Project Title: Mule Account Identification

WHAT ARE MULE ACCOUNTS?

- Mule accounts, also known as money mule accounts, are bank accounts used by criminals to launder illicit funds. They are often set up by unsuspecting individuals lured by promises of easy money or coerced into participation. The person who opens and uses the account, known as a "money mule," may or may not be aware that their account is being used for illegal activities

WHY IS THE MONEY MOVED?

- The funds moved by these money mules are the proceeds of criminal activity. By transferring this money through a network of mule accounts, the money mules launder the funds so that they can be re-integrated into legitimate economy.

HOW ARE MULE ACCOUNTS CREATED?

There are various ways by which a mule account can be created and they are listed as follows:

1. Recruitment:

Promises of Easy Money:

Scammers target individuals with promises of easy income or job opportunities, often disguised as legitimate business ventures.

Fake Job Offers:

Online ads or social media postings may create the illusion of legitimate job openings, leading unsuspecting individuals to become mules.

Direct Contact:

Scammers may directly contact individuals, claiming to need help with a business transaction or payment testing, and requesting their bank account details.

2. Fund Collection:

Depositing Stolen Funds:

Once the mule account is set up, stolen or illicit funds are deposited into the account, often through wire transfers or online payment methods.

Source of Funds:

These funds can originate from various scams, fraud, or stolen credit cards.

3. Money Movement:

Transferring Funds:

The scammer instructs the mule to transfer the money to another account or cryptocurrency wallet, sometimes across international borders.

WHAT ARE DIFFERENT TYPES OF MONEY MULES?

1. Unwitting participants

These individuals are often unaware that they are part of a criminal scheme. They might be lured through job advertisements, online friendships, or romantic relationships, believing they assist in legitimate business activities.

2. Witting participants

These participants know they are part of an illegal operation but might not know the full extent of the criminal activity. They are often motivated by financial incentives or are under some form of duress.

3. Complicit criminals

These are individuals or entities fully aware of their involvement in criminal activities. They are an integral part of the criminal network, often orchestrating the laundering process.

WHAT ARE THE WAYS OF IDENTIFYING A MULE ACCOUNT?

Identifying mule accounts involves detecting unusual transaction patterns and other indicators of potential money laundering activity. Banks and financial institutions use various methods, including rule-based monitoring, behavioral biometrics, and AI-powered solutions, to flag suspicious accounts.

- Unusual Transaction Patterns:
  Sudden spikes in activity, especially in dormant or newly opened accounts, are a red flag. Frequent, large deposits and withdrawals, or rapid movement of funds with little or no balance retention, also indicate potential mule activity.

- Location Behavior:
  Discrepancies between the customer's declared address and the location where they access their account can be a sign of a mule account.

- Account Information Changes:
  Frequent updates to account details, such as address, email, or phone number, can suggest attempts to assume a false identity.

- Fragmented Transactions:
  Breaking large transactions into smaller ones to stay below reporting thresholds is a common tactic used to avoid detection.

- Cross-Border Transactions:
  Transactions involving high-risk jurisdictions or cross-border transfers can be a sign of money mulling.

- Multiple Accounts Under a Single Identity:
  Fraudsters often create multiple mule accounts using the same identity or stolen credentials to spread their activities

- Inbound Payment Monitoring:
  Monitoring incoming payments to identify unusual transactions or patterns.

- Behavioral Biometrics:
  Analysing how customers interact with their devices and accounts to identify anomalies in their behaviour, such as login patterns or screen touch patterns.

- AI-Powered Detection:
  Utilizing AI to analyse large volumes of transaction data and identify patterns that are difficult for humans to detect.

- Graph Visualization:
  Using graph technology to visualize relationships between customers, transactions, and accounts to identify potential links between mule accounts.

- External Data Source:
  Leveraging data from social media and other sources to gain a more comprehensive understanding of customer behaviour and activity.

- Real-time Due Diligence:
  Implementation of real-time checks and verification procedures to identify potential mule accounts as they are opened.

SOME PLATFORMS WHICH CAN BE USED TO DETECT MULE ACCOUNTS AND FORM A FRAMEWORK ON

- Supervised Machine Learning

  Algorithms: Random Forest, XGBoost, Logistic Regression, Neural Networks

- Unsupervised Anomaly Detection

  Techniques: Isolation Forest, Autoencoders, DBSCAN, One-Class SVM

- Graph Based Analysis

  Tools: Matplotlib

Since here at ICICI Bank, we already have a current framework to identify mule accounts which is made using Supervised Machine Learning- Logistics Regression. It is performed using a mathematical Sigmoid Function and a principle of Maximum Likelihood. It optimises Cosine Function using gradient decent or specialization solver method which decides the category of variables present in the function which gives out the score on the basis of different parameters of the customer like personal information, transactions (incoming and outgoing), personal relational with the bank and more. On the basis of these parameters the model displays a score ranging from 1 to 10 using banding which shows a certain level of risk that a particular account portrays or holds.

I will be using Unsupervised Anamoly Detection (used to detect outliers which are important for a problem statement) to identify mule accounts within a DataSet which can be efficient and give out more accurate results than the current framework. It uses a Mathematical formula - S(m, n) to compute an anomaly score and a threshold value is given, above which the data is declared as an outlier. Formation of an isolated tree method is used in which the average search of depth for the probability of becoming a data point is found out from the isolated tree and then the average depth of all the data points is found out.
If E(h(x))<<c(m) = S(m, n) tends to 1 then it is declared as an outlier.
The output values will be in the form of 1 (normal data point) and -1 (outlier).

Alternative Comparison can also be done using the given methods of Unsupervised Anamoly Detection:
One-Class SVM
DBSCAN
Local Outlier Factor (LOF)

DETECTION OF MULE ACCOUNTS USING UNSUPERVISED ANOMALY DETECTION: ISOLATION FOREST

This report presents a data-driven approach to detect mule accounts, bank or wallet accounts used to facilitate money laundering or fraudulent transactions using the Isolation Forest algorithm. This unsupervised learning method helps identify anomalous behaviour without requiring labelled data.

Objective:
To identify anomalous account behaviours that may suggest mule activity.
Implement a scalable, unsupervised machine learning model using Isolation Forest.
Evaluate the effectiveness of the model for flagging accounts for further investigation.

Isolation Forest is an anomaly detection algorithm based on the principle that anomalies are:

- Few and different.
- Easier to isolate from the rest of the data.

Key benefits:

- Unsupervised: No need for labelled data.
- Efficient: Works well with high-dimensional datasets.
- Interpretability: Anomaly scores help prioritize further manual inspection.

Feature	Name Description
account_id	Unique ID for the account
num_transactions	Number of transactions per week/month
avg_transaction_amt	Average amount transferred per transaction
unique_recipients	Number of distinct destination accounts
login_frequency	Daily/weekly login count
account_age_days	Account's age in days
location_entropy	Diversity in transaction locations
tx_time_skewness	Skewness in time of transactions
incoming_outgoing_ratio	Ratio of incoming to outgoing funds

Methodology
Data Preprocessing:

- Handle missing values.
- Normalize numerical features using Min-Max or Standard Scaler.
- Encode categorical features.
- Drop irrelevant or high-cardinality identifiers.

Evaluation Strategy
Since it's unsupervised:

- Use of domain knowledge or expert labeling to validate a sample.
- Compare Isolation Forest with other anomaly detection methods (e.g., One-Class SVM, DBSCAN).
- Track detection precision in post-analysis investigations.

Model Interpretation

- Accounts labeled 1 under is_mule and are potential mule accounts.
- Visualize with t-SNE or PCA to see cluster separation.
- Review feature distributions for flagged accounts.

Results

- Percentage of accounts flagged as anomalies: tending to 2% (based on contamination for eg=0.02)

Key anomalous traits observed:

- High number of outbound transfers.
- Very recent account creation.
- Unusual recipient diversity.

Detected accounts can be prioritized for KYC re-verification or transaction blocking.

Post-Detection Analysis

- Visualize anomalies using PCA or t-SNE.
- Compare distributions of flagged vs. normal accounts.
- Identify key distinguishing features for flagged accounts.

Conclusion

This report demonstrates the potential of Isolation Forest as an effective tool for detecting mule accounts in financial systems. By analyzing account behavior patterns and flagging anomalies, institutions can proactively intervene and reduce fraud risk. The approach is unsupervised, scalable, and interpretable making it ideal for practical deployment in fraud detection systems.

Data Source Details:

- Isolation Forest. IEEE ICDM
- Sci-kit learn documentation - https://scikit-learn.org/
- Industry Case Study on mule account fraud detection.

MODEL ANALYSIS

Dataset Description:
The dataset contains account-level transaction summaries and labels for classification:

Dataset Summary
Data Source: Provided Excel file – “Development Base – Khushi – 11June25”

Total Records: 6667

Target Column: TARGET
- GOOD: 5565 accounts (83.4%)
- BAD: 1102 accounts (16.6%)

To facilitate model evaluation:
TARGET_BINARY was created, mapping:
- 'GOOD' → 0
- 'BAD' → 1

Feature Selection Strategy
The features used in the Isolation Forest model were selected based on their individual F1 scores when tested independently for their predictive power in identifying mule accounts.

Method:
- Each feature was used individually to train a simple Isolation Forest model.
- For each model, predictions were made, and the F1 score was calculated using the actual target labels.
- The top-performing features (with the highest F1 scores) were selected for the final model.

Features Used
The following 9 transaction features were used in the model with their rationales

1.INFLOW_CNT_7D (0.6033):
- Count of incoming transactions in the last 7 days
- Strong correlation with abnormal incoming transaction behavior

2.UPI_IMPS_INFLOW_VALUE_7D (0.6806):
- Total value of UPI/IMPS inflows over 7 days
- High-value inflows are often suspicious in mule cases

3.UPI_IMPS_OUTFLOW_VALUE_7D (0.6369):
- Total value of UPI/IMPS outflows over 7 days
- Captures fund disbursement patterns

4.PER75_WITHDRWAL_NEXT1HR (0.6297):

- 75th percentile of expected withdrawals in the next 1 hour
- Indicates rapid withdrawal behavior common in mule activity

5.INFLOW_CNT_1D (0.5380):
- Count of incoming transactions in the last 1 day
- Helps spot sudden transaction spikes

6.IMPS_INFLOW_CNT_7D (0.6550):
- Number of IMPS inflow transactions in 7 days
- Adds granularity to inflow count by channel

7.IMPS_INFLOW_VALUE_7D (0.6422):
- Total IMPS inflow value over 7 days
- Focus on the monetary volume through IMPS

8.IMPS_OUTFLOW_CNT_7D (0.6277):
- Number of IMPS outflow transactions in 7 days
- Outflow count helps in balance movement analysis

9.IMPS_OUTFLOW_VALUE_7D (0.6263):
- Total IMPS outflow value over 7 days
- Total outflow value is key for anomaly scoring

Data Preprocessing
- Missing Values: Filled using median imputation (X.fillna(X.median(numeric_only=True)))
- Scaling: All features scaled using StandardScaler to normalize values before feeding into Isolation Forest.
- Feature Matrix: X_scaled prepared with selected features.

MODEL DETAILS
Model Used: Isolation Forest (unsupervised anomaly detection)

Parameters:
- n_estimators: 85
- contamination: 0.11
- max_samples: 20
- random_state: 48
- max_features: 1.0

Anomaly prediction:
- -1 → anomaly (classified as mule account)
- 1 → normal

Later on,

- -1→ anomalies were replaced to 1 (mule accounts)
- 1 → normal accounts were replaced to 0 (normal data points)

Results & Performance

Predicted Mule Accounts: 734 out of 6667

Confusion Matrix:

	Predicted Legit	Predicted Mule
Actual GOOD	**5483**	**82**
Actual BAD	**450**	**652**

Classification Metrics:

- Accuracy: 92.02%
- Precision: 88.83%
- Recall: 59.17%
- F1 Score: 71.02%

Visualization Outputs
Precision-Recall Curve

- Helps understand model performance under different thresholds.

Anomaly Score Histogram

- Red distribution of all scores.
- Vertical blue line showing 5% quantile cutoff for high suspicion zone.

Appendix: Tools & Libraries Used

- Python Libraries: pandas, numpy, matplotlib, seaborn, sklearn
- Model: IsolationForest from sklearn.ensemble
- Evaluation: precision_score, recall_score, f1_score, confusion_matrix, classification_report

THE RATIO OF GOOD ANOMALIES TO TOTAL NUMBER OF MULE ACCOUNTS IS 11.1%