**Bitcoin Price Prediction on the variables of block-chains**

**Abstract**

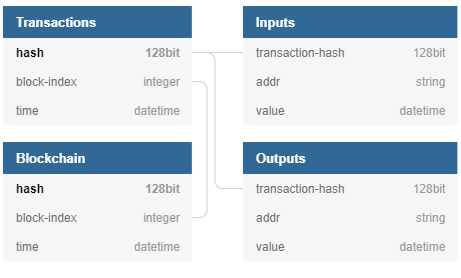
This project aims to extend cryptocurrency usage to remote areas and encourage companies to accept bitcoin transactions. In 2018, the famous futurologist made a prediction that cryptocurrencies are to replace 25% of fiat assets by 2030. One of main characteristics of cryptocurrency holding back is its volatility. To help extend the usage of cryptocurrencies, it is necessary to do bitcoin price predictions, as a next step for understanding cryptocurrencies and redesigning them in such a way that it won’t become as volatile as the current ones.

Cryptocurrencies work by someone broadcasting transactions and putting them into block. To solve the problem of double-spending, cryptocurrency miners verify the block and put those blocks into the history of verified blocks: block-chains. Since the transactions are done with the help of block-chains, one interesting assumption is that the cryptocurrency prices should somehow depend on the data inside the block-chain, because, after-all block-chain is the history of transactions.

**Introduction**

There are 4 steps to help achieve the goal: Data Gathering, Data Extraction, Data Processing/Feature Generation and Modeling the prediction. The data gathering is the process of acquiring the block-chain data and price history. The price history is publicly available on Kaggle [1]. And the bitcoin-blockchain data is available everywhere, since it is the only way to verify bitcoin transaction. For this paper, data is requested from blockchain API[2]. The main focus would be the cryptocurrency Bitcoin, and the time period: between 2021-January – 2021 March.

**Data Extraction**

The bitcoin blockchain is filled with unnecessary details, so here is the metadata for the extracted data from raw bitcoin transactions. So the entire blockchain is broken into 4 tables: Blockchain, Transactions, Inputs, Outputs. Below is show the schema of the tables. Created with the help of dbdiagram.io

**Data Processing and Feature Generation**

After Data Extraction, it would easier for this stage to combine Inputs and Outputs into a single table: Inp\_Outs table (Input values are positive and Outputs values are negative). After combining and observing the data, the following steps are taken to filter the data.

* The Inp\_Outs which have 0 values are dropped
* Inp\_Outs rows, which have the same address and transaction-id would be replace with one Inp\_Out transaction with the sum of the duplicates

In this stage all 5 tables (Price History and 4 tables above ) needed to be combined into one single table from which point on, the variables would be considered as features and the target variable would be the bitcoin price. The final table would look like the table below.

Graphical user interface

Description automatically generated with low confidence

Created with the help of dbdiagram.io

**References**

[1] Bitcoin historical prices Prices. Up to March 2021 by the time of writing this article. Kaggle Platform https://www.kaggle.com/mczielinski/bitcoin-historical-data

[2] Bitcoin Blockchain ResAPI https://www.blockchain.com/api/blockchain\_api