**VPBank Technology Hackathon 2025**

General Brief

Please fill up this table and use this document as a template to write your proposal.

|  |  |
| --- | --- |
| **Challenge Statement** | Write your Challenge Statement Title here |
| **Team Name** |  |

**Team Members**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Full Name** | **Role** | **Email Address** | **School Name**  **(if applicable)** | **Faculty / Area of Study** | **LinkedIn Profile URL** |
| Member 1 |  |  |  |  |  |
| Member 2 |  |  |  |  |  |
| Member 3 |  |  |  |  |  |
| Member 4 |  |  |  |  |  |
| Member 5 |  |  |  |  |  |

**Content Outline**

|  |  |
| --- | --- |
| Page No. | |
| **Solutions Introduction** |  |
| **Impact of Solution** |  |
| **Deep Dive into Solution** |  |
| **Architecture of Solution** |  |

**Solutions Introduction**

Our **AI-Powered Real-Time Banking Fraud Detection System** leverages an **Adaptive Variable Integration Module** to detect fraud, including account takeovers and money laundering, while autonomously adapting to new patterns without frequent manual updates. Using an **Isolation Forest** model, it processes transactions in near real-time (NRT) via **AWS Kinesis Data Streams**, with **Amazon SageMaker** for model training and **AWS Step Functions** for automated actions. A **React-based dashboard** on **AWS Amplify** visualizes insights **(LINH)**. Synthetic data with rule-based flagging (e.g., Transaction\_Amount\_VND > 300M VND) is used for the hackathon, but the system is designed to train on real banking data in production, with AWS enabling scalability when deployed.

### Main Features:

1. **NRT Processing**: Handles transactions in 5-minute batches via Kinesis.
2. **AI-Driven Detection**: Uses Isolation Forest for fraud detection, trainable on real data.
3. **Adaptive Integration**: Incorporates user-suggested variables (e.g., IP geolocation) to stay current.
4. **AML Capabilities**: Detects money laundering with features like Transaction\_Chain\_Length.
5. **Automated Actions**: Triggers responses (e.g., account freezing) via Step Functions.
6. **Interactive Dashboard**: React-based visualization on Amplify **(LINH)**.
7. **Explainability** (RECONSIDERATION): Optional LLM explanations, under review.

[Data Flow Diagram showing Kinesis ingestion, SageMaker AI detection, and Amplify dashboard]

**Impact of Solution**

### Benefits to Society and Target Audience

* **Security**: Protects customers from fraud, building trust in VPBank.
* **Compliance**: Supports AML/KYC with robust detection.
* **Customer Experience**: Minimizes false positives for seamless transactions.
* **Efficiency**: Automates fraud detection, reducing manual effort.

### Why Our Solution is Superior

Unlike competitors’ AI systems requiring regular manual updates, our **Adaptive Variable Integration Module** autonomously incorporates user feedback and new fraud patterns, ensuring relevance with minimal maintenance. Synthetic data uses rule-based flagging for hackathon training, but the AI-driven core is built for real historical data, making it practical for VPBank. NRT processing via Kinesis offers efficiency, and AWS ensures scalability in production.

### Competitive Advantage

* **Low-Maintenance Adaptability**: Evolves autonomously, reducing update frequency.
* **AML Focus**: Targets money laundering with specialized features.
* **Actionable**: Executes real-time preventive actions.
* **Customizable Visualization**: React dashboard on Amplify **(LINH)**.
* **Practical Design**: Adapts to real banking data with AWS scalability.

| **Aspect** | **Traditional Approach (Vietnam Context)** | **Static AI Fraud Detection** | **Adaptive AI Fraud Detection** |
| --- | --- | --- | --- |
| **Detection Method** | Manual reviews by staff + rule-based flagging (e.g., transactions > 300M VND) | Predefined ML models with fixed patterns | Continuously learning ML models with user feedback (e.g., IP geolocation) |
| **Adaptability** | Low; relies on manual updates and static rules | Moderate; requires periodic retraining | High; evolves autonomously with new data |
| **False Positives** | High; frequent due to rigid thresholds | Moderate; depends on initial training data | Low; refines over time, reducing errors |
| **Speed** | Slow; delayed by manual checks | Real-time; limited to initial model patterns | Real-time; quickly adapts to new fraud patterns |
| **Maintenance** | High; labor-intensive and rule updates needed | Moderate; periodic model updates required | Low; self-adapting, minimal manual effort |

**Deep Dive into Solution**

### Solution Overview

The system ingests transactions via Kinesis, detects fraud using Isolation Forest, and triggers actions via Step Functions. The **Adaptive Variable Integration Module** ensures the model evolves with user-suggested variables, offering a competitive edge. A React-based dashboard on Amplify visualizes results **(LINH)**. Synthetic data with rule-based flags (e.g., High\_Frequency\_Flag) is used for the hackathon, with the system designed for real banking data.

### Level 2 Data Flow

1. **Ingestion**: Synthetic data streams into Kinesis in NRT batches, adaptable to real data.
   * Features: Transaction\_ID, Transaction\_Amount\_VND, Device\_ID, Cross\_Border\_Flag.
2. **Processing**: Lambda enriches synthetic data with flags; real data skips this.
3. **Detection**: Isolation Forest assigns anomaly scores, rules used only for synthetic labels.
4. **Adaptation**: Users suggest variables via dashboard **(LINH)**, integrated after evaluation.
5. **Actions**: Step Functions triggers actions (e.g., account freezing) if scores exceed thresholds.
6. **Visualization**: Data stored in S3, visualized on Amplify dashboard **(LINH)**.
7. **Retraining**: SageMaker updates model with new data and variables.

[Level 2 Data Flow Diagram showing Kinesis, SageMaker, Step Functions, and Amplify]

### Adaptive Variable Integration Module

This module ensures the system stays ahead of fraud without frequent manual updates.

#### Why It’s Unique

Competitors’ AI systems rely on fixed features, requiring manual retraining for new fraud tactics. Our module dynamically integrates user-suggested variables (e.g., browser fingerprint) and trend-identified attributes, maintaining accuracy with minimal intervention. This is ideal for banking, where fraud evolves rapidly and proprietary data access is limited.

#### Workflow

1. **Monthly Trend Update**:
   * Collects confirmed fraud cases from user reports or historical data.
   * Retrains Isolation Forest model in SageMaker, evaluates performance, deploys if improved, notifies via SNS.
   * Pseudocode: Collect cases, retrain model, evaluate, deploy, notify.
2. **Variable Suggestion & Evaluation**:
   * **Capture**: Users submit variables (e.g., “IP geolocation”) via dashboard **(LINH)**.
   * **Normalization**: NLP standardizes variables (e.g., “login location” ≈ “GeoIP”).
   * **Testing**: Shadow model in SageMaker evaluates impact on AUC/precision/recall.
   * **Approval**: Admins approve/reject via dashboard **(LINH)**.
   * **Integration**: Retrains model with approved variables, notifies users **(LINH)**.
   * Pseudocode: Submit variable, normalize, test in shadow model, approve, retrain, notify.
3. **Variable Management**:
   * Stores metadata (submitter, approval history) in S3.
   * Auto-suggests existing variables to avoid duplication.
   * Tracks performance impact.
   * Pseudocode: Store metadata, suggest variables, update history.

### C:\Users\phong\Downloads\diagram-export-6-28-2025-10_06_28-AM.png

### Supporting Features

* **Explainability** (RECONSIDERATION): Optional LLM explanations, under review.
* **Adaptability**: Synthetic data for hackathon, real data for production.
* **Explainable AI**: SHAP analysis highlights variable importance.
* **Error Handling**: Step Functions retries failed actions.

**Architecture of Solution**

### AWS Infrastructure

The system uses AWS for an efficient, AI-driven solution, optimized for hackathon constraints but scalable in production with AWS.

### Architecture Diagram

### C:\Users\phong\Downloads\diagram-export-6-28-2025-9_45_59-AM.pngAWS Services

1. **AWS Kinesis Data Streams**: Ingests synthetic data in NRT batches, adaptable to real data.
2. **AWS Lambda**: Enriches synthetic data, invokes SageMaker, uses minimal permissions.
3. **Amazon SageMaker**: Trains Isolation Forest model, supports retraining and shadow models.
4. **AWS Step Functions**: Orchestrates actions and variable approvals.
5. **Amazon SNS**: Sends fraud and update alerts.
6. **Amazon S3**: Stores data and metadata.
7. **AWS Amplify**: Hosts React dashboard, customizable by Linh **(LINH)**.
8. **Amazon EC2** (optional): Supports Amplify if needed **(LINH)**.

### Integration

* **Kinesis-Lambda-SageMaker**: Enables efficient AI detection.
* **Lambda-Step Functions-SNS**: Triggers rapid actions and alerts.
* **S3-Amplify**: Stores data, hosts dashboard **(LINH)**.
* **SageMaker-Step Functions**: Manages adaptive retraining.