Capstone Project Concept Note and Implementation Plan

Anomaly Detection in Financial Transactions

Group 14 Members

1. Ayana Damtew Tajebe
2. Beamlak Dejene
3. Anwar Mohammed Koji
4. Anwar Gashaw Yimam
5. Amen Zelealem Tadese

Concept Note

1. Project Overview

This capstone project addresses the urgent problem of money laundering, which enables crime, corruption, and economic instability. Traditional rule-based detection systems have high false positive rates and struggle to adapt to evolving laundering strategies. Our project proposes an AI-driven real-time anomaly detection system leveraging machine learning and deep learning techniques, aligned with Sustainable Development Goals (SDG 16: Peace, Justice and Strong Institutions, and SDG 8: Decent Work and Economic Growth). The project aims to significantly improve the detection of illicit financial activities, reduce false alarms, and promote financial system transparency.

2. Objectives

- Develop a real-time system for detecting suspicious financial transactions.
- Reduce false positives compared to traditional rule-based systems.
- Utilize transaction metadata and contextual information (e.g., geopolitical data) for improved accuracy.
- Ensure data privacy and regulatory compliance through federated learning methods.

3. Background

Financial crimes like money laundering are increasingly sophisticated, making traditional detection systems less effective. Although solutions exist using rule engines, they fail to catch novel fraud patterns. Machine learning models, particularly those using unsupervised learning and deep learning, offer scalable, adaptive, and data-driven alternatives. Our system builds upon recent research in graph-based deep learning and federated learning to provide a more robust and privacy-preserving solution .

4. Methodology

- **Isolation Forest**: For unsupervised anomaly detection.
- **Graph Neural Networks (GNNs)**: To detect complex transactional relationships.
- **Federated Learning**: To enable cross-institution training while preserving data privacy.
- **Preprocessing**: Data normalization, anonymization, and feature engineering for improved model input quality.
- **Real-time processing**: Using streaming techniques and lightweight models.

5. Architecture Design Diagram

- **Data Ingestion Layer**: Collects real-time transaction data streams.
- **Preprocessing Layer**: Normalizes, encodes, and anonymizes incoming data.
- **Anomaly Detection Engine**: Runs Isolation Forest and GNN models to flag suspicious activities.

- **Federated Learning Coordinator**: Facilitates model training across multiple banks without sharing raw data.
- **Alert System**: Notifies financial compliance officers for further investigation.

6. Data Sources

We use synthetic datasets such as PaySim and AML-Bench, and publicly available datasets like the Kaggle Credit Card Fraud dataset. These datasets include transaction amounts, timestamps, sender/receiver IDs, and location data. Preprocessing includes normalization of transaction amounts, anonymization of personal identifiers, and encoding of transaction types for model readiness.

7. Literature Review

The literature supports the move from simple statistical approaches to machine learning and deep learning techniques for fraud detection. Research by Chandola et al. (2009) and Ahmed et al. (2016) highlights the advantages of clustering and classification models. Recent studies show the effectiveness of GNNs and federated learning in uncovering complex fraud patterns and ensuring data privacy.

Implementation Plan

1. Technology Stack

- **Programming Language**: Python
- **Libraries**: Scikit-learn, TensorFlow, Keras, PySyft, Pandas, NumPy
- **Frameworks**: Flask (for API), Apache Kafka (for streaming real-time transactions)
- **Visualization**: Matplotlib, Seaborn
- **Version Control**: Git, GitHub
- **Deployment**: AWS EC2 or Heroku

2. Timeline

Task	Start Date	End Date
Data Collection & Preprocessing	May 1, 2025	May 10, 2025
Baseline Model Development (Isolation Forest)	May 11, 2025	May 20, 2025
Advanced Model (GNN) Implementation	May 21, 2025	June 5, 2025
Federated Learning Setup	June 6, 2025	June 15, 2025
Model Evaluation and Tuning	June 16, 2025	June 22, 2025
Deployment and API Integration	June 23, 2025	June 30, 2025
Final Report and Presentation	July 1, 2025	July 5, 2025

3. Task Distribution Matrix

Team Member	Responsibility
Ayana Damtew Tajebe	Data Preprocessing, Visualization
Anwar Gashaw Yimam	Isolation Forest Model Development
Amen Zelealem Tadese	Graph Neural Network Implementation
Anwar Mohammed Koji	Federated Learning Setup
Beamlak Dejene	Deployment, Documentation, and Integration

4. Milestones

- Completion of preprocessing pipeline.
- Isolation Forest model achieving >80% anomaly detection recall.
- GNN model identifying complex laundering rings.
- Federated training across multiple synthetic nodes.
- Deployment of a real-time API that flags suspicious transactions.

5. Challenges and Mitigations

Challenge	Mitigation
Class imbalance in data	Use anomaly detection instead of direct classification.
Real-time latency constraints	Optimize model size and use batch processing for GNNs.
Data privacy issues	Employ federated learning and data anonymization.
Model interpretability	Integrate explainable AI methods (e.g., SHAP, LIME).

6. Ethical Considerations

- **Data Privacy**: We ensure full compliance with privacy regulations by using anonymized datasets and federated learning.
- **Bias Mitigation**: Regularly audit models for bias against any user group.
- **Impact Assessment**: Ensure flagged transactions are reviewed by human compliance officers to prevent wrongful actions based on automated systems.

7. References

- Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys (CSUR)*, 41(3), 1–58.
- Bolton, R. J., & Hand, D. J. (2002). Statistical fraud detection: A review. *Statistical Science*, 17(3), 235–255.
- Ahmed, M., Mahmood, A. N., & Hu, J. (2016). A survey of network anomaly detection techniques. *Journal of Network and Computer Applications*, 60, 19–31.
- Bhatia et al. (2023). Graph neural networks for anti-money laundering. *arXiv:2305.12345*.
- Liu et al. (2024). Federated learning for privacy-preserving AML detection. *IEEE Access*.
- Kaggle. (2016). *Credit Card Fraud Detection Dataset*.
