

**BT4012: Fraud Analytics**

**Group Project Final Report**

**Project Name: Detecting Fraudulent Accounts on Ethereum**

**Group 1**

| **Name** | **Matriculation Number** |
| --- | --- |
| Alex Ye Hein Aung |  |
| Chiu Xi, Brandon |  |
| Ho Cheuk Hei | A0217759N |
| Yan Yushan |  |
| Yu Shi Jie | A0223841L |

**1.Introduction**

Cryptocurrencies are digital assets powered typically by blockchain technology and have gained immense popularity in recent years due to their decentralized nature, diverse range of use cases, and substantial market values. However, this very decentralization presents significant problems such as fraudulent activities, including theft and money laundering, as transactions can be conducted anonymously. Since 2021, more than 46,000 individuals have fallen victim to cryptocurrency scams, resulting in a collective loss of over $1 billion, a striking 60-fold rise compared to 2018. The Federal Trade Commission (FTC) has identified Bitcoin, Tether, and Ethereum as the top choices for scammers to receive payments. Recent incidents, such as the misappropriation of over $1-2 billion in client funds by FTX, underscores the critical need for vigilance within the cryptocurrency ecosystem.

Ethereum is a decentralized blockchain platform that establishes a peer-to-peer network that uses smart contracts to securely execute and allow participants to transact with each other without a trusted central authority. Anonymity is one reason why it is challenging to detect frauds on Ethereum as transactions are conducted pseudonymously by cryptographic addresses, making it difficult to trace the culprit. Transactions are irreversible making it difficult to detect and recover stolen funds. Ethereum’s smart contracts are also susceptible to fraudulent activities such as reentrancy attacks. In addition to the above factors, the decentralized nature of ethereum makes it vulnerable to frauds and a challenge to detect.

In response to this growing and challenging threat, we have undertaken a project to enhance security and protect the assets of cryptocurrency users. Leveraging supervised machine learning techniques, including Logistic Regression, Naive Bayes, SVM, XGboost, LightGBM, MLP, Tabnet, and Stacking, our aim is to detect and predict fraudulent accounts specifically on the Ethereum blockchain. By doing so, we intend to fortify the fraud detection capabilities of crypto exchanges and wallets, empowering users to navigate the cryptocurrency landscape with confidence to safeguard their financial interests. Our primary goal is to achieve an F1 score of over 90% for our machine learning models, ensuring a high level of accuracy in predicting fraudulent Ethereum transactions and accounts. This project aims to serve as a crucial defense against the ever-evolving landscape of cryptocurrency fraud.

**2. Data**

2.1 Data Sources and Description

The dataset used is from Kaggle and the data is extracted from the Etherscan API through REST API’s and Web Scraping. Etherscan is the leading blockchain explorer, search, API and analytics platform for Ethereum aiming to provide equitable access to blockchain data. This dataset contains 9841 observations and a total of 49 variables, including the target variable “FLAG” and a unique identifier “Address” (Figure 1.1). Each row is an observation of a unique Ethereum account where each variable indicates the respective aggregate statistic over all transactions performed by that account (Figure 1.2). However, the dataset is highly unbalanced with only 22.14% of accounts flagged as fraud.

2.2 Data Preprocessing and Exploratory Data Analysis (EDA)

**Appendixes**

| **Data Source** | **Data Link** | **Number of Observations** | **Number of Variables** |
| --- | --- | --- | --- |
| Kaggle | <https://www.kaggle.com/datasets/vagifa/ethereum-frauddetection-dataset/data> | 9841 | 49 |

**Fig 1.1: Data Sources**

| **Variable** | **Data Type** | **Description** |
| --- | --- | --- |
| Address | object | The address of the ethereum account |
| FLAG | int | whether the transaction is fraud or not |
| Avg min between sent tnx | float | Average time between sent transactions for account in minutes |
| Avg\_min\_between\_received\_tnx | float | Average time between received transactions for account in minutes |
| Time\_Diff\_between\_first\_and\_last(Mins) | float | Time difference between the first and last transaction |
| Sent\_tnx | int | Total number of sent normal transactions |
| Received\_tnx | int | Total number of received normal transactions |
| Number\_of\_Created\_Contracts | int | Total Number of created contract transactions |
| Unique\_Received\_From\_Addresses | int | Total Unique addresses from which account received transactions |
| Unique\_Sent\_To\_Addresses20 | int | Total Unique addresses from which account sent transactions |
| Min\_Value\_Received | float | Minimum value in Ether ever received |
| Max\_Value\_Received | float | Maximum value in Ether ever received |
| Avg\_Value\_Received | float | Average value in Ether ever received |
| Min\_Val\_Sent | float | Minimum value of Ether ever sent |
| Max\_Val\_Sent | float | Maximum value of Ether ever sent |
| Avg\_Val\_Sent | float | Average value of Ether ever sent |
| Min\_Value\_Sent\_To\_Contract | float | Minimum value of Ether sent to a contract |
| Max\_Value\_Sent\_To\_Contract | float | Maximum value of Ether sent to a contract |
| Avg\_Value\_Sent\_To\_Contract | float | Average value of Ether sent to contracts |
| Total\_Transactions(Including\_Tnx\_to\_Create\_Contract) | int | Total number of transactions |
| Total\_Ether\_Sent | float | Total Ether sent for account address |
| Total\_Ether\_Received | float | Total Ether received for account address |
| Total\_Ether\_Sent\_Contracts | float | Total Ether sent to Contract addresses |
| Total\_Ether\_Balance | float | Total Ether Balance following enacted transactions |
| Total\_ERC20\_Tnxs | float | Total number of ERC20 token transfer transactions |
| ERC20\_Total\_Ether\_Received | float | Total ERC20 token received transactions in Ether |
| ERC20\_Total\_Ether\_Sent | float | Total ERC20token sent transactions in Ether |
| ERC20\_Total\_Ether\_Sent\_Contract | float | Total ERC20 token transfer to other contracts in Ether |
| ERC20\_Uniq\_Sent\_Addr | float | Number of ERC20 token transactions sent to Unique account addresses |
| ERC20\_Uniq\_Rec\_Addr | float | Number of ERC20 token transactions received from Unique addresses |
| ERC20\_Uniq\_Rec\_Contract\_Addr | float | Number of ERC20token transactions received from Unique contract addresses |
| ERC20\_Avg\_Time\_Between\_Sent\_Tnx | float | Average time between ERC20 token sent transactions in minutes |
| ERC20\_Avg\_Time\_Between\_Rec\_Tnx | float | Average time between ERC20 token received transactions in minutes |
| ERC20\_Avg\_Time\_Between\_Contract\_Tnx | float | Average time ERC20 token between sent token transactions |
| ERC20\_Min\_Val\_Rec | float | Minimum value in Ether received from ERC20 token transactions for account |
| ERC20\_Max\_Val\_Rec | float | Maximum value in Ether received from ERC20 token transactions for account |
| ERC20\_Avg\_Val\_Rec | float | Average value in Ether received from ERC20 token transactions |
| ERC20\_Min\_Val\_Sent | float | Minimum value in Ether sent from ERC20 token transactions for account |
| ERC20\_Max\_Val\_Sent | float | Maximum value in Ether sent from ERC20 token transactions for account |
| ERC20\_Avg\_Val\_Sent | float | Average value in Ether sent from ERC20 token transactions for account |
| ERC20\_Uniq\_Sent\_Token\_Name | float | Number of Unique ERC20 tokens transferred |
| ERC20\_Uniq\_Rec\_Token\_Name | float | Number of Unique ERC20 tokens received |
| ERC20\_Most\_Sent\_Token\_Type | object | Most sent token for account via ERC20 transaction |
| ERC20\_Most\_Rec\_Token\_Type | object | Most received token for account via ERC20 transactions |

**Figure 1.2 Data types and Description**