Data Analysis Article

Cryptocurrency Data Analysis based on the Polarity Measurement using Twitter Data

Adam Ibnu Fiadi[a], Don Rui Tornado Rosa[b], Mochamad Hafidh Zuhdian[c]

[a] 19/440440/PA/19079

[b] 19/440446/PA/19076

[c] 19/440452/PA/19082

ARTICLE INFO

ABSTRACT

Specification Table

|	
Subject	Tweets with the topic of specified cryptocurrency (DigiByte)
Specific Subject Area	Sentiment analysis applied to the tweet data to measure the polarity of each tweet and then examine the impact to the current value of the spcryptocurrency
Type of Data	Pie Graph
	Line Graph
	Tables
How data were acquired	By using tweepy to connect the twitter developer API to the project and extract the data
Data format	Data acquired are in raw format
Parameters for data collection	Parameters for data collection is the polarity value of the sentiment analysis of TextBlob
Description of data collection	The data collected are from various amounts of text that the twitter users tweets daily. By specifying a keyword related to cryptocurrency, we can collect those data and convert it to raw format in the form of a csv file and processed in Microsoft Excel 2020.
Data source location	Twitter
Data Accessibility	The program code, raw data, and processed data can be retrieved from the Github repository https://github.com/NotAChigga/tweet-data-sentiment-analysis.git

Value of data
- These data are useful because we can detect the trend of a certain cryptocurrency based on the tweets we acquired at twitter
- Citizens who are plunging into the world of cryptocurrency can analyze trends from twitter for their own respective uses in the future
- These data represent an easy and simple way of detecting whether the value of a certain cryptocurrency can rise up or fall down in the market

1. Data Description

The data set provides insightful information about tweets based on the specified cryptocurrency coin. Before acquiring the data, we have to create a twitter developer app account in order to get the Twitter API that used to retrieve the tweets from twitter. After making the account, a link will be generated to acquire the twitter developer API. Each twitter developer account will have a unique API token. These tokens are not to be published in public for the sake of each account's privacy. The API generated will consist of 2 access token keys and 2 consumer keys used as an API to connect with our python project. In this project, we used two API to minimize the time used to extract the datas.

In order to connect the API, we will be using the tweepy library from python for accessing the Twitter API. This library is great for simple automation and creating twitter bots. When the connecting process is completed, we will have to specify what kind of data will we access and how many tweets will be fetched from it. Note that the twitter developer app has a regulation based on the number of tweets we can fetch at a time. The limit at which we were able to extract the datas is 2000 tweets per extraction with intervals of 15 minutes. In this example, our group will be collecting 10.000 tweets per day in the span of 1 week, so the total datas that is collected are 700.000 tweets with the keyword of a specified cryptocurrency, "Digibyte".
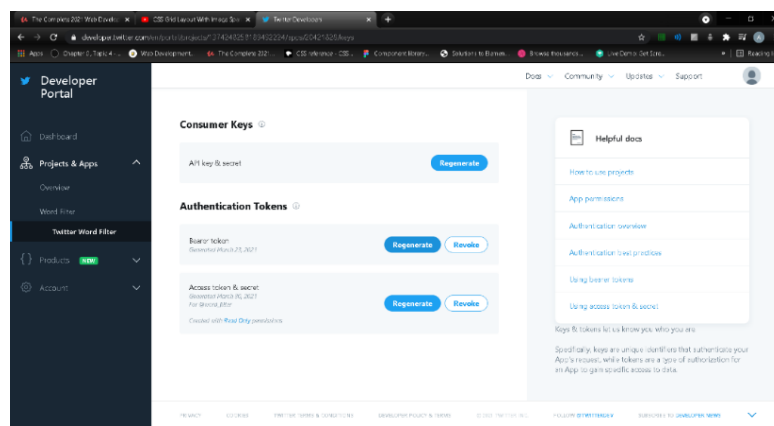
Fig 1. Twitter Developer API Dashboard

A	B	C	D
	tweet_id	text	created_at
0	1.377E+18	RT @RudyBouwman: We are	Wed Mar 31 23:57:57 +0000 2021
1	1.377E+18	RT @DigiByteCoin: After some hot f	Wed Mar 31 23:57:37 +0000 2021
2	1.377E+18	RT @RudyBouwman: We are	Wed Mar 31 23:55:14 +0000 2021
3	1.377E+18	Take a look at my latest #YouTube	Wed Mar 31 23:53:12 +0000 2021
4	1.377E+18	RT @RudyBouwman: No need to	Wed Mar 31 23:52:50 +0000 2021
5	1.377E+18	RT @RudyBouwman: We are	Wed Mar 31 23:48:46 +0000 2021
6	1.377E+18	RT @Jerry_nerdlab: #DigiByte first	Wed Mar 31 23:48:20 +0000 2021
7	1.377E+18	RT @AmoreLoveAmor: DigiByte is	Wed Mar 31 23:48:14 +0000 2021
8	1.377E+18	RT @DigiByteCoin: We've packed sc	Wed Mar 31 23:47:20 +0000 2021

Fig 2. Dataset fetched from twitter (100.000 tweets per day)

The dataset in Fig.2 contains 100.000 data tweets related to digibyte cryptocurrency from the 1st April 2021. It has 3 column attributes which are the tweet IDs, text, and the tweet date

created. Each tweet (text) has a unique tweet ID that can be used as an identifier and a key for retrieving the tweets. We also added a "created_at" column to determine the date for each tweets published.

2. Experimental design, materials and methods

The purpose of this project is to examine and analyze the trend of a certain cryptocurrency coin. Just like the stock market, the value of cryptocurrency will change depending on the ratings given from the community. A coin value will rise up if the number of demands are high and the supply is limited. Meanwhile if the demands are low, the value will fall down. One of the ways to measure the value trends is by collecting people's review about the certain cryptocurrency coin. This is done with the help of another python library called TextBlob.

Textblob is a python library for processing textual data. It provides a simple API for diving into common natural language processing (NLP) tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation, and more. The feature we will be using for the project is the sentiment analysis. Sentiment analysis can help us decipher the mood and emotions of general public and gather insightful information regarding the context. For more information, here is an example of sentiment analysis using Textblob.

SENTIMENT ANALYSIS OF A GIVEN SENTENCE

```
[8]  analysis = tb.TextBlob("Nice to meet you")
     print("Sentiment Analysis of the given word above: " , analysis.sentiment.polarity)

     Sentiment Analysis of the given word above:  0.6
```

SENTIMENT ANALYSIS OF A GIVEN SENTENCE

```
[10]  analysis = tb.TextBlob("He is a bad man")
      print("Sentiment Analysis of the given word above: " , analysis.sentiment.polarity)

      Sentiment Analysis of the given word above:  -0.6999999999999998
```

Fig 3. Example of Sentiment Analysis

As you can see on Fig 3, the sentiment analysis has a property named polarity. Polarity is float which lies in the range of [-1,1] where 1 means positive statement and -1 means a negative statement. With this sentiment, we can then identify the polarity of the twitter data we had fetched before. One problem my team encountered is that the TextBlob library can only analyze text data, meanwhile the tweets we retrieved consists of text and emojis. Therefore, in order for the sentiment analysis to work, we have to transform the emojis into a text using the emojis python library. Only then can the text data be analyzed using the polarity sentiment analysis.

**Fig 4.** Tweet text data before emoji transformation

Fig 5. Tweet text data after emoji transformation

After the emoji transformation, we can then analyze the whole tweet text data. To examine the total number of polarities in the dataset, my team specified a counter for positive, negative and neutral tweets depending on the text. For each 1 positive polarity tweet, the counter will iterate by 1 time, same as for the negative and neutral tweets. This counter will be outputted in the form of a pie charm and table according to the date each tweet is tweeted

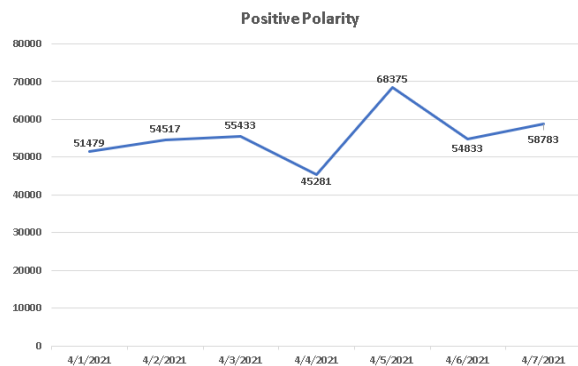| DATE	Positive Polarity	Negative Polarity	Neutral Polarity
4/1/2021	51479	7013	41508

4/2/2021	54517	5212	40271
4/3/2021	55433	4098	40469
4/4/2021	45281	7613	47088
4/5/2021	68375	7705	23920
4/6/2021	54833	8616	36549
4/7/2021	58783	4028	37189

Fig 6. Table of the total number of polarities for each date
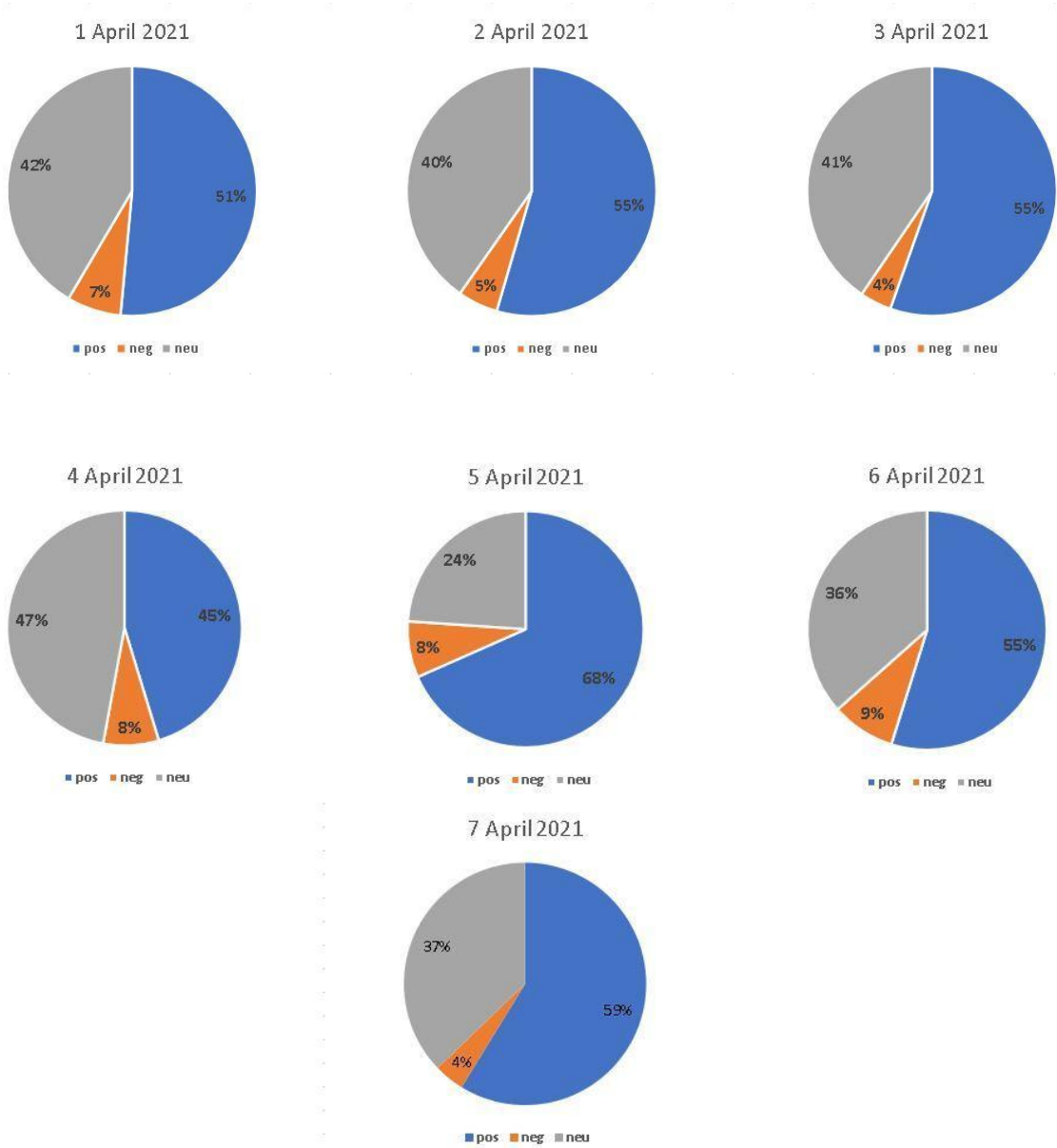
Fig 7. Pie Chart for the positive, negative, and neutral polarities from Day 1 - 7

With the data above, we can then compare the sentiment analysis we fetched and the data from the cryptocurrency market cap. As I said earlier above, my team analyzed the trend value of the digibyte cryptocurrency coin. For the sentiment analysis, only the positive polarity for each day will be measured and compared to the actual data of the coin. This is because the positive polarity reflects the movement of the digibyte coin value. Using the positive polarity, we can say that it can be a parameter for detecting whether the value will rise up or fall down. To prove the analysis, here is a graph based on the positive polarities and the digibyte coin value for each day.

Fig 8. Positive polarity line graph

Fig 9. Digibyte market value graph
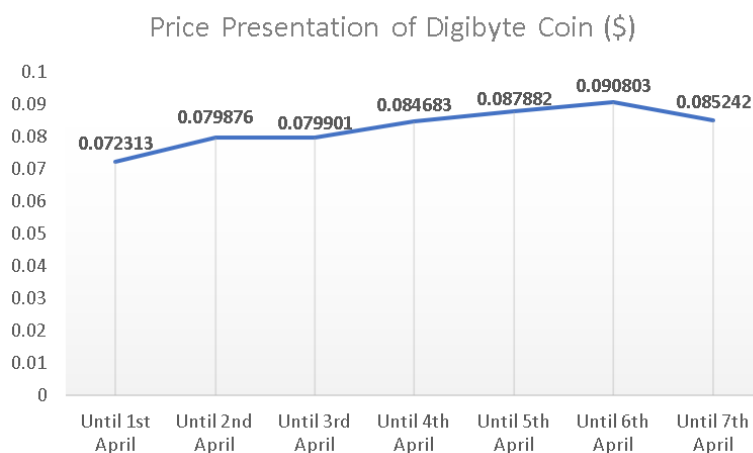
Fig 10. Price presentation of Digibyte cryptocurrency from 1st – 7th April 2021

Fig 8 shows that the positive polarity tweets had increased from April 1st to April 3rd and decreased the day after. But on April 4th , the polarity tweets increased by 50%, making it the highest polarity tweets within a week. Afterward, the polarity on April 6th faced a decrease and slightly increased the next day. If we compare the polarity tweet graph on Fig 8 with the digibyte market value graph on Fig 9, the difference of the graph is not too tenuous. Meanwhile if we compare the graph on Fig 8 with the actual digibyte price presentation on Fig 10, the difference can be seen immediately.

Comparing the April 3rd data on Fig 8 and Fig 10, the line chart seems to change in the opposite direction. The Fig 10 graph tends to increase slightly on April 3rd but suddenly decreases on the graph of Fig 8. If we look closely, this is because the value inconsistency of the actual digibyte market value graph on Fig 9, which tends to increase in the middle of the day and suddenly decreases again from April 3rd to April 4th . Another different measurement of the Fig 8 and Fig 10 graph is the data located between April 5th and April 7th. The problem is the same as before whereas the Fig 9 graph experienced an inconsistency of the data. Despite all the differences, the graph on Fig 8 and Fig 10 has the same chart movement of data from April 1st to April 3rd and April 4th to April 5th. This proves that the polarity sentiment analysis is quite significant but not completely consistent.

In conclusion, the data for 4 days experienced the same graph movement when comparing

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at

References