CYBERHACK 2025

INNOVATE – SECURE – PROJECT

"AI-Powered Detection of Rented Bank Accounts for Financial Fraud Prevention Using Machine Learning"

By- Suhani Gahukar, Purva Baghel, Poorva Pohekar

1. INTRODUCTION :

Financial fraud is becoming a serious threat to banks and financial institutions. Rented bank accounts are one of the new threats, in which fraudsters obtain access to genuine accounts to carry out illegal operations such as money laundering, scams, and unauthorized transactions. The rented accounts look authentic, so it becomes hard for conventional fraud detection mechanisms to detect them. The effects of such fraud are severe, causing financial losses to banks, loss of customer trust, and regulatory fines.

In order to fight against this problem, artificial intelligence (AI) and machine learning (ML) provide effective tools to detect fraud. Through the observation of transaction behaviors, login activity, and other important indicators, ML models are able to identify suspicious activities indicating an account is being abused. In contrast with rule-based detection systems, AI-driven methods are capable of responding to new methods of fraud and enhancing accuracy in the long term. This report discusses how machine learning can effectively be utilized to detect and prevent the abuse of rented bank accounts.

2. PROBLEM STATEMENT

Fraudsters rent bank accounts from real customers to avoid detection. These accounts may belong to individuals or businesses who unknowingly or knowingly allow others to use them. Signs of rented accounts include:

- Logins from different geographical locations within short time frames
- Unusually frequent high-value transactions
- Transactions to multiple unknown or unrelated recipients
- Changes in spending patterns that do not match the account holder’s history. Since fraudsters are continuously changing tactics, AI-driven fraud detection systems must adapt and improve over time.

3. PROPOSED SOLUTION

Machine learning can help identify rented accounts by analyzing transaction behaviors. The solution involves:

- **Extracting Key Features:** Important factors like login locations, transaction frequency, device details, and transaction recipients are monitored.
- **Selecting the Right Model:** Different ML models, such as Random Forest, XGBoost, and Neural Networks, are trained to recognize fraud patterns.
- **Anomaly Detection:** Unsupervised learning models, such as Isolation Forests and Autoencoders, identify unusual activities.
- **Behavioral Analysis:** A user's past activity is compared to new transactions to detect inconsistencies.
- **Real-Time Monitoring:** Suspicious activities trigger alerts, allowing banks to investigate and prevent fraud.

4. DATASET AND PREPROCESSING

A dataset containing real or synthetic bank transaction records is needed for training and testing the models. The data must be cleaned and prepared using the following steps:

- **Data Cleaning:** Removing duplicate entries, handling missing values, and correcting inconsistencies.
- **Data Transformation:** Converting data into a structured format that the ML model can process efficiently, such as scaling numerical values and normalizing transaction amounts.
- **Feature Selection:** Identifying the most relevant attributes, such as transaction frequency, location, and device information, to improve model accuracy and performance.
- **Encoding Categorical Data:** Converting non-numerical information like account type and transaction type into a machine-readable format using techniques such as one-hot encoding or label encoding.
- **Handling Imbalanced Data:** Applying techniques like oversampling fraud cases or under sampling legitimate transactions to ensure that the model learns effectively.
- **Splitting Data:** Dividing the dataset into training and testing subsets to evaluate model accuracy and prevent overfitting.

5. MACHINE LEARNING MODEL IMPLEMENTATION

Random Forest Classifier was used for identifying fraudulent transactions in rented bank accounts.

- Random Forest is an ensemble learning method, which creates multiple decision trees and aggregates their output to improve the accuracy and prevent overfitting.
- It is noise robust and performs well on tabular transaction data.

- Training and Testing Strategy
    - Data Preprocessing:
        - Dropped unnecessary columns (if any).
        - Encoded categorical variables using LabelEncoder().
        - Standardized numerical features with StandardScaler().
    - Data Splitting:
        - The dataset was split into 80% training and 20% testing using train_test_split().
        - A random_state=42 was set to ensure reproducibility.
    - Model Training:
        - A Random Forest Classifier with 100 estimators was trained.
        - The model learned transaction patterns to differentiate between fraudulent and legitimate rented accounts.
    - Predictions:
        - The trained model was used to predict fraudulent transactions in the test set.

6. EVALUATION METRICS

- Accuracy: Overall correctness of the model.
- Precision: Number of true positives over the total number of predicted fraud cases.
- Recall (Sensitivity): Number of actual fraud cases identified.
- F1-Score: A trade-off between precision and recall.
- Confusion Matrix: Contains values of true positive, true negative, false positive, and false negative.
- Output of the Code
- Accuracy: Calculated and displayed by using accuracy_score(y_test, y_pred).
- Classification Report: Reports precision, recall, and F1-score.
- Confusion Matrix: Shows false positive and false negative.

7. IMPLEMENTATION CHALLENGES AND SOLUTIONS

1. Data Privacy and Security Concerns

    - Data related to a financial transaction is sensitive; so proper anonymization techniques must be used.
    - Encryption and data handling protocols must be properly followed.
    - Compliance with regulations like GDPR and PCI DSS.

2. Model Interpretability and Bias Issues

Challenge: Random Forest is an ensemble method, which leads to lower interpretability

Solution:

- Feature importance can be used to derive most important fraud indicators.

- SHAP (SHapley Additive exPlanations) values can be used to explain the prediction of a model.
- Bias in the training data should be identified to not discriminate against lawful users.

8. CONCLUSION

The proposed fraud detection model based on the Random Forest Classifier successfully identifies patterns in rented bank accounts used for financial fraud. With the help of machine learning techniques, we have been able to achieve a high level of accuracy while minimizing false positives. The model was successful in distinguishing fraudulent transactions based on behavioral patterns such as multiple logins from different locations and unusual transaction frequencies. Data preprocessing steps such as feature scaling and categorical encoding also helped improve the model's predictive performance.

But data privacy, model interpretability, and the continuous development of new fraudulent schemes remain significant challenges. There is a strong need for developing explainable AI and safe practices in data handling for successful real-world banking implementation.

Future Work

Model Enhancement:

- Implement the LSTM network in the time series analysis of transactions.
- Utilize unsupervised learning with Autoencoders in the identification of fraud patterns unknown so far.

Real-time fraud detection system

- Implement the model as an online fraud detection system to check transactions in real-time.
- Leverage big data streaming platforms such as Apache Kafka for managing high-volume financial data.

Explanation and Bias

- Incorporate SHAP values to understand the importance of features in fraud detection.
- Balance datasets and fairness-aware algorithms to decrease model bias.

Blockchain Integration

- Apply blockchain-based identity verification to avoid fraudulent account rentals.
- Make use of tamper-proof transaction logs to increase transparency and accountability.

Federated Learning for Secure Model Training

- Train fraud detection models across different banks using Federated Learning in order to maintain the privacy of data.
- Cooperate with financial institutions to create a global fraud detection network.