**Fraud Detection in Blockchain Transactions Using Machine Learning**

**CSCI 5895v – Blockchain and Cryptocurrency Technologies**

**HtayNaing Oo**

**Y-00858527**

**Instructor: Professor Feng George Yu**

**4/25/2025**

**Introduction**

Blockchain Technology has introduced decentralized and transparent transactions systems across various industries. However, the rise in blockchain adoption has also seen a surge in fraudulent activities, especially in cryptocurrency transactions. Detecting illicit activity in a decentralized environment remains a technical challenge due to the pseudo-anonymity of blockchain users and the volume of transaction data. This project investigates how machine learning (ML) techniques can be applied to identify suspicious Bitcoin transactions using the Elliptic dataset, which is a publicly available collection of labeled transaction data.

**Dataset and Preprocessing**

The dataset used for this project is Elliptic dataset, which includes over 200,000 Bitcoin transactions represented with 166 features. Each transaction is labeled as either “licit”, “illicit”, or “unknown”. To prepare the data:

* Transactions labeled as “unknown” were excluded from analysis.
* The remaining labels were converted to binary values: 1 for illicit and 0 for licit
* The dataset was split into training and test sets (70/30 split).
* Features were numerical and standardized by default; thus, no additional normalization was required.

A screenshot of a computer

AI-generated content may be incorrect.A screenshot of a computer

AI-generated content may be incorrect.Here’s are screenshots of dataset information:

**Methodology**

We trained and evaluated two classification models:

1. **Random Forest Classifier**: A method that handles high-dimensional data well.
2. **XGBoost Classifier**: A gradient boosting model known for speed and predictive performance.

A screenshot of a computer program

AI-generated content may be incorrect.Both models were trained in the same split of data for fair comparison. Evaluation metrics included accuracy, precision, recall, and F1-score, with particular focus on recall for the illicit class.

A screenshot of a computer

AI-generated content may be incorrect.

**Results and Discussion**

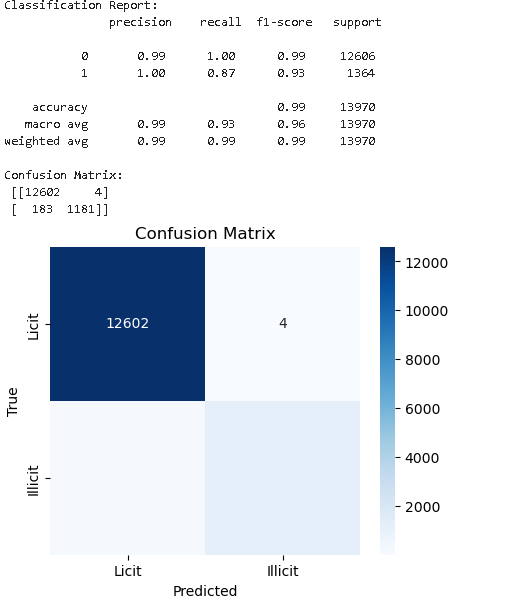
The Random Forest classifier achieved 99% accuracy with an F1-score of 93% for the illicit class. XGBoost slightly outperformed it with and F1-score of 94% and higher recall for illicit transactions (89% vs 87%).

|  |  |  |
| --- | --- | --- |
| **Metric** | **Random Forest** | **XGBoost** |
| Accuracy | 99% | 99% |
| Precision | 100% | 99% |
| Recall | 87% | 89% |
| F1-score | 93% | 94% |

Figure- The comparison of two models

The results confirm that both models are highly capable of detecting fraudulent transactions, but XGBoost provides a slightly better balance between sensitivity and precision.

A screenshot of a computer

AI-generated content may be incorrect.

**Conclusion**

This project demonstrates the effectiveness of machine learning for identifying fraud in blockchain-based systems. By leveraging transaction data and supervised learning algorithms, it is possible to flag suspicious activity with high accuracy. The use of XGBoost is particularly promising, offering robust performance even in imbalanced datasets. Future work can explore deep learning methods and real-time fraud detection integrations.

**References**

Elliptic Dataset: <https://www.kaggle.com/datasets/ellipticco/elliptic-data-set>

XGBoost Documentation: <https://xgboost.readthedocs.io/>

Scikit-learn Documentation: <https://scikit-learn.org/>