flagged-crypto-scraping

Introduction

This project involves the collection of cryptocurrency addresses flagged for involvement in activities such as dark web operations, scams, hacking, and other malicious activities. The primary focus is on major cryptocurrencies including Bitcoin, Ethereum, Litecoin, Tron, Solana, and any other significant ones where possible. The collected data will be used for transaction pre-screening to identify and mitigate risks associated with flagged addresses before processing transactions.

Data Source

Chainabuse.com: Chainabuse.com is a community-driven platform that features reported and flagged cryptocurrency addresses associated with scams, fraud, hacking, and other malicious activities. Users can report suspicious addresses and track flagged crypto addresses to stay informed and avoid potential threats.

Data Collection Process

- **Tools Used**: Scrapy, Selenium
- **Process**: The data collection involved analyzing the CSS used on Chainabuse.com to create CSS selectors for data extraction. Scrapy was used to extract content from the pages, while Selenium was employed to navigate through the pages as Chainabuse.com is a JavaScript-driven, dynamic website.
- **Preprocessing**: The data underwent cleaning to remove duplicates and ensure the uniqueness of addresses. This step was crucial to maintain the integrity and reliability of the dataset.

Data Description

The dataset consists of the following columns:

- `crypto_address`: The cryptocurrency address that has been flagged.
- `crypto_name`: The name of the cryptocurrency (e.g., Bitcoin, Ethereum, Litecoin).
- `flagging_reason`: The reason for the flagging, such as dark web activities, money laundering, hacking, terrorism, etc.

License

This project is licensed under the MIT License.

Author

@umairsiddique3171