Problem Statement

SiriusAI

REAL-WORLD FINANCIAL ANOMALIES

IDENTIFYING FRAUD BEFORE IT SPREADS

By :- Parth Tyagi

June 19, 2025

Fake Invoices & Trade-Based Money Laundering (TBML)

Example:

A company claims to import ₹50 lakh worth of furniture, but customs data shows only plastic chairs worth ₹2 lakh. The ₹48 lakh gap is laundered money.

Industry Use:

Common in cross-border import/export businesses. Used by criminal networks to justify illicit fund movement.

Risk Indicators:

Mismatch in invoice vs shipment type/value. Repeated transactions to/from high-risk jurisdictions.

Source: https://www.fatf-gafi.org

Round-Tripping Transactions

Example:

Money flows from Company A → Company B → Company C → back to Company A — creating artificial revenue or laundering funds.

Industry Use:

Found in stock manipulation and fake invoicing. Detected via entity relationship graph analytics.

Detection Techniques:

Graph-based transaction flow detection.

Risk Indicators:

Circular fund movement without valid invoices.

Source: https://www.acfe.com

Cuckoo Smurfing

Example:

Illegal remittance agent instructs someone to deposit ₹5 lakh into a legitimate account. Another person overseas receives equivalent amount, masked as unrelated transfer.

Detection Techniques:

Transaction synchronization and unrelated transfer matching.

Risk Indicators:

Unusual match of debit-credit amounts across borders.

Source: https://www.fatf-gafi.org

Politically Exposed Persons (PEP) Unusual Transactions

Example:

A mid-level politician’s relative receives large foreign transfers from tax havens, without declared sources.

Industry Use:

Detected via enhanced due diligence (EDD) and risk scoring.

Detection Techniques:

PEP list screening and transaction profiling.

Risk Indicators:

Unexplained foreign remittances, high-risk origin countries.

Gift Card / Voucher Laundering

Example:

Fraudsters purchase high-volume gift cards using illicit funds, then resell for cash or crypto.

Industry Use:

Used to bypass banking transaction monitoring systems.

Detection Techniques:

Track gift card volume, velocity, and resale activity.

Risk Indicators:

High-frequency card purchases and quick liquidation.

Source: https://www.nrf.com

Mule Account Networks

Example:

Multiple personal accounts receive identical amounts and quickly forward them to another account.

Industry Use:

Used to obscure money trail across many accounts.

Detection Techniques:

Behavioral clustering and transaction pattern mapping.

Risk Indicators:

Low balance retention, fast outflows, synchronized activity.

Synthetic Identity Creation

Example:

A fake profile is created using real PAN + fake mobile + address to pass KYC.

Industry Use:

Seen in BNPL apps, neobanks, and digital onboarding.

Detection Techniques:

Cross-check identity attributes across datasets.

Risk Indicators:

Inconsistent identity signals, repeated data reuse.

Structuring Across Locations

Example:

Multiple customers across branches deposit INR 49,500 daily to avoid CTR limits.

Industry Use:

Common method to evade reporting thresholds.

Detection Techniques:

Centralized data consolidation and behavioral models.

Risk Indicators:

Repeat sub-threshold deposits across geographies.

Source: https://www.fatf-gafi.org

Dormant Account Reactivation

Example:
A dormant account suddenly receives a large deposit and transfers it offshore.

Industry Use:
Used for mule activity or stolen account exploitation.

Detection Techniques:
Behavior deviation models and risk scoring engines.

Risk Indicators:
Dormant → active jump with high-value transfers.

Source: https://fiuindia.gov.in

Clustered Loan Defaults

Example:

10 MSMEs from same branch default, introduced by the same agent.

Industry Use:

Agent collusion or mass identity fraud.

Detection Techniques:

Link analysis between introducers and borrowers.

Risk Indicators:

Common introducer, time-based clustering.

Source: https://www.europol.europa.eu

Back-to-Back Multi-Bank Loans

Example:

Real estate firm borrows from 3 banks in 2 weeks without any project.

Industry Use:

Exploits delay in credit bureau updates.

Detection Techniques:

Velocity rules and interbank credit sharing.

Risk Indicators:

High borrow volume without activity.

Source: https://www.worldbank.org

Mass Digital Onboarding

Example:
500 accounts opened from the same IP with similar email patterns.

Industry Use:
Bot farms and fake identity creation.

Detection Techniques:
Device fingerprinting and email/domain patterns.

Risk Indicators:
IP repetition, username/email similarities.

Source: https://www.acams.org

Cross-Border Card Misuse

Example:

Card used in Delhi at 10 AM and in Dubai at 12:30 PM same day.

Industry Use:

Cloned card or stolen credentials.

Detection Techniques:

Impossible travel rules and geolocation logic.

Risk Indicators:

Physical impossibility of travel.

Source: https://usa.visa.com

Elderly Fraud Patterns

Example:

78-year-old’s account used for mobile wallet transfers from new device.

Industry Use:

Often overlooked fraud pattern.

Detection Techniques:

Device behavior and age-based profiling.

Risk Indicators:

Device mismatch, abnormal activity for profile.

Source: https://www.consumerfinance.gov

Account Takeovers (ATO)

Example:

Six accounts reset passwords and transferred to one payee within an hour.

Industry Use:

Credential compromise leading to theft.

Detection Techniques:

Login IP mismatch, beneficiary anomaly.

Risk Indicators:

Sudden profile change, same payee.

Source: https://www.fincen.gov

Transaction Velocity Anomalies

Example:

New account does 100 UPI transactions in 40 minutes.

Industry Use:

Bot or UPI spam fraud.

Detection Techniques:

Velocity thresholds, graph analysis.

Risk Indicators:

Too many UPI hits in short time.

Source: https://www.npci.org.in

SiriusAI

ISOLATION FOREST WITH CODES

By :- Parth Tyagi

June 19, 2025

Isolation Forest

Parameters

n_estimators: int, default=100

Why It Matters:

- Each tree helps randomly isolate data points.
- The model calculates the average path length across all trees to determine if a point is anomalous.

Value	Behavior
Small (e.g. 10)	Fast training, but less accurate scores
Medium (100)	Good balance (default)
Large (500+)	More stable results, but slower

More trees → better generalization, but also more computation.

max_samples

Each tree is trained on a subset of your data.
Smaller subsets → faster and help detect rare outliers.

Value Type	What Happens
"auto"	Use min(256, total samples) → safe & fast default
int (e.g. 100)	Use exactly 100 samples for each tree
float (e.g. 0.5)	Use 50% of the dataset (0.5 * total samples) per tree

contamination

contamination = 'auto' or float (e.g., 0.05)

The model what fraction of the data is expected to be anomalies (outliers).
It is used to set the threshold for labeling points as:
-1 → anomaly
1 → normal

Value	What it means
'auto'	Let the model decide threshold (based on data, good for unknown cases)
float (e.g., 0.1)	You tell the model: “Expect 10% of the data to be outliers”

max_features

max_features = int or float, default=1.0

Type	What it does
int	Use exactly that number of features (e.g., max_features=3)
float	Use a fraction of total features (e.g., 0.5 = 50% of all features)
1	Use **all** features (default and fastest)

bootstrap & n_jobs

bootstrap

bool, default=False
bootstrap: If True, trees use data with replacement (some rows can repeat); if False (default), each row is used only once per tree.

n_jobs:int, default=None

Number of CPU cores to use; None = 1 core, -1 = use all available cores.

random_state , verbose & warm_start

random_state : int,

Random_State instance or None, default=None
Controls randomness in selecting features and split points for each tree.
Set an int (e.g., 42) to get the same results every time you run the model.

verbose : int, default=0

Controls the verbosity of the tree building process. Controls how much progress info is printed while building trees — 0 means silent, higher numbers show more details.

warm_start: bool, default=False

When set to True, reuse the solution of the previous call to fit and add more estimators to the ensemble, otherwise, just fit a whole new forest.

Thank you