

# AI-Based Detection of Cryptocurrency Scams Using Machine Learning

Research Paper

Pooja Mehta<sup>1</sup>, Faizan Asad Gandhi<sup>2</sup>

<sup>1</sup>Assistant Professor, D. Y. Patil Deemed to be University, School of Humanities & Sciences, Belapur, Maharashtra, India.

<sup>2</sup>Student, Department of MCA, D. Y. Patil Deemed to be University, School of Humanities & Sciences, Belapur, Maharashtra, India.

\*\*\*

**Abstract** - Cryptocurrency has changed the way money moves online, because payments are quick, need no central bank and cross borders easily. The same qualities attract fraudsters, and scams such as phishing, Ponzi schemes, fake investment projects, fraudulent wallets and malicious smart contracts are now common. Fixed, rule-based fraud checks struggle because scammers keep changing their tricks. Machine learning can go through huge amounts of blockchain data and pick up the patterns that tend to appear around suspicious activity. In this paper we propose a framework that uses transaction, wallet, smart-contract, time-based and network features to spot cryptocurrency scams. Logistic Regression, Random Forest, XGBoost, LSTM and Graph Neural Networks can be compared using accuracy, precision, recall, F1-score and ROC-AUC. Explainable AI is included so that investigators can see why a wallet or transaction was flagged. The framework is meant to help researchers, exchanges, investigators and blockchain security teams.

**Key Words:** Artificial Intelligence, Cryptocurrency, Blockchain, Scam Detection, Machine Learning, Fraud Detection, Ethereum, Bitcoin.

## 1. INTRODUCTION

A cryptocurrency is digital money whose transactions are recorded and checked on a blockchain. Bitcoin, Ethereum and many other assets are now used for payments, investing, decentralized applications and financial services. Because many blockchains are public and traceable, the way people transact can be studied in detail. The same environment is also open to abuse. Common problems include phishing, Ponzi schemes, fake token projects, fraudulent investment offers, malicious smart contracts, suspicious wallet behaviour and wash trading. Several researchers have tried machine learning to catch phishing and Ponzi schemes on Ethereum and other blockchains [1], [2].

Checking this much data by hand is not realistic, whereas AI can go through it and highlight patterns that a person would likely miss. Supervised learning can separate known fraudulent activity from legitimate activity, while graph-based and time-based methods can capture the

links between wallets and the order of transactions. More recent studies have also looked at explainable AI and large language models for finding blockchain scams [3], [4].

## 2. LITERATURE REVIEW

### 2.1 Machine Learning for Ponzi Detection

Chen et al. studied how machine learning can detect Ponzi schemes on Ethereum. They showed that transaction and contract features are useful for automatic detection, and they pointed out problems such as imbalanced classes and information leakage [1].

### 2.2 LSTM-Based Detection

Wang et al. detected Ponzi schemes with an LSTM model trained on oversampled data. Their approach used account and smart-contract information, and it showed that deep learning on sequences can help with blockchain fraud [2].

### 2.3 Explainable AI

More recent work pairs XGBoost with SHAP explanations to make smart Ponzi detection easier to interpret. This matters because investigators want to know which features led to a fraud prediction, not just receive a yes or no answer [3].

### 2.4 Graph and Advanced AI Methods

A cryptocurrency network can be viewed as a graph, with wallet addresses as nodes and transactions as edges. For this reason, researchers have tried dynamic graph neural networks and other graph methods for phishing and fraud detection. Some newer studies also mix transaction sequences, graph structures and large language models [4], [5].

## 3. PROBLEM STATEMENT

The fast growth of cryptocurrency has given fraudsters more room to operate. Systems that depend on known patterns can miss new scam behaviour. This research asks how AI and machine learning can automatically find

suspicious transactions, wallets and smart contracts by looking at how they behave, when they act and how they connect to others in the network.

#### 4. OBJECTIVES

- To study the different kinds of cryptocurrency scams.
- To find the transaction, wallet, smart-contract and network features linked with fraud.
- To build an AI-based framework for detecting cryptocurrency scams.
- To compare machine-learning algorithms for classifying fraud.
- To deal with class imbalance in cryptocurrency fraud datasets.
- To judge the models with suitable performance metrics.
- To look at how explainable AI can clarify model predictions.
- To explore graph-based analysis for fraud detection.
- To offer a framework that supports cryptocurrency security and fraud investigation.

#### 5. TYPES OF CRYPTOCURRENCY SCAMS

##### 5.1 Phishing

Attackers trick users into giving away private information or approving harmful transactions.

##### 5.2 Ponzi Schemes

A fraudulent scheme that pays earlier investors with the money coming in from newer investors.

##### 5.3 Fake Investment Projects

Fake projects that lure people with false claims, unrealistic returns or misleading details.

##### 5.4 Malicious Smart Contracts

Some smart contracts contain code written to manipulate transactions or to gain access to funds without permission.

##### 5.5 Fraudulent Wallets

Some blockchain addresses are known to be tied to malicious or fraudulent activity.

##### 5.6 Pump-and-Dump Activities

An asset is hyped to push interest or price up artificially, and then the promoters sell off their holdings.

##### 5.7 Wash Trading

Fake trading activity is created to give a false picture of how active a market is.

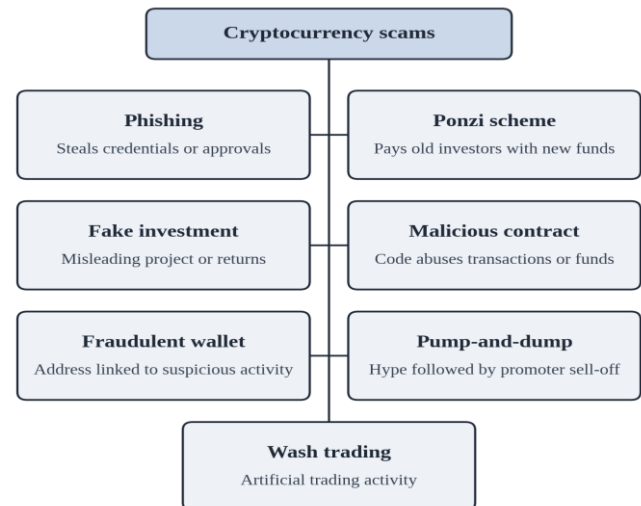


Figure 1. Overview of cryptocurrency scam categories

#### 6. PROPOSED METHODOLOGY

The proposed system works in these stages: Data Collection → Data Preprocessing → Feature Extraction → Feature Selection → Model Training → Scam Classification → Evaluation → Risk Prediction.

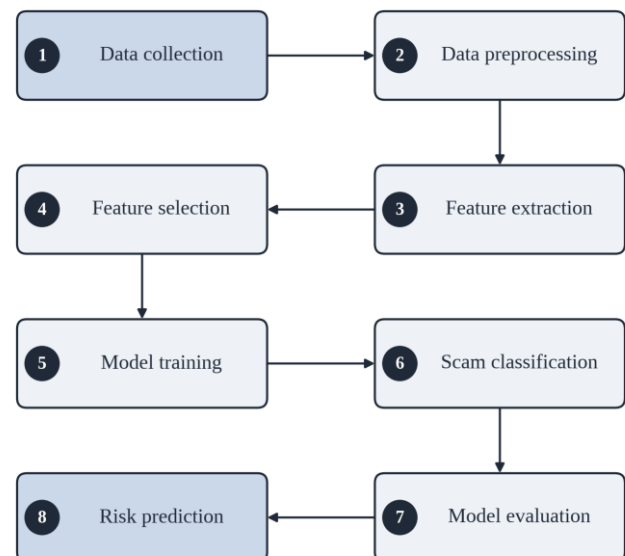


Figure 2. Proposed methodology workflow

##### 6.1 Data Collection

Blockchain transaction data can be taken from public research datasets and blockchain data sources. Useful

attributes include the wallet address, transaction amount and frequency, incoming and outgoing transactions, timestamps, contract details, gas usage, wallet interactions and scam labels.

One public example is the Elliptic++ dataset, a Bitcoin transaction dataset that suits graph-based fraud research. It holds more than 822,000 wallet addresses and over 1.27 million temporal interactions [6].

## 6.2 Data Preprocessing

- Remove duplicate records.
- Fill in or otherwise handle missing values.
- Turn categorical values into numbers.
- Normalize or standardize numerical features where needed.
- Drop features that are not relevant.
- Separate the legitimate and fraudulent labels.
- Handle class imbalance with a suitable method.
- Split the data into training, validation and testing sets.

## 7. FEATURE EXTRACTION

| Feature                   | Description                                                     |
|---------------------------|-----------------------------------------------------------------|
| Transaction Amount        | The value moved in a transaction.                               |
| Transaction Frequency     | How many transactions an address is involved in.                |
| Incoming Transactions     | Count or value of transactions received.                        |
| Outgoing Transactions     | Count or value of transactions sent.                            |
| Address Age               | How long the wallet or address has been active.                 |
| Unique Counterparties     | How many different addresses it has dealt with.                 |
| Average Transaction Value | The average value across its transactions.                      |
| Gas Usage                 | The computing cost of Ethereum transactions.                    |
| Contract Interaction      | How often it interacts with smart contracts.                    |
| Time Interval             | The gap between one transaction and the next.                   |
| Network Degree            | How many addresses it is connected to in the transaction graph. |

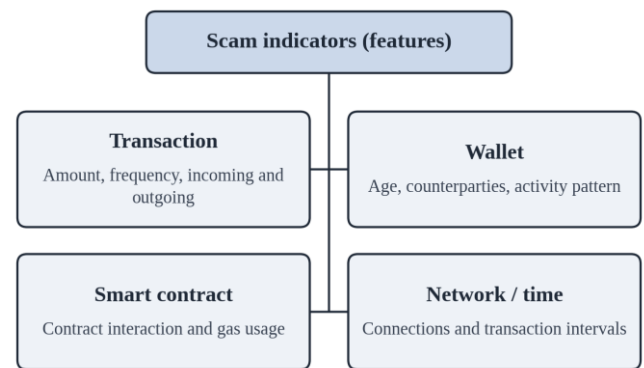


Figure 3. Feature groups used for scam analysis

## 8. PROPOSED AI MODELS

### 8.1 Logistic Regression

Logistic Regression can serve as a simple baseline. It estimates the probability that a transaction or wallet belongs to the fraud class.

### 8.2 Random Forest

Random Forest builds many decision trees and combines them. It can capture non-linear relationships and also reports which features matter most.

### 8.3 XGBoost

XGBoost is a gradient-boosting method that works well on structured transaction data. It can label records as legitimate or suspicious, and SHAP can be added to explain its output [3].

### 8.4 LSTM

Long Short-Term Memory networks read transactions in order, so they may pick up time-related patterns that show up in fraudulent behaviour [2].

### 8.5 Graph Neural Network

Graph Neural Networks study the links between addresses and transactions. They help when fraud shows up in the structure of the network rather than in a single transaction [4].

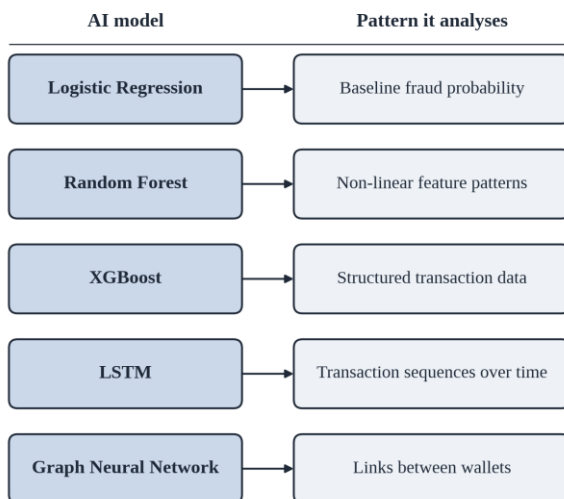


Figure 4. AI models and the patterns they can analyze

## 9. PROPOSED SYSTEM ARCHITECTURE

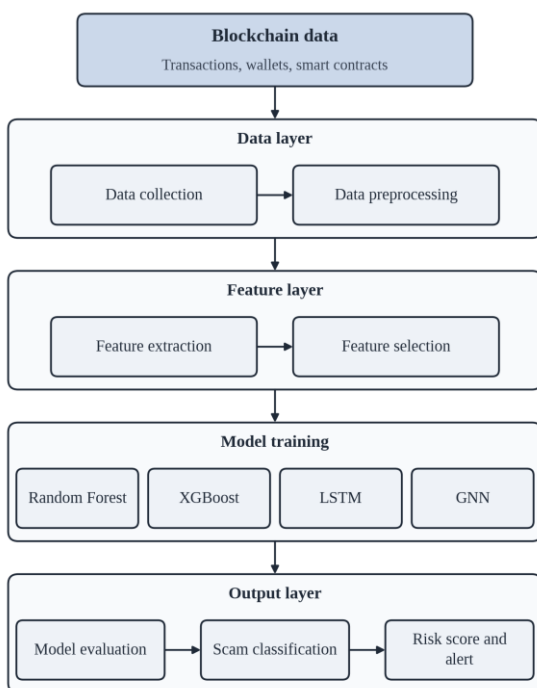


Figure 5. Proposed system architecture

Blockchain Data → Data Collection → Data Preprocessing → Feature Extraction → Feature Selection → Model Training (RF / XGBoost / LSTM / GNN) → Model Evaluation → Scam Classification → Risk Score and Alert

## 10. CLASSIFICATION PROCESS

The system can give a probability or risk score for each transaction or wallet. For demonstration, one possible scale is 0–30 for low risk, 31–60 for medium risk, 61–80

for high risk and 81–100 for very high risk. These cut-offs are only examples, and they need to be tested through experiments before anyone relies on them in a real setting.

## 11. EVALUATION METRICS

### 11.1 Accuracy

The share of all records that are classified correctly.

### 11.2 Precision

Out of the records predicted as scams, the share that really are scams.

### 11.3 Recall

Out of all the real scam records, the share that the model manages to find.

### 11.4 F1-Score

The harmonic mean of precision and recall.

### 11.5 ROC-AUC

Shows how well a classifier separates the two classes across different thresholds.

### 11.6 Matthews Correlation Coefficient

A balanced score for binary classification that is handy when the classes are imbalanced.

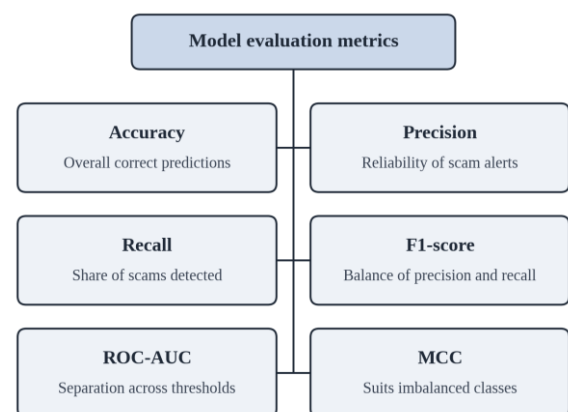


Figure 6. Evaluation metrics used to assess the models

## 12. PROPOSED EXPERIMENTAL RESULTS

The table below is a template for reporting results once the experiments are run. The TBD values should be replaced with real numbers from the chosen dataset and the implemented models.

| Model               | Accuracy | Precision | Recall | F1-Score |
|---------------------|----------|-----------|--------|----------|
| Logistic Regression | TBD      | TBD       | TBD    | TBD      |
| Random Forest       | TBD      | TBD       | TBD    | TBD      |
| XGBoost             | TBD      | TBD       | TBD    | TBD      |
| LSTM                | TBD      | TBD       | TBD    | TBD      |
| GNN                 | TBD      | TBD       | TBD    | TBD      |

### 13. ADVANTAGES

- Fraud detection can be automated.
- It can handle large blockchain datasets.
- It can find complex transaction patterns.
- It can study suspicious wallet behaviour.
- It can help spot possible fraud early.
- It cuts down the need for manual investigation.
- Models can be retrained when new patterns appear.
- Graph models can examine how wallets relate to each other.
- Explainable AI helps investigators understand predictions.

### 14. LIMITATIONS

- Results depend on how good and how broad the dataset is.
- Fraud labels may be incomplete or noisy.
- New scam methods may look different from the labelled examples.
- Different blockchain networks structure their data differently.
- Class imbalance can hurt model performance.
- Attackers may change their behaviour to avoid detection.
- Complex AI models can need a lot of computing power.
- A model's prediction should not be treated as proof of a crime.

### 15. FUTURE SCOPE

- Building real-time systems for detecting cryptocurrency scams.
- Combining graph neural networks with time-based transaction analysis.
- Using explainable AI to give investigators clearer alerts.
- Detecting scams across several blockchain networks.
- Adding analysis of smart-contract code.

- Using semi-supervised and unsupervised learning to find scams not seen before.
- Creating fraud-risk models that are updated continuously.
- Studying models that stay robust against evasion attempts.
- Connecting the system with blockchain monitoring and compliance tools.

### 16. CONCLUSION

Scams are a serious security problem for blockchains, because fraud can involve tangled transaction patterns, links between wallets, smart contracts and techniques that keep changing. AI and machine learning give us practical ways to study these patterns at scale. This paper proposed a framework built on data collection, preprocessing, feature extraction, classification, evaluation and risk prediction. Traditional models such as Logistic Regression, Random Forest and XGBoost can be used together with sequence and graph methods like LSTM and Graph Neural Networks, and explainable AI can add transparency. Future experiments should test these methods on carefully labelled blockchain data that is separated by time, and report precision, recall, F1-score, ROC-AUC and other suitable metrics.

### 17. REFERENCES

- 1) Chen, W., Zheng, Z., Cui, J., Ngai, E., Zheng, P., & Zhou, Y. (2021). Detecting Ponzi schemes on Ethereum: Towards a supervised learning approach. IEEE Access / related research literature on Ethereum Ponzi detection.
- 2) Wang, Z., et al. (2021). Ponzi scheme detection using oversampling-based LSTM approaches and smart-contract/account features. Knowledge-Based Systems, 228, 107252.
- 3) Liu, et al. (2024). IDPonzi: An interpretable machine learning approach for smart Ponzi scheme detection. Engineering Applications of Artificial Intelligence, 136, 108885.
- 4) Recent research on dynamic graph neural networks for Ethereum phishing/fraud detection. Pattern Recognition / related 2026 research literature.
- 5) Recent research on temporal transaction sequences, hypergraph representations and large language models for Ethereum phishing-scam detection. Expert Systems with Applications / related 2026 research literature.
- 6) Weber, M., Domeniconi, G., Chen, J., Weidele, D. K. I., Bellei, C., Robinson, T., & Leiserson, C. E. (2023). Anti-Money Laundering in Bitcoin: Experimenting

with Graph Convolutional Networks for Financial Forensics. arXiv:2306.06108 (Elliptic++ dataset).

- 7) Elliptic. (2019). Elliptic Data Set: Bitcoin transaction graph dataset for anti-money laundering research.
- 8) Nakamoto, S. (2008). Bitcoin: A Peer-to-Peer Electronic Cash System.
- 9) Zheng, Z., Xie, S., Dai, H., Chen, X., & Wang, H. (2017). An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends. IEEE International Congress on Big Data.
- 10) Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.