# Project

| | |
|---|---|
| Context | Algorithms JHU ML |
| | ☐ |
| Due Date | @April 27, 2022 |
| Progress | Due Urgent |
| Related to Journal (Property) | |
| Related to Master ToDo (TimeLine) | |

- Ethereum
  - Ethereum is a decentralized, open-source blockchain with smart contract functionality.
  - Ether is the native cryptocurrency of the platform.
  - Among cryptocurrencies, Ether is second only to Bitcoin in market capitalization.
  - Ethereum was conceived in 2013 by programmer Vitalik Buterin who was just 18 years old
  - Special because of smart contracts
- Detection
  - Individual Account
  - Overall
  - What makes it bad - Each transaction is immutable
  - Real Time data processing would allow us to flag transactions in real time

Columns:

In the blockchain, a transaction is an event.

Address - Transaction Hash - Reference number for parties involved in the transaction created whenever a transaction is initiated on the Ethereum blockchain. We use

etherscan.io to search

We used this hash to fetch other columns

The next few columns measure the time difference between transactions like average mins between transactions and the time difference between the first and last transaction

Smart Contracts - Smart contracts are a type of Ethereum account . This means they have a balance and they can send transactions over the network.

However they're not controlled by a user, instead they are deployed to the network and run as programmed.

User accounts can then interact with a smart contract by submitting transactions that execute a function defined on the smart contract.

You need to deploy your smart contract for it to be available to users of an Ethereum network by sending an ethereum transaction containing the compiled code of the smart contract

In a blockchain, addresses are unique identifiers associated with an entity, a wallet or a smart contract.

They are usually composed of an alphanumeric string with between 26 and 35 characters. The address and the public key can be shared with anyone with no security restrictions.

On the other hand, the private key cannot be shared and should be kept secure (unless you want to lose all your money).

The address derives from the public key, and there are 3 simple steps to create an address:

- 1- Creating a private key (ECDSA)
- 2- Take the public key from the private key (Public Key Infrastructure always have private/public key pairs)
- 3- Hash the public key to generate the address

The next few columns talk about the value sent and received including that for contracts and the total number of transactions

Ether - Total Ether balance - ***Ether is the*** transactional token that facilitates operations on the ***Ethereum*** network.

ERC20 Tokens:

In the Ethereum system, tokens represent a diverse range of digital assets, such as vouchers, IOUs, or even real-world, tangible objects.

Essentially, Ethereum tokens are smart contracts that make use of the Ethereum blockchain.

ERC-20 has emerged as the technical standard; it is used for all smart contracts on the Ethereum blockchain for token implementation and provides a list of rules that all Ethereum-based tokens must follow.

ERC-20 is similar, in some respects, to bitcoin, Litecoin, and any other cryptocurrency; ERC-20 tokens are blockchain-based assets that have value and can be sent and received. The primary difference is that instead of running on their own blockchain, ERC-20 tokens are issued on the Ethereum network.

The ERC-20 commands vital importance; it defines a common list of rules that all Ethereum tokens must adhere to. Some of these rules include how the tokens can be transferred, how transactions are approved, how users can access data about a token, and the total supply of tokens.

Consequently, this particular token empowers developers of all types to accurately predict how new tokens will function within the larger Ethereum system. This simplifies the task set forth for developers; they can proceed with their work knowing that each and every new project won't need to be redone every time a new token is released, as long as the token follows the rules. This compliance is also necessary; it ensures compatibility between the many different tokens issued on Ethereum.

In terms of implementation coding for ERC-20 tokens, the six basic coding functions are:

1. total supply
2. balance of
3. allowance
4. transfer
5. approve
6. transfer from

Dealing with categorical data such as ERC-20 most received type tokens:

ERC20_Most_Sent Toke Type has 305 values

ERC20_Most rec_token type has 467 values

We dropped the categorical data as Liverpool token which constituted a large share of values had an equal number of fraud and valid transaction, this led us to believe that using the categorical data would heavily bias our data.

Another great example is of blockwell.ai - Blockwell.ai has an advanced system for flagging fraudulent transactions associate with known bank accounts. Because of their size they also have an equal number of valid transactions, but the data differentiates between them marking them differently. If we make both of them equal we basically lose essence of the data.

As any ERC token type is yet to be found malicious cos of the active community behind it, any fraudulent token type is never adopted. Its kinda survival of the fittest.

The second option was to convert it to numerical data, but as the relation does not hold true i.e old token type is not better than the other.

Because of the small dataset of fraud transactions and the heavy bias due to categorical data we decided to drop it.