Project Proposal

Project Title

Detecting Anomalous Transactions for AML Using LSTM Autoencoders

Abstract

This project explores the application of deep learning techniques to the problem of anomaly detection in financial transactions, a key requirement in Anti-Money Laundering (AML) efforts. We propose using an LSTM-based autoencoder to learn normal transactional patterns and flag deviations that may indicate fraudulent or suspicious behavior. Unlike traditional rule-based systems, this model captures complex sequential dependencies across time, offering a robust and adaptive alternative for transaction monitoring.

LSTM (Long Short-Term Memory)

It is a type of **Recurrent Neural Network (RNN)** architecture used in deep learning, particularly suited for **sequence prediction problems**. Unlike traditional RNNs, LSTMs are designed to **learn long-term dependencies** and avoid problems like **vanishing gradients**, making them highly effective for tasks involving **time-series data**, **natural language processing**, and **transaction sequences**.

Objective

Develop an LSTM autoencoder to:

- Learn the normal sequence behavior of financial transactions.
- Identify anomalies without relying on labeled fraud data.
- Demonstrate its potential to flag money laundering-like behaviors in synthetic or anonymized transaction data.

Scope

- Focused on unsupervised learning, suitable for scenarios with limited labeled fraud data.
- Transactions modeled as time series sequences for individual accounts or user profiles.
- Visualization and evaluation of detected anomalies.

Deep Learning Techniques

- LSTM (Long Short-Term Memory) networks: for capturing temporal dependencies in transaction sequences.
- Autoencoders: for learning compressed representations and reconstructing input.
- Reconstruction Error: used to detect anomalous behavior based on thresholding.

Tools & Libraries

- Python 3.x
- TensorFlow/Keras – model building and training
- Pandas, NumPy – data preprocessing and sequence creation
- Matplotlib/Seaborn – visualization of loss, anomalies, and trends

Datasets

Options include:

1. PaySim (Synthetic Financial Data) – designed to simulate mobile money transactions.
   GitHub: https://github.com/EdgarSantosA/paySim
2. Kaggle Credit Card Fraud Detection – anonymized dataset with known outliers.
   https://www.kaggle.com/mlg-ulb/creditcardfraud

Methodology

1. Data Preprocessing
   - Normalize numerical values
   - Create sequences per account/user
   - Split into training (normal behavior) and test sets

2. Model Building
   - Design an LSTM autoencoder in Keras
   - Train on normal data to minimize reconstruction loss

3. Anomaly Detection
   - Calculate reconstruction error on unseen data
   - Flag sequences with error above a threshold as anomalous

4. Evaluation & Visualization
   - Plot loss curves and anomalies over time
   - Analyze flagged sequences and interpret patterns

Expected Outcomes

- A trained LSTM autoencoder capable of detecting outlier transaction patterns
- Reconstruction error distribution visualized and interpreted
- Practical demonstration of how deep learning can assist AML systems without labeled fraud data

Deliverables

- Python notebooks/scripts
- Trained model files
- Anomaly detection report with plots and analysis
- Summary presentation or report document

Conclusion

This project leverages the power of unsupervised deep learning to enhance AML efforts by detecting unusual transaction patterns using LSTM autoencoders. It serves as a practical step toward data-driven financial fraud detection in real-world systems.