Ideation: GenAI-based Fraud Detection for SecureBank

Team Hackaholics

12 June 2025

Abstract

This document outlines a GenAI-based fraud detection solution for SecureBank, targeting: (1) transaction-based fraud via a Random Forest classifier and per-user GMM-based activity anomaly detection; (2) KYC verification with liveliness detection and AI-driven record checks; (3) insider/bad-employee fraud via activity-log-based Random Forest classification. When suspicion arises, an AI agent retrieves relevant data and prepares a summary for human review.

1 Overview

The solution comprises three modules:

- **Transaction-based Fraud**: Random Forest on given transaction features + personalized GMM clustering per user.
- **KYC Verification**: Liveliness detection via webcam + ID image check; AI agent calls functions to verify records from data sources.
- **Insider Detection**: Random Forest on employee activity logs (role, login time, duration, accounts handled, etc.).

On any flagged case, an AI agent fetches relevant records for that user or employee and generates a concise summary for a human official.

2 Transaction-Based Fraud

2.1 Classifier Features

Train a Random Forest classifier using exactly the provided fields:

Table 1: Transaction Features

Feature	Description / Purpose
trans_date_trans_time	Timestamp of transaction
cc_num	Credit-card or account identifier
merchant	Merchant name/ID
category	Merchant category or transaction type
amt	Transaction amount
first, last	Customer name identifiers (or hashed)
gender	Customer gender
street, city, state, zip	Customer address or location fields
lat, long	Customer geolocation
city_pop	City population (contextual)
job	Customer occupation
dob	Customer date of birth
trans_num	Unique transaction identifier
unix_time	Numeric timestamp
merch_lat, merch_long	Merchant geolocation

2.2 Personalized Activity Detection

- For each user, model recent activity as a fixed-dimension vector (e.g., spend patterns, time patterns, category distribution, location patterns, transaction velocity).
- Fit a Gaussian Mixture Model (GMM) on the historical activity vectors of that user.
- For each new transaction, update the vector and compute likelihood under the user's GMM.
- If likelihood indicates an anomaly, mark the activity as suspicious.
- This per-user GMM update is fast and scalable, enabling on-the-fly detection.

2.3 Integration and Workflow

- Compute Random Forest fraud probability and GMM anomaly signal.
- If either exceeds its threshold, flag the transaction.
- Trigger the AI agent to retrieve that user's recent transactions and profile data, then prepare a summary report for human review.

3 KYC Verification

3.1 Liveliness and ID Check

- **Liveliness Detection**: Prompt user via webcam for simple gestures or movements; verify real person.
- **ID Capture**: Ask user to show an ID document; capture image(s).
- **AI Agent Verification**:

- Perform OCR/extraction on captured ID.
- Via function calls, check extracted fields against internal/external data sources.
- Summarize results: e.g., “Liveliness passed; ID fields match source records; verification OK” or “Discrepancy found: escalate.”

3.2 Workflow Table

Table 2: KYC Verification Steps

Step	**Action**
1. Liveliness Prompt	User performs movement; model checks for real presence.
2. ID Presentation	User shows ID to camera.
3. OCR & Extraction	Extract name, DOB, ID number, photo.
4. Data Source Checks	AI agent calls functions to verify against records.
5. Summary	AI agent composes concise approval or escalation note.

4 Insider / Bad-Employee Detection

4.1 Employee Activity Features

Collect exactly these fields from logs for modeling:

Table 3: Employee Activity Features

Feature	**Description**
role	Employee role or department
login_time	Timestamp of login
duration	Session duration
accounts_handled	Number of accounts accessed
(other logs)	Any additional logged actions if available

4.2 Random Forest Classifier

- Train a Random Forest on labeled data (benign vs. insider misuse) using the above fields.
- If an employee session or aggregated activity is flagged, trigger the AI agent.
- AI agent fetches relevant logs for that employee and generates a summary for review.

5 AI Agent for Summarization

- **Trigger**: Any flag from transaction RF, GMM anomaly, KYC mismatch, or insider RF.
- **Function Calls**: Retrieve from database:
  - For user: recent transaction history, profile info.
  - For KYC: prior verification records.
  - For employee: recent activity logs.
- **Output**: A concise natural-language summary report indicating key facts and recommendations, to assist a human official.

6 Conclusion

We present a focused ideation: use Random Forest classifiers for transaction-based and insider fraud using only the specified features; personalized GMM clustering per user for activity anomaly detection; KYC via liveliness plus AI-driven verification; and an AI agent summarizing flagged cases.