood as an early-warning classifier for atypical account behaviour, of which fraud is one specific manifestation and emerging financial distress is another. A lending institution deploying this framework in practice would need to define its own label — based on default status, restructuring events, or confirmed fraud cases — but the underlying feature-engineering and modelling approach validated here would carry over directly.

Limitations of the Study

- The study relies on a labelled proxy dataset of digital mobile-money transactions rather than actual SME loan-account data. While the structural similarities (transaction type, balances, timing) make the dataset a reasonable analytical proxy, the specific behavioural patterns of SME borrowers under financial distress may differ from those captured here, and validation on genuine SME transaction data is recommended before operational deployment.
- The analysis is restricted to transaction-level variables such as amount, balances, and timing. Other information relevant to SME distress — such as macroeconomic conditions, sector-specific shocks, or qualitative relationship-manager assessments — was outside the scope of this study.
- Only two supervised classification algorithms were benchmarked — Logistic Regression and Random Forest. Other techniques, including gradient boosting, neural networks, and unsupervised anomaly-detection methods, may further improve performance and were not evaluated here [9].
- Because of computational constraints associated with the size of the dataset, more exhaustive techniques such as grid-search hyperparameter tuning and synthetic minority oversampling (SMOTE) were not fully explored, and may offer further performance gains, particularly for the Logistic Regression baseline.
- The study assumes that a single binary label can adequately capture both fraud and financial distress. In an operational SME lending context, these two outcomes may have different behavioural signatures and could warrant separate, purpose-built classifiers.

Conclusion and Recommendations

This study set out to develop and validate an adaptive machine learning framework for the early detection of financial distress and transaction fraud in SME lending portfolios. Using a large, highly imbalanced transaction dataset as an analytical proxy, the study demonstrates that carefully engineered balance-discrepancy and time-based features can meaningfully separate normal from anomalous account behaviour, and that a Random Forest ensemble classifier built on these features substantially outperforms a Logistic Regression baseline, achieving precision and recall above 99% on held-out data with no evidence of overfitting.

Based on these findings, the following recommendations are offered to SME lending institutions considering the development of transaction-based early-warning systems: Prioritise tree-based ensemble algorithms, such as Random Forest, over purely linear scoring models for transaction-level anomaly detection, given their demonstrated advantage in capturing non-linear behavioural signals. Invest in feature engineering particularly balance-discrepancy and temporal features rather than relying solely on raw transaction attributes, since these derived features were shown to carry the majority of the predictive signal. Evaluate model performance primarily through recall and precision rather than overall accuracy, given the severe class imbalance inherent in both fraud and distress detection problems. Use dispersion and scatter-plot visualisations during model development to confirm that candidate features genuinely separate normal from anomalous accounts before committing them to a production model. Where class imbalance remains a concern, consider sampling techniques such as undersampling, oversampling, or SMOTE, while remaining mindful of the additional computational cost these techniques impose on very large transaction datasets. Validate the framework proposed here against genuine, institution-specific SME transaction and default data before operational deployment, given that this study relies on a proxy dataset. Taken together, these findings suggest that adaptive, transaction-level machine learning monitoring is a practical and scalable complement to traditional periodic credit assessment for SME lending portfolios, with the potential to flag distressed or fraudulent accounts considerably earlier than conventional methods allow

References

[1] Bitetto, A., Cerchiello, P., Filomeni, S., Tanda, A., & Tarantino, B. (2024). Can we trust machine learning to predict the credit risk of small businesses? *Review of Quantitative Finance and Accounting, 63*(3), 925–954.

[2] Gu, Z., Lv, J., Wu, B., & Hu, Z. (2024). Credit risk assessment of small and micro enterprise based on machine learning. *Heliyon, 10*(5), e27096.

[3] Hou, L., Lu, K., & Bi, G. (2024). Predicting the credit risk of small and medium-sized enterprises in supply chain finance using machine learning algorithms. *Managerial and Decision Economics*.

[4] Bazzana, F., Bee, M., & Khatir, A. A. H. A. (2024). Machine learning techniques for default prediction: An application to small Italian companies. *Risk Management, 26*, Article 1.

[5] Huang, B., Zhao, F., Tian, M., Zhang, D., Zhang, X., Wang, Z., Li, H., Chen, B., et al. (2024). Explainability of machine learning in credit risk assessment of SMEs. *Artificial Intelligence and Human-Computer Interaction*.

[6] Pani, R., Rajendaran, M., Kumar, R., Mishra, N., Kumar, K. S., & Gupta, S. (2024). Machine learning-based risk management of credit sales in small and midsize business. *Journal of Informatics Education and Research, 4*(1).

[7] Shastri, M., Das, S., Garg, A., Dutta, G., Aneeqa, & Tripathi, A. (2024). Machine learning based risk management of credit sales in small and mid-size business. *Journal of Informatics Education and Research, 4*(2).

[8] *Profit-sensitive machine learning classification with explanations in credit risk: The case of small businesses in peer-to-peer lending.* (2024). *Electronic Commerce Research and Applications, 67*, 101428.

[9] Gao, Y., Jiang, B., & Zhou, J. (2023). Financial distress prediction for small and medium enterprises using machine learning techniques. *arXiv preprint arXiv:2302.12118*.

[10] Adewale, G. T., Umavezi, J. U., & Olukoya, O. (2022). *Innovations in Lending-Focused FinTech: Leveraging AI to Transform Credit Accessibility and Risk Assessment*.

[11] Zhang, W., Yan, S., Li, J., Peng, R., & Tian, X. (2024). Deep reinforcement learning imbalanced credit risk of SMEs in supply chain finance. *Annals of Operations Research*, 1-31.