# Can we determine the Doge Price direction based on influential tweets?

Elangovan, Vishnu
velango3@ur.rochester.edu

Konjety, Keerthi Sreenivas
kkonjety@ur.rochester.edu

## Abstract

***The crypto currency Dogecoin's price has seen great fluctuations in april and may of 2021. Our study tries to observe the trends in twitter sentiments and the significant impact of Elon Musk's twitter activity on Dogecoin over a series of two events. Two such event's we analyzed is when, after Elon Musk claimed via Twitter that "SpaceX is going to put a literal Dogecoin on the literal moon" and later called himself "DogeFather" as he announced publicly that he would be hosting the Saturday Night Live on 8 May 2021. By doing sentiment analysis using Textblob on relevant tweets before Musk's tweet , during the period between his tweet and his actual appearance on the show and after the show , we found that Tweet volume , twitter sentiments and Google trends collectively play as predictor variables of price direction. By using these insights, a person is able to make better informed purchase and selling decisions related to Dogecoin/ or other crypto currencies when they are being talked about.***

## 1. Introduction

Dogecoin: Crypto currency is a form of tradable digital money built on blockchain technology. Doge coin is one of the crypto currencies that has been in the spotlight in 2021, becoming one of the top 10 coins that was traded with a market cap of over $28 billion Doge coin's price has seen a sudden rise during April and May 2021. A lot of discussion surrounded the coin on social media like Twitter and Reddit.

Elon Musk, the CEO of Tesla and SpaceX, an active Twitter user, has shown his interest towards cryptocurrencies like bitcoin and made his opinions public during several instances. Him being a social media influencer, his tweets have influenced the public's opinion towards buying cryptocurrencies.On Jan 29, 2021 , he changed his twitter bio to bitcoin , after which the price of Bitcoin rose from $32,000 to $38,000.

On 28th,April 2021, Elon Musk tweeted "The Dogfather SNL May 8". He called himself the Dogefather and conveyed to the audience that he would be hosting the Saturdays Night Live show on May 8th. The price of the coin rises from 28th of April to May 8th. In the show, he indicates through his words "Hustle" that he purposely pushed the price of the coin. After which, there is a sudden drop in the price of the dogecoin.

Twitter is a platform where we have publicly available opinions of the crowd. Millions of tweets are posted by users daily and these tweets are a good measure of general public opinion of a particular topic. Tweet volumes on a particular topic also is a good indicator of knowing what the public is interested in and what is trending currently.Twitter data has a lot of words and sentences that can give us meaningful insights. Sentiment Analysis is a well known research area in NLP studies which helps us understand and gain insights from text data. As the word means , it helps us to learn the sentiment of a user like positive,negative and neutral opinions on a certain topic.

As a matter of fact , Google has been the No 1 search engine in the world for a while now , and it's no surprise it's been used to infer what most people search on the internet in any part of the day. The search data provides credible values and insights in understanding what is trending locally and globally. Google provides this data publicly and it's fascinating to observe that it acts as a metric to predict cryptocurrency price directions with its relationship with interests of the public.

In our experiment, we are trying to observe the tweet patterns of twitter users, from 28th,April 2021 to 8th may, 2021 during which the price of the coin rises, and 8th may 2021 to 17th may 2021, during which the price of the coin drops. These two drastic events will be compared in terms twitter sentiments, tweet volumes, word clouds and top words used by twitter users. Further, we will be comparing the google trends data, tweet volumes, and tweet sentiments to infer patterns in the change of price of doge coin.

Thus this study contributes to the existing research on the influence of social media on the price prediction of a cryptocurrency and the investor attention in this market.

## 2. Terminology:

For the purpose of this study we use the terms Event 1 to refer the time interval between 28th,April 2021 to 8th may, 2021 and Event 2 to refer to the time interval between 8th may, 2021 to 17th may, 2021.

## 3. Related Work

This paper builds on the ideas of a wide range of research and topics. Behavioral economists such as Daniel Kahneman and Amos Tversky have discovered that emotions, as well as values, influence decisions, including those having economic consequences [7]. Emotions have a substantial influence on decision-making, according to R. J. Dolan's study on "emotions, cognition, and behavior" [8]. These researchers' findings show that demand for items, and thus prices, cannot be driven solely by economic fundamentals, and that profits can be determined using technologies like sentiment analysis.

Later study discovered that the information acquired online influenced people's purchase decisions in particular. Galen Thomas Panger discovered that people's mood on Twitter is related to their normal emotional state. Furthermore, social networking sites like Twitter have been demonstrated to have a relaxing influence on end-user emotional states rather than an elevating effect [9]. Chen and colleagues , did a text analysis on Seeking Alpha, an investor-led social media site, and discovered that the opinions expressed in the article are related to results and can even anticipate earnings surprises [10]. Paul Tetlock discovered that media pessimism about the stock market had an impact on trade volumes [11]. Lastly , according to a poll conducted by Gartner, the majority of consumers use social media to influence their buying decisions [12].

Other researchers are extensively studied the efficacy of sentiment analysis in tweets. Kouloumpis and colleagues discovered that Standard natural language processing approaches such as sentiment scoring at the sentence and document levels have been shown to be ineffective due to the simplicity of the tweet and the uniqueness of the language utilized [13]. Alexander Pak and Patrick Paraubek demonstrated that sorting tweets into good, negative, or neutral categories allows for successful emotion analysis [14]. Others, including O'Connor found that sentiment expressed in the tweet mirrored public opinion on a variety of topics in a national poll [15]. In comparison to national surveys, her research revealed sentiment analysis as a cost-cutting option, with the indication that emotions from tweets properly reflect the sentiments of more individuals on the topic. It implies that it can be used to forecast changes in demand and, as a result, prices.

Other than Twitter and social media, web data was a rich source of research. The idea is derived from Ettredge et al, as far as we know. They discovered that job searches were linked to unemployment rates [4]. Bordino and his colleagues found that the volume of queries was found to be related to the volume of NASDAQ equities traded [16]. Hyunyoung Choi and Hal Varian also conducted a concrete study on Google Trends data, finding that a simple seasonal automatic regression model using Google Trends data as input increased the model accuracy by 5 %to 20% more than a model that did not use Google Trends data. [3]. Asur and his colleagues discovered that the number of tweets regarding recently released movies turned out to be a good predictor of box office earnings. [17].Bollen et al. Using a self-organizing fuzzy neural network with Twitter sentiment as input, predicted the price movements in 30 Dow Jones stocks and obtained an accuracy of 86.7 [20] .

With the advent of cryptocurrencies, similar work has been done to see if such methods are effective in predicting changes in the price of cryptocurrencies. Evita Stenqvist and Jacob L onn o article, "Predicting Bitcoin Price Movements with Twitter Sentiment Analysis", describes the process of collecting Bitcoin and Bitcoin price tweets from May 11 to June 11, 2017. The tweets have been excluded from non-alphanumeric symbols (using "" and "@" as examples of symbols removed). Posts that were later deemed irrelevant or too influential were removed from the analysis. The author then uses the Valence Perception Dictionary and the Emotional Regulatory Authority (VADER) to analyze the emotion of each tweet and categorize it as negative, neutral, or positive. The final analysis retained only tweets that could be considered positive or negative [21].We used a similar study inspired from her work.

Connor Lamon et al. used sentiments from headlines and tweets, and predicted changes in Bitcoin, Litecoin (one of many alternative cryptocurrencies currently on the market) and Ethereum. The study found that logistic regression was the best way to classify these tweets and was able to correctly predict 43.9% price increases and 61.9% declines [22]. Colianni et al. collected tweets from November 15, 2015 to December 3, 2015, and used Naive Bayes and Support Vector Machines to classify tweets, and there were 255 improvements in accuracy. [23] Finally, Shah et al succeeded in establishing a trading strategy using historical prices and Bayesian regression analysis [24]

Other areas of research in this area include various applications of neural networks. Kimoto et al created a synchronization system for stock market transactions on the Tokyo Stock Exchange using modular neural networks and profitability through simulated stock purchases [25] . Guresen et al compared the performance of different neural networks for stock price prediction and found that neural networks with Multilayer Perceptron (MLP) performed the best [26]. Xu et al Used stock trading volume as input to the neural network, and found a modest improvement in mid and long-term prediction performance[27].

The research presented in this paper builds on all of the above, but predicts cryptocurrency price movements by taking web search data (in the form of Google Trends),tweets and tweet volume as inputs for our analyzer and it is unique in this way in solving the problem.In addition, we will show why sentiment analysis is not very useful in terms of predictive capabilities of cryptocurrencies, even though it has potential in other areas.

# 4. Methodology

## 4.1. Data Sources:

Different data sources were considered as input. Tweets, google trends data, doge price data and tweet volume data to observe a pattern in price change and public opinion change.

## 4.2. Tweet Extraction:

We have extracted 115,387 tweets for Event-1 and 255,000 tweets for Event-2 tweets using a twitter API researcher account. These tweets relevant to our topic of interest are considered by taking ”Doge”, ”Dogecoin”,”Elon Musk”, ”Musk” as our keywords and hashtags. We filtered the scrapping of data by restricting it to only the English language for easy interpretation. Once we tested our approach and experimented with different keywords and hashtags , we automated the script to run every 15 minutes by dividing into 6 hours time period blocks of data each day to minimize any redundancy and increase the volume of our data.

## 4.3. Tweet Volume:

Using twitter API documentation code, we have extracted the tweet volumes of ”Doge” . We converted the .json format output into .csv and visualized it using python and matplotlib.

## 4.4. Google Trends data:

Google trends provides readily downloadable search index values in format of csv files. We have used google trends plots to compare the corresponding spike and drop in the price of doge.

## 4.5. Analysis of Tweets:

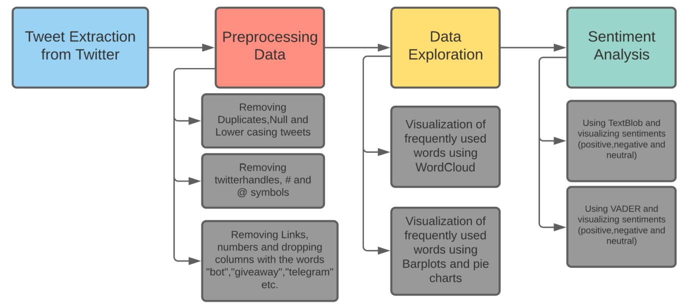


Figure 1. Step by step process of sentiment analysis

### 4.5.1 Step1: Preprocessing

Event 1 tweets and event 2 tweets were preprocessed so that only clean data is sent for analysis. ( Figure 1 )Preprocessing consisted of several steps. Following are the preprocessing steps performed in order.

- Removing Null Values
- Dropping duplicates records
- Lowercasing the tweets
- Removing @ and twitter handles
- Removing # symbols
- Removing Links
- Removing numbers
- Dropping records with some spam words - giveaway, telegram, bot, follow and share, earn free, video
- Tokenization
- Removing stop words
- Removing punctuations

### 4.5.2 Step2: Data Exploration

- Using the word cloud module in python we have created a visualization of most repeated words in the tweets. The more frequently it appears in the data, the bolder and bigger is the word.
- Analysing top 10 words in the twitter data: We used Counter functions from collections module to count the frequency with which each word has occurred

- We analyse the most used hashtags during this period using a barplot visualization. We pass the original tweet data without processing to count the most frequent hashtags.

#### 4.5.3 Step3: Sentiment Analysis

Firstly, we use textblob to perform sentiment analysis. TextBlob is a python library that provides text mining, text analysis and text processing modules in python. Texblob performs analysis at sentiment level by taking the dataset input and polarity is calculated in it sentence by sentence. We determine the polarity of the dataset by counting the number of positive and negative. Polarity is determined using sentiment() function. To this sentiment function we pass polarity ( range: -1 to 1) and subjectivity (0 to 1), 0 being most objective and 1 being most subjective.

Next, we apply the VADER sentiment analysis algorithm to the data. VADER is a lexicon and a rule based tool that uses sentiment lexicon and lexical features labelled across semantic orientation (positive or negative). VADER is quite popular for its success with social media data, movie review data, and e-commerce product review data. VADER not only tells us whether the data is positive or negative, it tells us the extent to which it is positive and negative. It is fast and does not require any training as it is constructed from valence-based and human curated gold standard lexicon. We show the results of VADER in a pie chart representing the amount of positive, negative and neutral tweets.

## 5. Experiment and Results:

Doge coin price graph is shown in the graph below. Figure 2, shows a spike during may 2nd week around which event 1 and event 2 have been seperated.

Tweet volumes during Event 1 and Event 2 can be seen in figure 3. We observe that the amount of tweets about the Dogecoin have increased till Saturday Night (May 8th, 2021), and price also increases correspondingly. After Elon reveals it's a hustle, we observe a dip in the volume of tweets about doge coins and the price also drops soon after the saturday night live.

We observe that the google trends search index (figure 4) shows that the search index has been rising from around 25th April, 2021 and drops after 9th May,2021.

From the wordcloud of event 1(Figure5 ) , we see that "Buy" and "Moon" are the most frequent words. These words indicate that people were inclined towards buying the coin during event 1. "Doge to the moon" was a common phrase used by the users during this time, as Elon tweeted regarding this sometime earlier than event1(Figure 6). Hence, we can infer that Elon has


Figure 2. Doge coin price graph

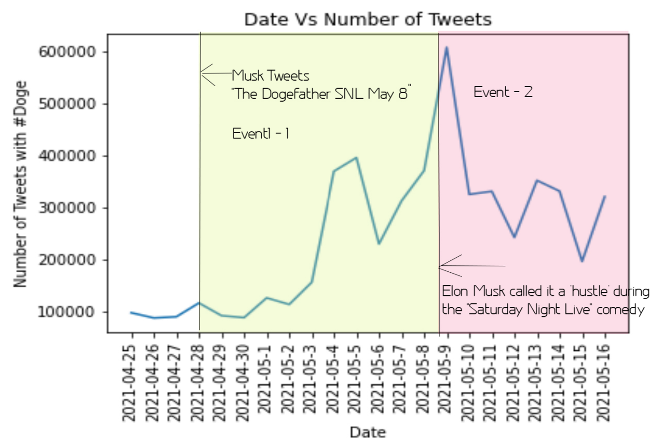


Figure 3. Tweet Volume Graph

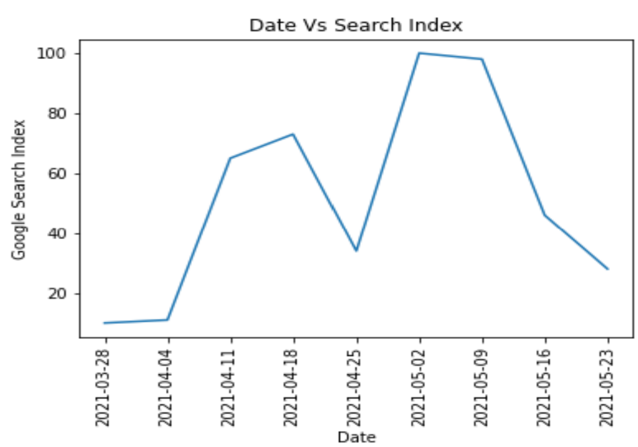


Figure 4. Google trends for the word Doge

influenced users towards buying the coin.


Figure 5. WordCloud for event 1


Figure 6. Tweet by ElonMusk

From the wordcloud of event 2 (Figure 7) , we see that "Sell ","One" and "drop" are the most frequent words. These words indicate that people did observe the drop in the prices and have started talking about it . After Elon's conversation about dogecoin in the Saturday night live's show, the price crashes, and people start talking about it. From many tweets, we have observed that "One down" was a common phrase being used, meaning "at a disadvantage in a game or a competitive situation". We can infer that people observe this drop and are inclined towards selling the coin.

The figure 8 shows the most commonly used words in doge coin related tweets during event1. We excluded bitcoin and btc on wordcloud to differentiate some commonly used words during a special circumstance like event1 and 2. We observe btc,bitcoin,eth other popular cryptocurrencies being used in the tweets, indicating the users must have a good population of cryptocurrency followers/buyers.

The figure 9 shows the most commonly used words in doge coin related tweets during event2. Drop, sell, get and one were some of the commonly used words in the tweets. Buy was still in the top 10 words used, indicating that some proportion of the tweets were hopeful that the price may increase again. [h]


Figure 7. WordCloud for event 2

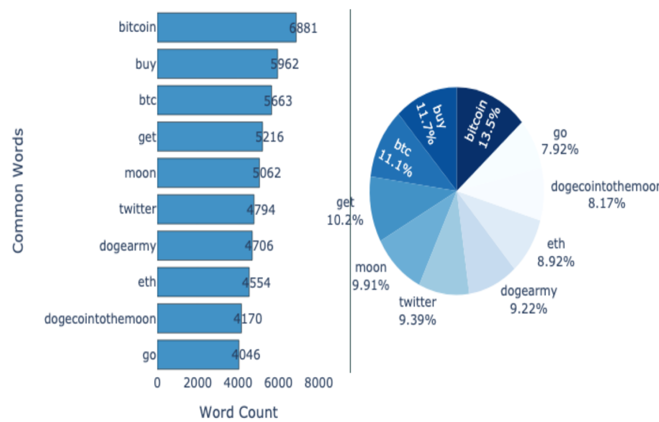


Figure 8. Top 10 words in tweets around event2

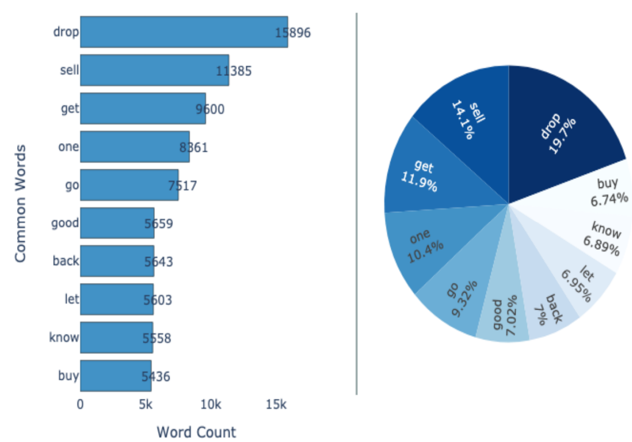


Figure 9. Top 10 words in tweets around event2

Sentiment analysis using Textblob for event1 and event 2 show similar proportions of positive, negative and neutral tweets (Figure 10 and 11), indicating not much difference in opinion. But there is a spike in the negative tweets and positive tweets together.

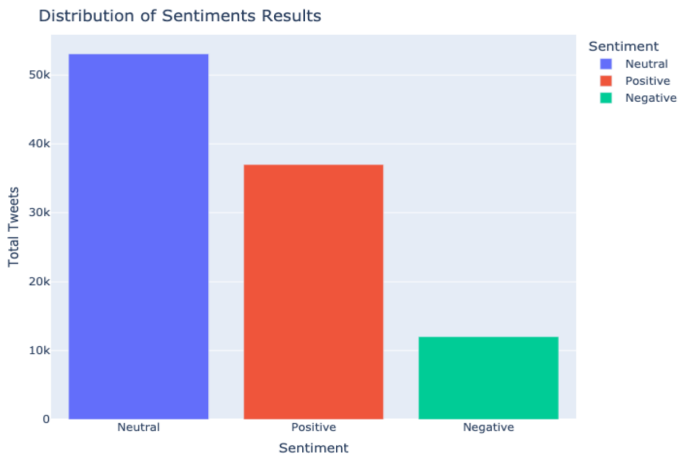


Figure 10. Event 1 Sentiments (TextBlob)

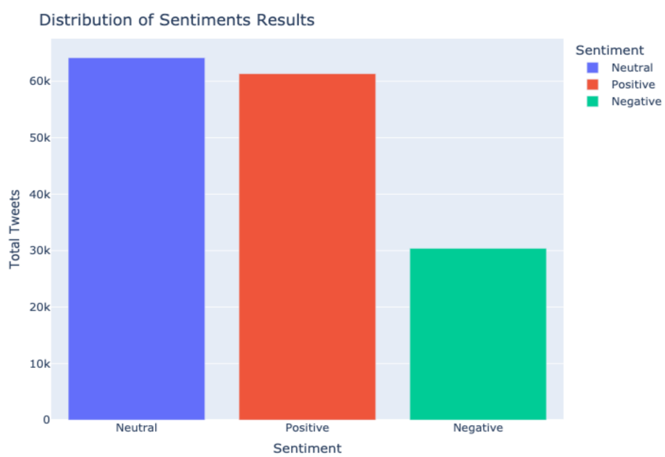


Figure 11. Event 2 Sentiments (TextBlob)

Sentiment analysis using using VADER for event 1

## 6. Conclusion

In this paper , we found that it is not possible to show the relationship between the price movements and twitter sentiments floating over the crypto-tastemakers like Elon Musk. We were able to process enough tweets and pass it to textblob and VADER to understand the public's sentiment in two events. We observed that the overall sentiment remained majorly positive, even though the prices plunged after event 2.

There are many areas of improvement to our study , by improving the quality of the input data , by increasing the number of sources where such content is gathered.

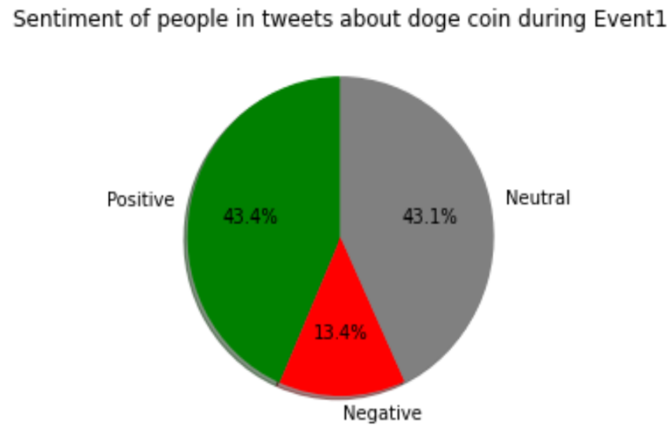


Figure 12. Event 1 Sentiments(VADER)

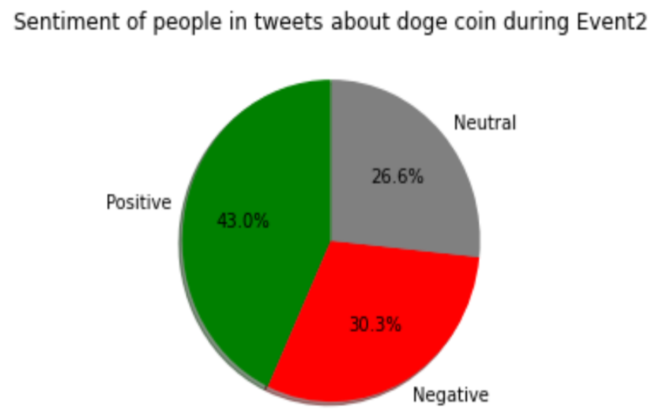


Figure 13. Event 2 Sentiments (VADER)

Other areas of work can be done by using more specialized models like LSTM's and T-MLP's , which can inherit the "moods" of the market and adapt according to it.

## References

[1] Nakamoto, S.: Bitcoin: A peer-to-peer electronic cash system. In: Cryptography Mailing list at https://metzdowd.com. (03 2009)

[2] Kristoufek, L.: What are the main drivers of the bitcoin price? evidence from wavelet coherence analysis. PLOS ONE 10(4) (04 2015) 1–15

[3] HYUNYOUNG, C., HAL, V.: Predicting the present with google trends. Economic Record 88(s1) 2–9

[4] Ettredge, M., Gerdes, J., Karuga, G.: Using web-based search data to predict macroeconomic statistics

[5] Miraz, M.H., Ali, M.: Applications of blockchain technology beyond cryptocur- rency. CoRR abs/1801.03528 (2018)

[6] Hutto, C., Gilbert, E.: Vader: A parsimonious rule-based model for sentiment analysis of social media text. In: 8th International AAAI conference on weblogs and social media (IDWSM),. (2014)

[7] Kahneman, D., Tversky, A.: Prospect theory: An analysis of decision under risk. Econometrica 47(2) (1979) 263–291

[8] Dolan, R.J.: Emotion, cognition, and behavior. Science 298(5596) (2002) 1191– 1194

[9] Panger, G.T.: Emotion in Social Media. PhD thesis, University of California, Berkeley (2017)

[10] Chen, H., De, P., Hu, Y.J., Hwang, B.H.: Customers as advisors: The role of social media in financial markets (2011).

[11] Tetlock, P.C.: Giving content to invsotry sentiment: The role of media in the stock market. The Journal of Finance (2007)

[12] Gartner: Gartner says majority of consumers rely on social networks to guide pruchase decisions (2010)

[13] Kouloumpis, E., Wilson, T., Moore, J. In: Twitter Sentiment Analysis: The Good the Bad and the OMG! AAAI Press (2011) 538–541

[14] Pak, A., Paroubek, P.: Twitter as a corpus for sentiment analysis and opinion mining. In: LREC. (2010)

[15] O'Connor, B., Balasubramanyan, R., Routledge, B.R., Smith, N.A.: From tweets to polls: Linking text sentiment to public opinion time series. In: Proceedings of the Fourth International AAAI Conference on WEblogs and Social Media. (2010)

[16] Asur, S., Huberman, B.A.: Predicting the future with social media. In: 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology. Volume 1. (Aug 2010) 492–499

[17] Sul, H., Dennis, A.R., Yuan, L.I.: Trading on twitter: The financial information content of emotion in social media. 2014 47th Hawaii International Conference on System Sciences (2014) 806–815

[18] de Jong, P., Elfayoumy, S., Schnusenberg, O.: From returns to tweets and back: An investigation of the stocks in the dow jones industrial average. Journal of Behavioral Finance 18(1) (2017) 54–64

[19] Bollen, J., Mao, H., Zeng, X.: Twitter mood predicts the stock market. Journal of Computational Science 2(1) (2011) 1 – 8

[20] Stenqvist, E., Lo nno , J.: Predicting bitcoin price fluctuation with twitter sentiment analysis. KTH Royal Institute of Technology School of Computer Science and Communication (2017) 3–28

[21] Lamon, C., Nielsen, E., Redondo, E.: Cryptocurrency price prediction using news and social media sentiment. Master's thesis, Standord (2015)

[22] Colianni, S., Rosales, S.M., Signorotti, M.: Algorithmic trading of cryptocurrency based on twitter sentiment analysis (2015)

[23] Shah, D., Zhang, K.: Bayesian regression and bitcoin. 2014 53nd Annual Allerton Conference on Communication, Control, and Computing (Allerton) (2014) 409–414

[24] Kimoto, T., Asakawa, K., Yoda, M., Takeoka, M.: Stock market prediction system with modular neural networks. In: 1990 IJCNN International Joint Conference on Neural Networks. (June 1990) 1–6

[25] Guresen, E., Kayakutlu, G., Daim, T.U.: Using artificial neural network models in stock market index prediction. Expert Systems with Applications 38(8) (2011) 10389 – 10397

[26] Zhu, X., Wang, H., Xu, L., Li, H.: Predicting stock index increments by neural networks: The role of trading volume under different horizons. Expert Systems with Applications 34(4) (2008) 3043 – 3054