Detection of Illicit Transactions and Wallets in the Bitcoin Network using Machine Learning

Agenda

- Company Overview
- Dataset Overview
- Problem Statement
- ML Models
- Insights

Presented by: Group 11

ELLIPTIC

Company Overview

- **Founded in 2013 & headquartered in London**
- A leading company in blockchain analytics and crypto compliance
- **Provides tools to help:**
  - Financial institutions
  - Crypto businesses
  - Government agencies
- **Core services include:**
  - Real-time monitoring of cryptocurrency transactions
  - Risk assessment of digital wallets
  - AML and sanctions compliance solutions
- Enables multi-blockchain tracing to identify high-risk fund flows
- Collaborates with top institutions like MIT and IBM on financial crime research
- Published open datasets (e.g, Kaggle Elliptic Dataset) for AI and machine learning in fraud detection
- Expanded global presence with a regional HQ in the UAE, serving clients across the Middle East and beyond

ELLIPTIC

Problem Statement

The rapid rise in Bitcoin transactions has led to a surge in illicit activities, including money laundering, fraud, and terrorist financing. Regulators and financial institutions struggle to detect such activities in real time due to:

- The pseudonymous nature of wallet addresses
- The massive volume and complexity of blockchain transactions
- Limited interpretability of traditional detection systems

Objective:

- Develop machine learning models to classify **illicit transactions and suspicious wallet addresses** in the Bitcoin network using the Elliptic dataset.

Dataset Overview

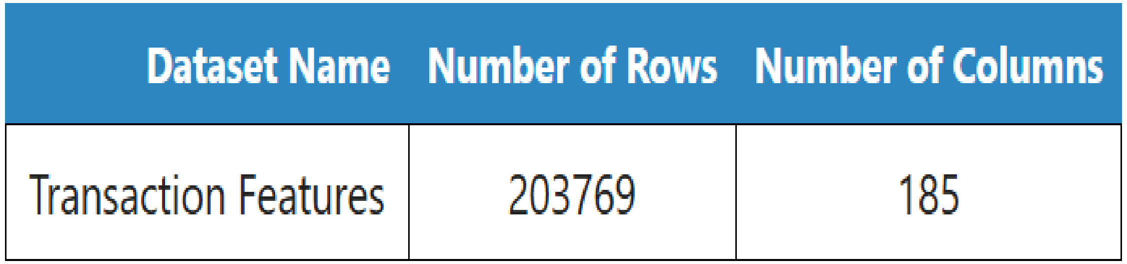
Dataset Description

Two core categories:

- **Features**
  - Transaction features contains 184 features for 203,769 transactions
  - Captures what happened in each transaction
  - Wallet features contain 56 features for 1,268,260 wallet addresses
  - Captures how wallet entities behave overtime
- **Classes**
  - Labels each transaction and wallet address as either 1 (Illicit), 2 (licit) or 3 (unknown)
  - Supports detection of suspicious transaction patterns
  - Enables identification of fraudulent wallets

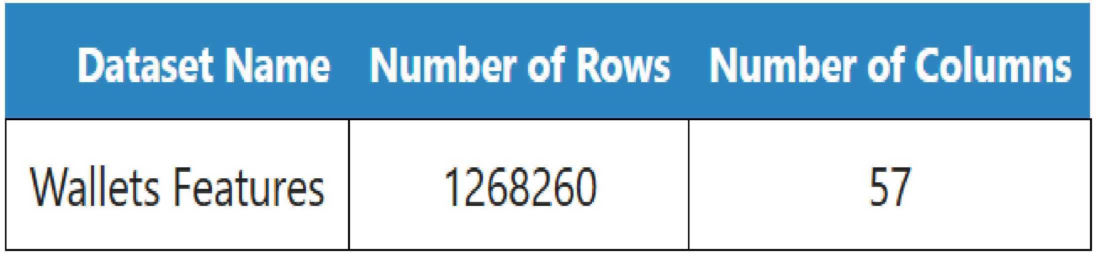
Dataset Characteristics

Transaction Features:



Dataset Name	Number of Rows	Number of Columns
Transaction Features	203769	185

Wallet Features:



Dataset Name	Number of Rows	Number of Columns
Wallets Features	1268260	57

ELLIPTIC

Machine Learning Models Used

Transaction Analysis

Model Name	Precision	Recall	F1 Score	Micro-Avg F1
Random Forest	0.965	0.719	0.824	0.980
XGBoost	0.922	0.730	0.815	0.978
LightGBM	0.608	0.740	0.667	0.951
Multilayer Perceptron (MLP)	0.622	0.597	0.609	0.949
Logistic Regression	0.323	0.704	0.443	0.883

Wallets Analysis

Model Name	Precision	Recall	F1 Score	Micro-Avg F1
Random Forest	0.909	0.780	0.840	0.989
XGBoost	0.893	0.808	0.848	0.989
Multilayer Perceptron	0.842	0.412	0.553	0.976
LightGBM	0.384	0.919	0.542	0.944
Logistic Regression	0.491	0.057	0.102	0.964

Dataset Link: https://www.kaggle.com/datasets/ellipticco/elliptic-data-set/data

ELLIPTIC

Insights

Goal: Detecting Illicit Transactions

- **Recall was the top priority,** in fraud detection missing illicit activity is riskier than flagging false positives.
- **XGBoost and Random Forest** consistently delivered the best balance of precision and recall across both transaction and wallet levels.
- **LightGBM showed exceptional recall**, making it valuable when detection sensitivity outweighs precision.
- **Logistic Regression underperformed**, especially in wallet-level tasks, highlighting its limitations for complex, imbalanced datasets.
- **We recommend XGBoost** for its strong balance of precision and recall, making it ideal for effective fraud detection.