Analysis on Crypto Currency Market Cap

Data Mining on Ethereum Blockchain BigData

Block

2013171033 Hanseung Oh

2016121114 Hyeonjae Cho

2019840738 XU KEVIN CHEN

Prof. Seongbae Cho

2019-1 Data Mining Term Project

Contents

1. Problem Definition

1-1. Purpose

1-2. Data

2. Related work

3. Methodology

4. Expected Result

5. Role

6. Schedule

[2019-1 Data Mining]

Analysis on Ethereum

Market Cap

2013171033 Oh Han Seung
2016121114 Cho Hyeon Jae
2019840738 XU KEVIN CHEN

1. Problem definition

1-1. Purpose

Given the increased interest in Blockchain business, the need to learn about Blockchain is also increasing. Among the different Blockchain applications, our team is particularly interested in Ethereum, the second largest CryptoCurrency in the world. However, there is difficulty in predicting and analyzing the main factors which impact the capitalization of the Ethereum market.

Our team's goal is to find some main factors and make it easy for those who are interested in Ethereum to understand and predict Ethereum in order to help facilitate their decision-making process. In order to achieve this goal, we can use the characteristic of transparent and immutable network of Ethereum by obtaining those network's log data, like a transaction log. In short, by analyzing the Ethereum data, we can find main factor valuables, and using those valuables, we can predict a capitalization of each Ethereum market.

1-2. Data

In order for our project to be accurate, obtaining a reliable data set is an essential step. This can be achieved through BigQuery (Google's Big data platforms Ethereum data set, provided by Google), reliable because Google operates Ethereum's full node directly and collects that data in real time. In addition to this, Google processes the data by step of ETL (Extract, Transform, Load) so that the data is more user-friendly. To compare our predictions on market capitalization with real market capitalization, we will also need data information about real market capitalization. We can get this data

from coinmarketcap, a reliable site providing information on current cryptocurrencies. We will pre-process the data we collected and obtain a refined data set that will be used in our project. It is possible we will collect data information such as the dates of Ethereum's transaction, the size of ethereum traded, and the transaction's direction from where to where and so on.

2. Related work

Some research has been carried out on this subject in recent years. There are indeed a lot of expectations around this subject because anyone who is using Cryptocurrencies hopes to facilitate their decision-making process.

For example, there already exist some research on large Bitcoin transactions. This has helped to gain a better understanding of the transactions. Bitcoin holds the largest part of Cryptocurrency market capitalizations.

Furthermore, some work on Ethereum has also been carried out. For instance, the goal of our problem could have been the clustering of Ethereum Addresses by analysing the transaction activities in order to find some interesting patterns. This work could assist in predicting the type of an Ethereum's user such as whether it is an exchange or a miner.

3. Methodology

Our goal is to predict the numerical target variable Y : *Market Cap* ($). The predictors are then all other variables. Here are the steps that we will follow according to the data mining process :

1. First, we must define and understand the purpose. This step has already been achieved in the Problem definition.
2. Second, we want to obtain the data that will be used in our problem. This step has already been completed in the Problem definition.
3. Explore, clean, and pre-process data :
   In this part, we will first study the different variables and their relevance for the purpose of our problem. We can then decide to include or not the irrelevant variables in our dataset.
   Moreover, we can handle the missing data and the outliers in the dataset.
   Since some variables are categorical, it will be necessary to create dummy variables so that we can apply the different algorithms which require numerical variables.

4. Reduce the data dimensions :
   Each record is characterized by X variables. Thus, it seems pertinent to perform a principal components analysis in order to reduce the data dimensions.
5. Determine task (classification, clustering, etc.) :
   A supervised prediction task will then be carried out. The final model will be able to predict the variable *Market Cap* which is a numerical variable.
6. Partition data :
   We will randomly separate the data into different parts in order to avoid overfitting and to be able to assess our final model. Here are the several parts :
   - Training dataset : 50% of the original data,
   - Validation dataset : 30% of the original data,
   - Test dataset : 20% of the original data.
7. Choose the techniques :
   Since our task is a supervised prediction task of a continuous variable, we will use a supervised learning algorithm to build a prediction model : linear regression.
8. Iterative implementation and "tuning"
   In order to assess and tune our models, we will use the Validation dataset. We will try to predict the target variable in this dataset and we will compare the different errors computed with the model.
9. Evaluate results – compare models :
   After building several predictive models for the problem, we will need to select the best algorithm to deploy. That's why we will use different algorithms for selecting the best set of variables for predicting the outcome variable.
10. Deploy the model - scoring :
    Finally, we will run our final model on the Test dataset. This step allows us to be able to assess the performances of our model on real records.

4. Expected Results

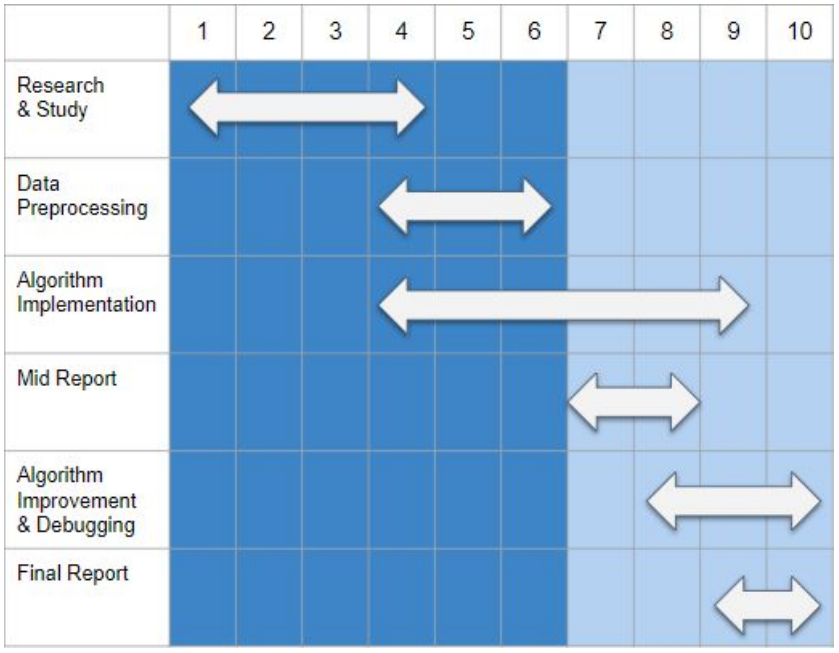
Like most cryptocurrencies, Ethereum has a high price volatility. This comes from being influenced by exchanges and speculations. Through exchanges, lots of investors dump or buy large amounts of cryptocurrencies. In the last 30 days, a 5.26% volatility index was recorded - which measures

the price volatility versus that of fiat currency using USD - according to Bitcoin World Wide[1]. Over the last 60 days, the index was recorded as 4.35%. CryptoCurrency, as name refers, also serves the role of currency. However, with volatile price, no one can use cryptocurrencies as a means of exchange or investment. With this project, first and foremost, we will discover the most influential variable that impacts Market Cap, which represents the value of a certain platform. Based on the conclusion of this project, investors will be able to predict Market Cap more conveniently and precisely despite the price volatility.

5. Role

All of the team members play pivotal roles in analyzing this project. Hyeonjae mainly organizes the proposal and powerpoint templates. Hanseung researches for background knowledge of this project. Kevin analyzes the process in statistical views. All of the team members continuously discuss about intermediate outputs and analyze the data.

6. Schedule



[1] please refer to : https://www.buybitcoinworldwide.com/ethereum-volatility/