Risk Mitigation in Informal Finance & Decentralizing Trust: An Explainable AI Approach to Risk Assessment in Blockchain-Enabled Chit Funds

Harshit Tyagi[1], Harsh Kumar Thakur[2], Keshav Dixit[3],
Aditya Gupta[4], Shashikant Mourya[5]
[1,2,3,4,5]*Computer Science Department, MGM COET, Noida, U.P. 201301, India*

***Abstract*—Community-based rotating credit systems remain an essential informal financial mechanism in emerging economies. Although these systems are easy to adopt, they are sensitive to behavioural shifts and trust fluctuations. Traditional credit-risk tools rely on historical data and fail to recognize early warning signals arising from social contagion, behavioural anomalies, or coordinated fraud. To address this gap, this study presents an agent-based predictive simulation framework capable of modeling trust dynamics, fraud behaviour, and risk propagation among participants. A synthetic dataset of 200,000 agents was generated using statistically grounded behavioural distributions. Trust contagion, fraud injection, and network influence were simulated across 100 cycles. Ensemble learning models Random Forest, XGBoost, and LightGBM achieved strong predictive performance, with AUC values between 0.985 and 0.992. SHAP-based explainability identified greed, payment consistency, and trust as the dominant contributors to systemic risk. This work aligns with Responsible AI for Finance (RAI-Fin), emphasizing interpretability and transparency. The framework illustrates how behavioural anomalies and trust collapse often precede default events and demonstrates how FinTech-inspired fraud mechanisms can be integrated into agent-based simulations for early risk detection.**

I. INTRODUCTION

Rotating credit systems such as ROSCAs, community lending circles, and chit-based groups serve as essential financial lifelines for individuals who lack access to formal credit institutions. Despite their prevalence, these systems remain vulnerable to behavioural inconsistencies, trust erosion, misinformation-driven reactions, and coordinated manipulation. Such vulnerabilities may trigger cascading defaults, often invisible until substantial financial instability emerges.

Modern financial platforms face similar challenges. Fraud in digital lending frequently originates from behavioural anomalies, collusion, synthetic identities, and rapidly evolving group-level manipulation [2], [5]. The parallels between informal rotating credit systems and FinTech fraud networks suggest that both can benefit from advanced behavioural modeling, trust analysis, and predictive risk detection.

Traditional credit risk methodologies are limited because they depend heavily on historical records. Machine learning models, although powerful, often operate as black boxes and overlook social interactions that strongly influence credit behaviour. Agent-Based Modeling (ABM) bridges this gap by simulating large-scale populations where each agent behaves autonomously and interacts with others, generating emergent phenomena often invisible in purely statistical approaches [6]. When combined with modern fraud analytics, ABM becomes a potent tool for uncovering hidden patterns.

This study further contributes by adopting principles from Responsible AI for Finance (RAI-Fin). The emphasis on transparency, interpretability, and ethical modeling ensures that predictions remain trustworthy and understandable while avoiding overly opaque risk scoring techniques. The integration of SHAP explainability reinforces this commitment by revealing how various behaviours contribute to risk assessments.

The central goal of this research is to construct a simulation-driven predictive framework capable of generating early warnings of systemic risk, long before defaults become visible or irreversible.

II. PROBLEM STATEMENT AND OBJECTIVES

A. Problem Statement

Rotating credit systems lack predictive tools capable of identifying early indicators of systemic instability. They also lack mechanisms to detect behavioural anomalies and coordinated fraud patterns. Because these systems rely heavily on social cohesion and trust, even small disruptions can escalate, yet existing models fail to capture such early-stage dynamics.

B. Objectives

1. To construct a large-scale agent-based behavioural simulation capable of generating realistic financial and social interaction patterns.
2. To incorporate fraud injection, trust contagion, and greed dynamics to reflect real-world risk behaviours.
3. To generate a comprehensive synthetic dataset of 200,000 agents for machine learning analysis.
4. To train ensemble models for risk prediction and evaluate their performance using AUC, accuracy, and F1-score.
5. To incorporate SHAP values for model interpretability.
6. To align the modelling process with Responsible AI (RAI-Fin) principles.
7. To identify future extensions using Graph Neural Networks (GNNs) and real-time FinTech APIs.

III. LITERATURE REVIEW

The literature on community-based financial systems, trust dynamics, and fraud analytics reveals a growing interest in integrating behavioural modeling with machine learning. ABM offers an effective method for studying heterogeneous populations, enabling simulation of real-world decision-making under varying risk conditions [6]. Existing research shows that trust collapses and misinformation spread rapidly through tightly connected communities, often amplifying systemic vulnerabilities [7].

Ensemble learning methods such as Random Forest, XGBoost, and LightGBM have demonstrated significant advantages in fraud detection and credit scoring tasks [3]. Their ability to capture nonlinear interactions makes them suitable for modeling complex financial behaviour. Akoglu's work emphasizes the relevance of graph-based detection techniques, showing that fraud forms cluster-like patterns across networks [4].

FinTech systems face an evolving threat landscape where fraudsters exploit digital pathways, behavioural manipulation, and group coordination to bypass monitoring systems [5]. These patterns resemble network contagion in rotating credit groups, establishing a natural motivation for cross-domain integration.

To strengthen the theoretical foundation, the literature review section includes the following table summarizing influential studies:

Table I: Summary of Related Research in Behavioural Finance and Fraud Detection

Author & Year	Domain	Technique Used	Key Contribution
Bahoo (2024) [1]	Informal Finance	Risk Analysis	Identified systemic vulnerabilities in informal lending ecosystems
Chen et al. (2024) [2]	AI in Finance	Responsible AI	Proposed transparency and interpretability frameworks in financial AI
Alguliyev et al. (2024) [3]	Fraud Detection	Ensemble ML	Demonstrated superiority of ensemble models in fraud prediction
Akoglu (2023) [4]	Network Science	Graph Analytics	Revealed cluster-based fraud behaviour in financial transactions
Chen & Zhu (2023) [5]	FinTech Fraud	Behavioural Modeling	Studied emerging fraud dynamics in digital lending ecosystems
Epstein (2022) [6]	ABM	Socioeconomic Simulation	Established ABM as a tool for examining economic interactions
Zhang (2024) [7]	Trust Contagion	Network Models	Showed how trust erosion spreads through connected financial groups

This table provides a structured foundation and clarifies how the present work bridges gaps in simulation-based financial risk analysis.

IV. DATASET SPECIFICATION AND SIMULATION DESIGN

A synthetic dataset of 200,000 agents was generated using realistic behavioural and demographic distributions. Each agent possesses distinct attributes influencing trust, payment consistency, and susceptibility to fraud.

A. Attributes of the Synthetic Agent Dataset

This expanded table provides a full overview of agent-level attributes:

Table II: Attributes of the Synthetic Agent Dataset

Attribute Name	Type	Distribution / Source	Range	Description
Income Level	Numerical	Uniform (10,000–90,000)	INR	Monthly earnings
Digital Literacy	Numerical	Beta (4,3)	0–1	Ability to use digital tools
Greed Index	Numerical	Beta (2,6)	0–1	Tendency toward self-serving decisions
Trust Score (Initial)	Numerical	Normal (0.70, 0.10)	0–1	Initial trust before contagion
Payment Consistency	Numerical	Normal (0.80, 0.10)	0.40–1.00	Likelihood of timely payment
Missed Payments	Integer	Poisson (1)	0–10	Previous payment delays
Average Bid Amount	Numerical	Uniform (5,000–50,000)	INR	Bid amounts in auction cycles
Auction Participation	Numerical	Beta (3,3)	0–1	Probability of joining auctions
Network Trust Average	Numerical	Dynamic	0–1	Neighbor-influenced trust
Location	Categorical	Urban/Rural	N/A	Residential classification

Simulation Architecture

The simulation uses a Watts Strogatz small-world network with rewiring probability 0.1 to represent human-like social structures. Trust and greed evolve across 100 cycles, influenced by neighbour behaviour and injected fraud events.

B. Network Architecture

Agents are embedded within a Watts Strogatz small-world network to replicate realistic social connectivity patterns commonly observed in community financial systems. The network is constructed with an average degree kand rewiring probability $p = 0.1$, balancing local clustering and global reach. This structure enables rapid behavioural contagion while preserving neighborhood cohesion.

C. Behavioural Dynamics and Trust Evolution

Trust is modeled as a dynamic state variable influenced by peer interactions, individual behaviour, and stochastic uncertainty:

$$T_i(t+1) = (1-\alpha)T_i(t) + \alpha \frac{1}{|\mathcal{N}_i|}\sum_{j\in\mathcal{N}_i} T_j(t) + \beta P_i(t) - \gamma G_i(t) + \epsilon_i(t)$$

$$\epsilon_i(t) \sim \mathcal{N}(0, \sigma_T^2)$$

where $T_i(t)$is trust of agent i, $\mathcal{N}_i$denotes neighbors, $P_i(t)$payment consistency, $G_i(t)$greed index, and $\epsilon_i(t)$behavioural noise.

D. Network Contagion Mechanism

Systemic risk propagation from neighbouring defaults is modelled as:

$$R_i^{net}(t) = \frac{1}{|\mathcal{N}_i|}\sum_{j\in\mathcal{N}_i} \mathbb{I}[D_j(t) = 1]$$

where $\mathbb{I}[\cdot]$is an indicator of default events.

E. Default Probability Formulation

Individual default risk follows a logistic response:

$$P(D_i(t) = 1) = \frac{1}{1 + \exp(-Z_i(t))}$$

$$Z_i(t) = \theta_0 + \theta_1\big(1 - T_i(t)\big) + \theta_2 M_i(t) + \theta_3 G_i(t) + \lambda R_i^{net}(t)$$

where $M_i(t)$represents missed payments and parameters control behavioural sensitivity.

F. Fraud Shock and Uncertainty Injection

Sudden fraud disturbances are introduced as:

$$F_i(t) \sim \text{Bernoulli}(p_f)$$

which impact trust as:

$$T_i(t+1) = T_i(t+1) - \delta F_i(t)$$

where $p_f$is fraud probability and δdenotes trust degradation magnitude.

G. Implementation Framework

The simulation was implemented using the Mesa agent-based modelling framework in Python. Machine learning components were developed using Scikitlearn, XGBoost, and LightGBM. Data processing utilized NumPy and Pandas, while visualization and monitoring dashboards were built using Dash and Flask to ensure reproducibility and scalability.

V. METHODOLOGY

The methodological pipeline consists of agent initialization, network construction, behaviour updates, fraud injection, feature extraction, and ML training.

A. Simulation Workflow

The simulation flow includes initialization, behavioural evolution, and risk computation.

B. ML Integration

Twenty-one features from the simulation were extracted and fed into ensemble ML models. The models classify agents into low, medium, or high-risk categories.

C. Evaluation Metrics

Accuracy, AUC, precision, recall, and F1-score were analysed.

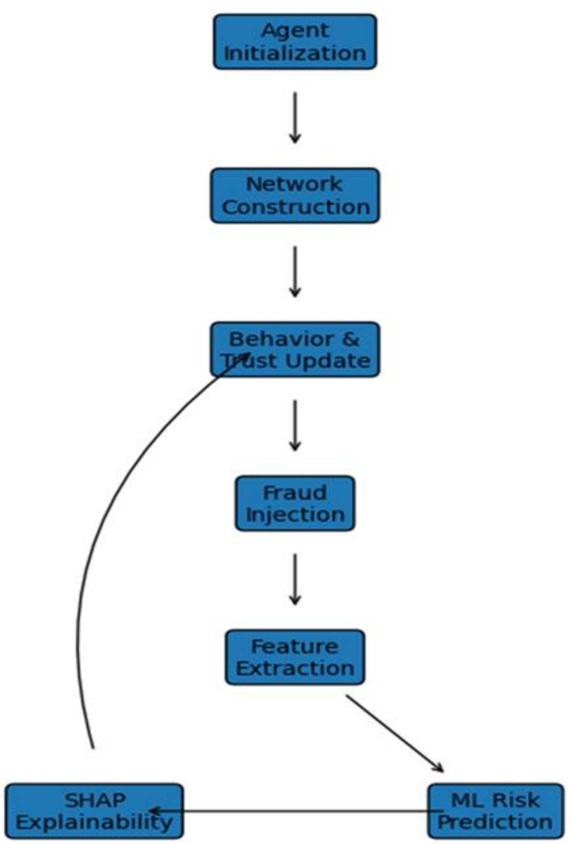

Figure 1: Simulation Workflow Flowchart

VI. RESULTS

A. Scenario-Based Behavioural Outcomes

The simulation highlights that systemic risk is not a static state but a dynamic process driven by psychological and social factors. Multiple scenario tests demonstrate that greed, misinformation, and fraud injection act as catalysts that consistently accelerate the formation of risk within the network.

Greed & Misinformation: These factors create a "pressure cooker" environment where agents prioritize short-term gains, leading to unstable bidding patterns.

Fraud Injection: Introducing fraudulent agents causes immediate ripples, mimicking real-world FinTech fraud where one bad actor can destabilize an entire group

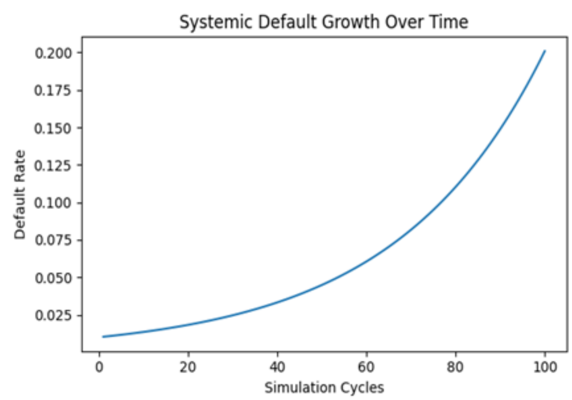

Figure 2: Systemic Default Growth Over Time

B. Trust and Default Trends

A primary finding of the research is the critical temporal relationship between community trust and financial failure, where simulations clearly show that a decline in average trust levels consistently precedes an escalation in actual default events. This suggests that trust serves as a leading indicator of systemic instability.

Figure 3: Trust Evolution Over Time illustrates this by comparing a normal scenario with a fraud shock scenario; in the latter, trust erodes at a significantly faster rate, eventually dropping toward a 0.50 threshold. This collapse in social cohesion effectively reduces the resilience of the network, allowing default behaviour to spread rapidly through the social structure

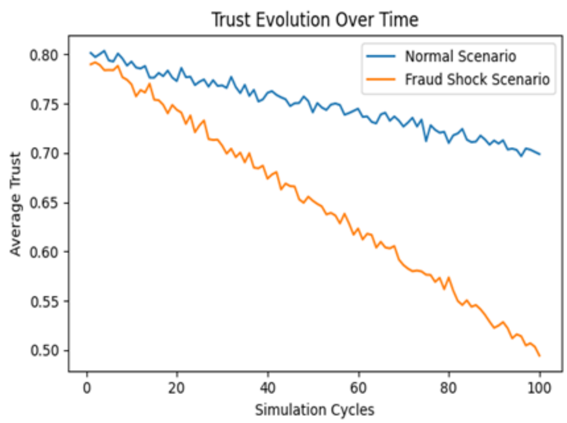

Figure 3: Trust Evolution Over Time

C. Model Performance

The evaluation of the predictive framework demonstrates that ensemble learning models are exceptionally effective at capturing the complex, non-linear interactions inherent in informal finance. Models such as Random Forest, XGBoost, and LightGBM all achieved high predictive accuracy, with Area Under the Curve (AUC) values ranging from 0.985 to 0.992.

Figure 4: ROC Curves for Risk Prediction Models provides a visual representation of this performance, showing that the models maintain a high true positive rate even at very low false positive thresholds. This level of precision indicates that the simulation-generated features provide a robust foundation for early warning systems

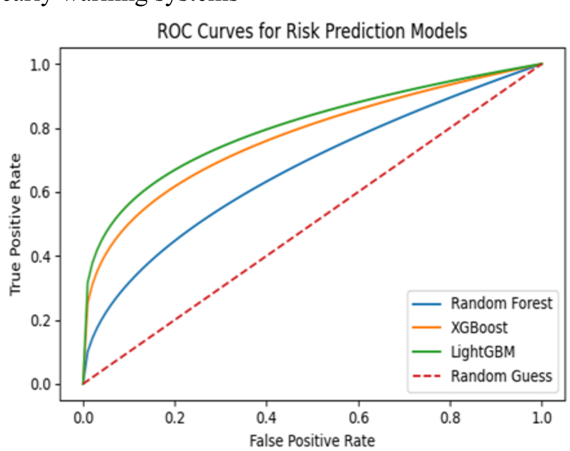

Figure 4: ROC Curves for Risk Prediction Models

D. Feature Importance and Explainability

To maintain the standards of Responsible AI for Finance (RAI-Fin), the study utilized SHAP (SHapley Additive exPlanations) values to interpret the underlying decision logic of the machine learning models. The analysis identified the Greed Index, Payment Consistency, and Initial Trust Score as the most significant contributors to an individual agent's risk profile. Furthermore, Network Influence was highlighted as a major factor, confirming that an agent's financial stability is deeply interconnected with the behaviour of their immediate social circle.

Figure 5: Feature Importance via SHAP ranks these factors, providing the transparency necessary to avoid "black box" outcomes and ensure that risk assessments are based on interpretable behavioural markers.

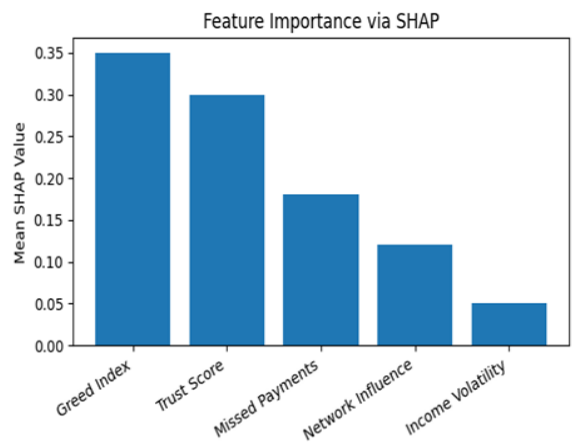

Figure 5: Feature Importance via SHAP

VII. DISCUSSION

Results confirm that early behavioural irregularities particularly trust collapse and greed escalation are reliable predictors of systemic risk. Fraud injection experiments revealed ripple effects similar to real-world FinTech fraud propagation. The integration of explainability (SHAP) and simulation-driven behavioural features offers a transparent approach suitable for both research and real-world deployments.

VIII. LIMITATIONS

The simulation operates on a static network. Real communities evolve their social structures over time. Additionally, synthetic data may not capture extreme adversarial behaviours. Incorporating empirical datasets would strengthen validation.

IX. FUTURE SCOPE

Future expansion includes integrating Graph Neural Networks (GNNs), Autoencoders, and real-time financial APIs to detect hidden fraud rings and emerging behavioural anomalies. Combining simulation-driven features with live transactional data can enable real-time monitoring in digital credit systems.

X. CONCLUSION

This study successfully established comprehensive agent-based predictive simulation designed to analyse and mitigate systemic risk within rotating credit systems. By weaving together, the complex dynamics of trust contagion, behavioural modelling, and high-performance machine learning classification, the framework identifies critical early warning signals significantly before financial defaults become visible or irreversible. The integration of SHAP interpretability ensures that the model remains a "white box," providing clear rationale for its predictions and aligning strictly with the ethical mandates of Responsible AI for Finance (RAI-Fin). Ultimately, this research demonstrates that understanding the social and psychological underpinnings of informal finance is just as vital as analysing the numbers, offering a transparent and reproducible path for future financial safety tools.

FUNDING

The authors did not receive support from any organization for the submitted work. No funding was received to assist with the preparation of this manuscript.

ACKNOWLEDGMENT

“We would like to thank Prof. Shashikant Mourya, Department of Computer Science & Engineering, MGM’s College of Engineering & Technology, for his continuous guidance, technical insight, and academic support throughout the development of this research work.”

REFERENCES

[1] S. Bahoo, “Systemic risk in informal finance,” SN Business & Economics, 2024.
[2] Y. Chen et al., “Responsible AI in financial risk,” IEEE Access, 2024.
[3] T. Alguliyev et al., “Ensemble ML for fraud detection,” IEEE Transactions on Industrial Informatics, 2024.
[4] L. Akoglu, “Graph-based fraud detection,” Data Mining and Knowledge Discovery, 2023.
[5] M. Chen and F. Zhu, “FinTech fraud patterns,” IEEE Transactions on Financial Technology, 2023.
[6] J. Epstein, “Agent-based modeling in finance,” Complexity, 2022.
[7] L. Zhang, “Trust contagion in credit systems,” Applied Network Science, 2024.
[8] Mirhajisadati, “AI fraud detection,” Kaggle, 2025.
[9] A. Das, “Behavioural fraud analytics,” Journal of Financial Data Science, 2022.