As per UGC guidelines an electronic bar code is provided to seure your paper

International Journal for Modern Trends in Science and Technology, 8(05):542-551, 2022

ISSN: 2455-3778 online
DOI: https://doi.org/10.46501/IJMTST0805082
Available online at: http://www.ijmtst.com/vol8issue05.html

Prediction of Crypto-Currency Trends by Tweet Sentiments and Volume Analysis

Ayush Sultania | Bhaskar Kapoor

Department of Information Technology, Maharaja Agrasen Institute of Technology, New Delhi.
Corresponding author Email ID: sultania.ayush27@gmail.com

To Cite this Article
Ayush Sultania, Bhaskar Kapoor. Prediction of Crypto-Currency Trends by Tweet Sentiments and Volume Analysis. International Journal for Modern Trends in Science and Technology 2022, 8(05), pp. 542-551. https://doi.org/10.46501/IJMTST0805082

Article Info
Received: 22 April 2022; Accepted: 16 May 2022; Published: 25 May 2022.

ABSTRACT

In this work, we put forward a mechanism for deducing Ethereum and Bitcoin price changes using Google Trends and Twitter data. They are the two most valuable cryptos, with a total market cap of $160 billion dollars. On the other hand, both daily and long-term valuations of Bitcoin and Ethereum have seen a lot of price volatility. Twitter is quickly becoming a news source, informing people about the currency and its expanding popularity, and so influencing purchasing decisions. As a result, a cryptocurrency user or trader might gain an advantage in purchasing or selling by immediately comprehending the effect of tweets on price direction. By examining tweets, we discovered that tweet quantity, not tweet sentiment (which is always positive despite the price trend), is a better indicator of price direction. We have been able to correctly anticipate the trend of price variations using a linear model that incorporated Google Trends data and tweets as input. Using this model, a user may make more informed Bitcoin and Ethereum purchase and selling decisions.

KEYWORDS: Tweets, Sentiment Analysis, Vader, Tweepy, Bitcoin, Ethereum

1. INTRODUCTION

As of May 2018, Bitcoin and Ethereum were the two most valuable cryptos with respect to market cap with a market value of 160 billion USD (combined) [1]. Bitcoin was responsible for almost $115 billion of the total. Because of their enormous worth, some people regard these currencies as real currencies, while few regard them as openings for investments. As a result, both currencies' values have fluctuated dramatically over short time frames. In 2017, the increase in price of a single Bitcoin was by 2000%, being $862.9 on January 9, 2017 to $17550 on December 11, 2017. On 5th February, 2018, eight weeks later, single Bitcoin price rose to $79643, more than half its prior value. The blockchain is a promising technology that supports cryptos and makes it more probable that they will be utilized in the future in some capacity, and that their use will increase.

Cryptocurrency value volatility creates insecurity for both stakeholders and those who want to utilize it as money instead of as an investment. In comparison to fiat currencies like the US dollar (USD) or the Japanese yen, cryptocurrency is a comparatively recent store of value (example: Bitcoin - launched in 2009 [1]). It's up for discussion what drives price variations in this new store of value. According to researcher LadislavKiroufek, Bitcoin is a one-of-a-kind asset in that its price behaves like both a traditional financial asset and a speculative one [2]. Crypto values are exceedingly difficult to

anticipate because they do not behave like traditional currencies.

We describe an approach for anticipating bitcoin price fluctuations in this paper. Methods based on tweet sentiment analysis are being investigated in order to do so. This includes gathering and storing tweets that mention Ethereum or Bitcoin with the help of Twitter's API and a Python library "Tweepy"[2]. They were then evaluated to generate a daily score of sentiment, which is then compared to the day's price fluctuations to see if there was any correlation between Twitter sentiment and Bitcoin price swings.

[3] and [4] found that online pursuit information (on account of Ettredge et al.) and Google Trends information (on account of Choi and Varian) could be utilized to anticipate an assortment of macroeconomic elements, including auto deals and joblessness rates. We noticed associations among Bitcoin and Ethereum esteems and Google Trends information in the wake of doing our exploration. At long last, we examine the aftereffects of a straight model involving the volume of tweets about every cryptographic money as sources of info.

The results of our research demonstrate that sentiment analysis is less effective in predicting bitcoin price movements while prices are declining. This is because, regardless of price movements, tweets about cryptocurrencies tend to be objective (i.e., without a distinct feeling) or positive. Despite the fact that their worth has risen dramatically in recent years, they are still negligible in terms of usage when compared to traditional fiat currencies. Bitcoins are also a part of a larger technology (the blockchain). As a result, rather than being regarded as a store of value like a traditional stock, Twitter activity surrounding them can be driven by people who have a specific interest in the currency or technology. As prices rise and fall, tweet volume and Google Trends data reflect the cumulative interest in cryptocurrency ownership.

Section 2 provides a quick rundown of crucial points that the reader will need to be aware of in order to follow along with the study. Cryptocurrency, sentiment analysis, Twitter and Google Trends are just a few examples. Section 3 reviews earlier work, covering the diverse study domains on which this report is based. The data collection process and the final data set are described in Section 4. The research techniques and how we chose the model's inputs are discussed in Section 5. The outputs of the model are described in Section 6. Section 7 talks about the ethical implications for using data obtained from the public domain. Section 8 encapsulates our findings as well as future research potential.

2. CRYPTO, TWEETS AND SENTIMENT ANALYSIS BACKGROUND

The research described ahead in this article necessitates a comprehension of where and why the specific data was aggregated, as well as how cryptos differ from typical fiat currency or company stocks traded on regular stock exchanges. In this section, we'll go into the history of these data sources and why they were picked, so that the reader can understand the final analysis.

Cryptos and Blockchain

In this research, we examine market capitalization statistics for the world's two largest cryptocurrencies. Bitcoin is the most valuable, just ahead of Ethereum. Bitcoin was the world's first crypto. Bitcoin was designed by an individual or a group of individuals known as "Satoshi Nakamoto" and was available in 2009[3]. "Satoshi Nakamoto" published a paper titled "Bitcoin: A Peer-to-Peer Electronic Cash System" in conjunction with the launch of Bitcoin, which emphasized on a peer-to-peer transaction mechanism based on electronic currency (crypto) that could be transferred straight from sender to receiver without the need for a third party for verification. This invention has been made possible by the usage of "blockchain," which functions as an open ledger on a peer-to-peer network in which all payments are confirmed by the net and can't be falsified [1].

Blockchain has a wide range of applications outside of peer-to-peer payment systems. Blockchain offers safety, anonymity, and a distributed ledger, making it ideal for healthcare, internet-of-things uses, distributed storage systems, and other applications [5]. As a result of the blockchain's wide range of applications, several more blockchains and cryptocurrencies have been established (currently, 1,658 cryptos exists[4]). Cryptos are related to the blockchain because they offer a financial incentive for machines to run and validate the blockchain, as well as the electricity needed to do so. Cryptocurrencies will become more popular as the use of blockchains grows. This provides them intrinsic value, but how much that

value is determined by a variety of things. Because this is a new sort of money and a new type of store of value, gaining a better knowledge of what can drive price shifts is valuable.

Twitter

Twitter was started in July 2006 as a social media program (which includes other apps/websites like Instagram, Facebook, LinkedIn, and others) as well as a microblogging platform. Microblogging is a type of blogging that allows for more frequent and smaller updates than blogging[5]. Twitter allows users to send public messages (referred to as "tweets") of up to 140 characters[6] in length. That restriction was doubled to 280 characters in November of 2017. Users can also include "hashtags" in their tweets, which consist of the "#" sign followed by a string of characters. This is used to identify a tweet's topic or theme and make it searchable. This will be utilized later in the data section when we gather the tweets.

Twitter has gained in popularity significantly since its inception in 2006. On January 15, 2009, a US Airways airplane crashed into the Hudson River, demonstrating the company's reach and might. Traditional media agencies were not the first to report the news, as an image uploaded on Twitter did[7]. There are 330 million monthly active users on Twitter, 1.3 billion accounts have been created, 83 percent of the world's leaders have a Twitter account, roughly 23 million of Twitter's active users are bots rather than humans, and 500 million tweets are posted every day[8]. As a result of all of these outstanding numbers, Twitter may be a highly rich source of information about how people feel about almost any topic. It's also easy to examine how those feelings alter over time if you know when a tweet was sent. This makes Twitter a useful resource for gathering text data on a topic like cryptocurrency in order to investigate possible price correlations.

Sentiment Analysis

According to estimates, world's 90 percent data was generated in the last two years[9]. Unstructured text data, whether in the form of tweets, internet articles, emails, text messages, or other formats, accounts for a significant amount of that data. Because of this massive volume of unstructured data, "natural language processing" (NLP) has emerged as a field of research or growth. It is a set of techniques for analyzing and comprehending text by computers[10].

We use a set of NLP technologies known as "sentiment analysis" in this paper. It is the process of deriving and assessing subjective feelings or opinions contained in a written material. This can be done in a multiple ways. In this analysis, we used the "VADER" (or "Valence Aware Dictionary and sEntimentReasoner") [6], discussed in further depth in the methods section. The ultimate purpose of this study is to use sentiment analysis to assess whether collected tweets are typically favorable or negative in their views on cryptocurrency. Furthermore, we want to apply sentiment analysis to distinguish between subjective tweets(opinionated) and objective(just data with no polarity).

Google Trends

Nearly every facet of daily life now involves the internet in many regions of the world. Google is by far the most used search engine on the internet, accounting for 74.52 percent of all searches on internet[11]. As a result, Google search data can provide a wealth of information about what the globe is interested in, as well as the extent of it. Data available through "Google Trends" reveals how prevalent certain search phrases are in contrast to others. This serves as a proxy for overall interest in cryptos at any one time, which may have an impact on their prices as interests fluctuates.

3. RELATED WORK

The study incorporates concepts from a variety of studies and themes. Emotion, not just value, has been shown by behavioral economists such as Amos Tversky and Daniel Kahneman to affect decisions, even ones having financial ramifications [7]. Emotions have a strong influence on decision making, according to [8]. These researchers' findings suggest that need for a commodity, and hence value, can be wedged by variables other than basic economic principles, which opens up the possibility of gaining an advantage utilizing technology like sentiment analysis.

Researchers later discovered that information acquired online had an impact on people's purchasing decisions. Galen Thomas Panger established a relationship between people's overall mood and their Twitter emotion. He also discovered that, rather than exacerbating the

end-emotional user's distress, social media such as Twitter had the opposite impact [9]. Chen et al. conducted text based study on "Seeking Alpha," a social media oriented at stakeholders, and discovered that opinions conveyed in "Seeking Alpha" articles were linked to yields and might even forecast earning wonders [10]. In a similar line, [11] discovered that high levels of stock market media negativity had an impact on trading volumes. Finally, according to a survey conducted by Gartner, most buyers rely on social platforms to inspire their purchasing choices [12]. Other academics have looked into the effectiveness of tweet sentiment analysis. Because of the brief nature of tweets and the distinctiveness of the language employed, [13] discovered that typical natural language processing methods (like document and sentence level sentiment score) were useless. Separating tweets into good, negative, or neutral categories can yield useful sentiment assessments, as Alexander Pak and Patrick Paroubek shown [14]. [15] discovered that tweet mood reflected public opinion in national surveys on a variety of themes. Their research revealed sentiment analysis as a cost-effective alternative to national polling, but the implication that sentiments from tweets properly represent the feelings of the larger population on issues suggests that it might also be used to predict demand, and hence product price changes.

Internet data, in addition to social media, has been a fertile research ground. Ettredge et al. were the first to discover that internet search data might be utilized to forecast macroeconomic variables, finding that queries related to employment were connected with unemployment rates [4]. Bordino et al. discovered a link between enquiry volumes and stock trading volumes on the NASDAQ [16]. Hyunyoung Choi and Hal Varian studied Google Trends data and found that basic seasonal auto-regressive models with Google Trends data as inputs outperformed models without Google Trends data by 5-20% [3]. The quantity of tweets regarding newly released movies correctly predicted box-office profits, according to [17].

The correlation between valence scores on tweets concerning S&P 500 firms and stock prices was studied by [18]. Pieter de Jong and colleagues analyzed every minute stock value and tweets for 30 stocks and discovered that tweets were responsible for 87 percent of market returns. They also researched for evidence that stock prices were impacting tweets, but they couldn't find any [19]. A self-organizing fuzzy neural network with Twitter sentiment as an input to forecast price variations in the DOW Jones with 86.7% accuracy was used in [20].

Since the development of cryptos, comparable research has been conducted to see if such algorithms can reliably forecast bitcoin price variations. In the study "Predicting Bitcoin price volatility with Twitter sentiment analysis," it is described how the tweets about Bitcoin and Bitcoin prices for a month (11/05/2017 to 11/06/2017) were collected. Tweets with non-alphanumeric symbols were removed (examples of removed symbols - "#", "@", etc). The research was then refined to remove any tweets that were either irrelevant or overly impactful. Then VADER (or "Valence Aware Dictionary and sEntimentReasoner") was used to classify each tweet's sentiment (3 types). Only the tweets considered as positive or unfavorable were kept in the end [21]. Connor Lamon and colleagues used emotion analysis to estimate price fluctuations in Bitcoin, Litecoin and Ethereum. They discovered that logistic regression did the best in categorizing the tweets, correctly predicting 43.9 percent of price increases and 61.9 percent of price falls [22]. From November 15, 2015 to December 3, 2015, Colianni et al. gathered tweets and categorized them using Naive Bayes and Support Vector Machines, yielding a 255 increase in accuracy [23]. Finally, Shah et al. [24] used historical prices and Bayesian regression analysis to develop a trading method.

DATA

Several distinct sources of data are explored as potential inputs to the model to solve the challenge of predicting bitcoin price changes. The sentiment analysis of collected Ethereum or Bitcoin tweets is the first input taken into consideration. The third factor was tweet volume, which was based on Google Trends data. This section is about how these multiple data sources was collected, cleaned and changed as needed.

Collection of Tweets using Twitter's API

Finding the bitcoin hashtag was the first step towards acquiring the required tweets. Tweepy, an open-source

Python tool for accessing the Twitter API, is used to collect Twitter data. Tweepy allows you to filter by hashtags or words. Cryptocurrencies of interest can be mentioned in a variety of ways in tweets. Using a hashtag ("#") followed by the words "bitcoin" or "ethereum" is the easiest technique. A hashtag plus the currency abbreviation ("#btc" for Bitcoin and "#eth" for Ethereum) is another possibility. Early collections of tweets using only the hashtags "#bitcoin" and "#ethereum" quickly amassed a massive amount of information. Because they had little doubt, these hashtags were chosen as the sole ones we will utilize to collect tweets using Tweepy. For each tweet, the user ID, a permanent identification, the time stamp, and the number of times it was "retweeted" (tweet getting shared by someone so that their followers could see it) and "favorited" were all gathered.

Because tweets can be multilingual, the tweets for this analysis are filtered for English. The tweets were acquired using a script that ran every 15 minutes and captured 1,500 tweets each time. The software was designed to run every 15 minutes automatically to collect tweets for our inquiry. The process was repeated for a total of 60 days, with the amount of tweets collected varying depending on the number of active sessions at the moment. The final tweet dataset had a total of 30,420,054 tweets.

Tweets Processing for Analysis

Additional processing is required after the tweets have been collected. Tweets are constructed in such a way that they don't include enough "information" to perform sentiment analysis.

The sentiment analysis of the collected tweets was done by VADER (or "Valence Aware Dictionary for sEntiment Reasoning"). VADER analysis has the ability to classify text as negative, positive or neutral, as well as quantify the intensity, or polarity, of the words employed. VADER's language and ratings are also customized to microblog and social media situations, which helps us achieve our aims. To remove clutter from the analysis, we cleaned the acquired tweets beforehand.

Tweets have a lot of clutter, such as URLs, hashtags, and emotions. Analysis of Twitter sentiment is challenging due to these characters. How effectively the data is preprocessed defines the efficacy of the future phases in the process of sentiment analysis. Preprocessing generally comprises removing capitals to make it easier to write common terms like such, as well as stemming words to eliminate tense because they all transmit the same idea (eat, ate, and eating all represent similar idea). With the help of freely available pre-processing programs and regular expressions, we were able to accomplish this. Regular expressions are a set of patterns for detecting and removing erroneous text patterns. The # tags, quotation marks, and "?" were removed using regular expressions, which affect the sentiment analysis results. Regular expressions were also used to delete links.

Collection and Adjustment of Google Trends Data

Google has been providing trends data since 2004[12]. It is an impartial set of search data. Rather than search volumes, Google provides a search volume index (SVI). Search volume index is produced using relevant calculations. The proportion of a search phrase across all topics is then scaled from 0 to 100[13]. The SVI that is returned when trends data is searched for more than 90 days is aggregated on a weekly basis. Adjustments were made to compare SVIs across time periods. We conducted our investigation using Erik Johansson's technique.

The approach consists of four main steps. Gather all of the daily SVI data you'll need in 90-day increments, then combine them into a single increment that spans the whole time period of interest. Second, to calculate the weekly SVI, line up the data for the same whole time period but aggregated on a weekly basis. Divide the weekly SVI value by the daily SVI value where the dates overlap to calculate an adjustment factor. To achieve the final result, multiply the daily SVI measurements by the adjustment factor[14]. The Google Trends query produced a result of 1 when the SVI was less than 1. We modified that number to 0.5 to allow for an adjustment calculation. Because Google could not provide any additional information on the actual amount, the halfway value of 0.5 was used.

We used the least confusing terms available, "bitcoin" and "ethereum," to collect Google Trends data, just as we did with tweets. The two currencies' abbreviations, "BTC" and "ETH," weren't used.

Tweet Volume Collection

We used Tweepy's API to collect them, which has a one time limit of 1,500 tweet records. While this produced a huge data collection of tweets that was an arbitrary sample of tweets at the time the algorithm was running, it prevented us from knowing the total number of tweets mentioning bitcoin on any given day.

Since April of 2014, however, www.bitinfocharts.com has provided free access to the daily volume of tweets concerning both of these cryptocurrencies. The entire volume of Ethereum and Bitcoin tweets was calculated using this website.

DATA ANALYSIS

The data was collected, cleaned, and changed as necessary before being assessed to see if it would be a useful input to the final model. Following the cleansing of tweets, there are two difficulties that remain. To begin, it must be assessed how many tweets have any sentiment at all. When the majority of the tweets aren't objective, a sentiment analysis of them adds little to the model. Second, it must be proven that there is a link between the emotion of bitcoin tweets and changes in cryptocurrency prices. Otherwise, the model will be filled with noise. The same can be said for Google Trends data and tweet volume. If there does not appear to be a link between these measures and price changes, they will not be useful model inputs. This section explains the procedures for determining which data are appropriate for use as model inputs.

Sentiment Analysis of Tweets

Humans aren't the only ones who send out tweets. Bots account for a significant number of users and tweets. Bots account for up to 23 million of Twitter's active users, according to estimates. If the tweets were sent by bots with favorable or negative emotion about cryptocurrencies, people's desire may still get influenced to possess cryptocurrencies and, as a result, affect prices. Many of the tweets, on the other hand, are devoid of sentiment and instead deliver merely data or serve as advertisements. Aside from bots, the subject matter is a source of concern. Cryptocurrency discussions are sometimes quite neutral in character. The present price of a single btc token in dollars is a cold, hard fact with no emotional significance. As a result, sentiment analysis of the tweets may only offer the model with limited information. The information gathered from tweets through sentiment analysis is still of little use after pre-processing the collected tweets, according to results of algorithm. Overall, tweets were either generic or ads generated by bots. Only half of the tweets received an objective VADER score on any given day, according to Bitcoin (Fig. 1) and Ethereum (Fig. 2). Other than that, all of the tweets were impartial.

Furthermore, in tweets where extracted VADER score was objective, scores significantly below the 0.5 criteria were observed, whether positive or negative. Gamma kernels for objective and VADER's neutral sentiment scores demonstrate negligible overlap in distributions for all the tweets with a polar score over 0.0. While positive and negative sentiment was found in the Bitcoin (Fig. 3) and Ethereum (Fig. 4) distribution plots, VADER sentiment analysis revealed that tweets were more neutral than objective overall.

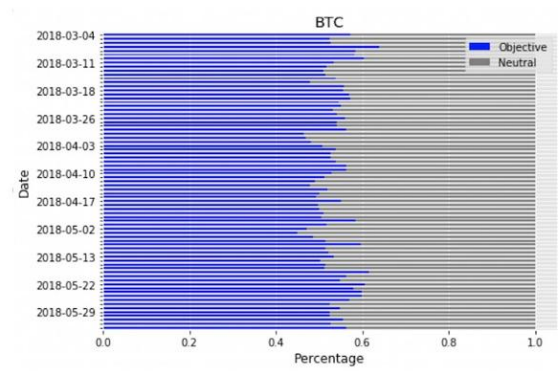

Fig. 1. Bitcoin percent of objective (having a positive or negative sentiment) versus neutral tweets *(blue bar represent objective tweets, gray bars represent neutral tweets)*. Charts created in Python.

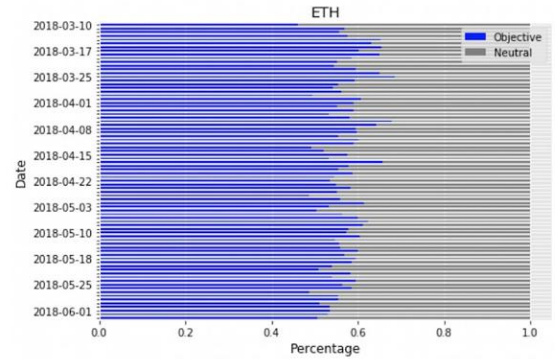

Fig. 2. Ethereum percent of objective (having a positive or negative sentiment) versus neutral tweets *(blue bar represent objective tweets, gray bars represent neutral tweets)*. Charts created in Python.

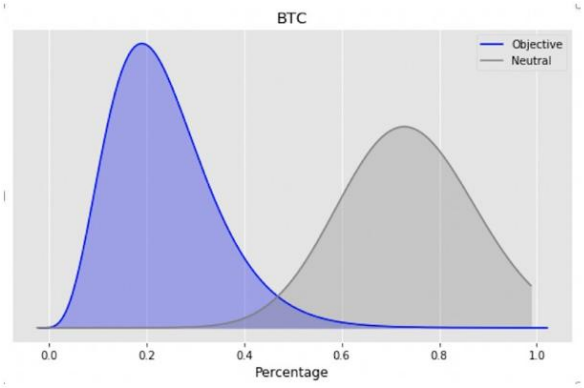

Fig. 3. *Bitcoin objectivity distribution (blue is objective, gray is neutral).* Chart created in Python.

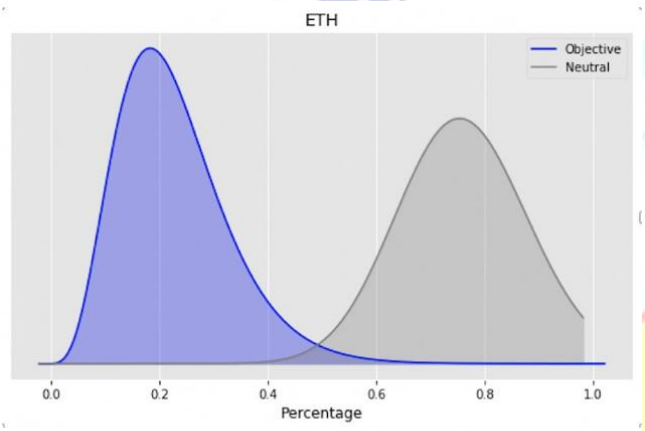

Fig. 4. *Ethereum objectivity distribution (blue is objective, gray is neutral).* Chart created in Python.

Even though only half the number of tweets aggregated contain a non-neutral sentiment, which is insufficient to contemplate the tweet objective in its entirety, the non-neutral sentiment may still offer useful evidence to the model if there is a relation between price changes and sentiment. Figures 5 and 6 depict stats from 4th March 2018 to 3rd June 2018. The blue line, which in both figures corresponds to the left vertical axis, represents the everyday variance (Bitcoin in fig. 5, and Ethereum in fig. 6). The violet line on the vertical axis on the right indicates the sentiment of tweets throughout the day. With 11 of the 19 days sighted a price decline and 8 of the 19 days sighted a price gain, this era demonstrates constant day-to-day price volatility. The tone of the tweets, on the other hand, has been consistent. There was only one day in the case of Bitcoin, March 7, 2018, when sentiment fell below zero. Regardless of price fluctuations, Ethereum sentiment has never dropped below 0.

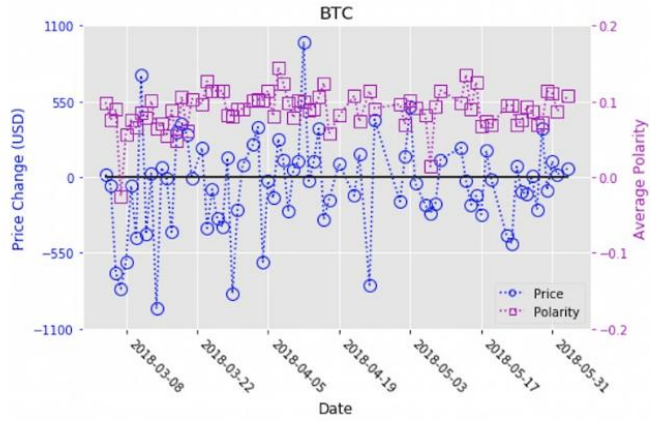

Fig. 5. *Bitcoin price change (blue line, left vertical axis)* and *Daily Average Tweet Polarity (purple line, right vertical axis)* by *date (horizontal axis).* Chart created in Python.

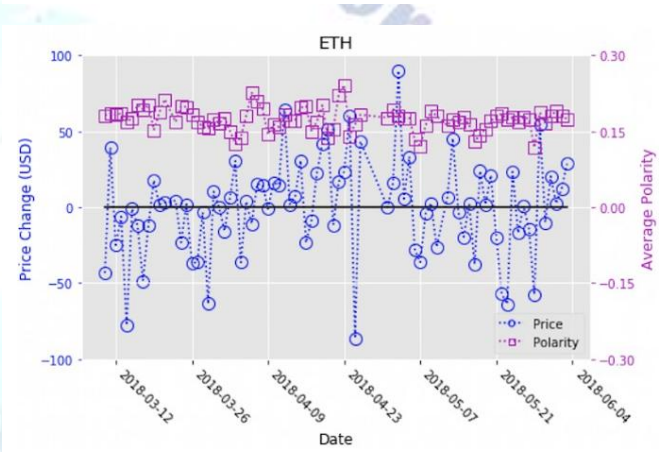

Fig. 6. *Ethereum price change (blue line, left vertical axis)* and *Daily Average Tweet Polarity (purple line, right vertical axis)* by *date (horizontal axis).* Chart created in Python.

The investigation shows that Twitter sentiment is inconsistent with price fluctuations while prices are dropping from March 4, 2018 to March 24, 2018. Because there is no evident link between tweet sentiment and price changes, sentiment analysis won't be employed as an input in the model.

Crypto Prices and Google Trends

A correlation was run for both currencies to see if there was a link between Google Trends search data and cryptocurrency price fluctuations. The "Pearson R" and the p-value are the two basic measures used to establish this. The "Pearson R" is a measure of the association's strength. Its value is between -1 and 1. A positive number indicates that the two variables are positively associated, or that raising one causes the other to increase (we can't say that one variable causes the other to change because this is correlation, not causation). A negative value, on the other hand, denotes a negative relationship between the two variables, or that an increase in one variable's value is accompanied by a decrease in the other. The

p-value denotes the likelihood that these correlation measures were identified by chance. As a result, the lower the p-value, the more confident we can be that a correlation exists and is not a fluke. Figure 7 depicts the relationship between Bitcoin and Google trends data. The price is highly associated with Google Trends data, as seen by the line chart. The pattern is true in both rising and falling price eras. The correlation's Pearson R is 0.817, with a p-value of 0.000.

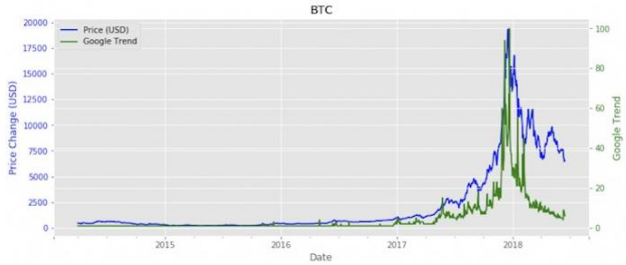

Fig. 7. Google Trends SVI values *(green line)* for Bitcoin after performing Erik Johansson's adjustment to compare daily data across time horizons longer than 90 days. Bitcoin price shown by the blue line. The plot was created in Python using matplotlib.

Crypto Prices and Volume of Tweets

Tweet volume is the last model input to examine. Despite price fluctuations, tweet sentiment tended to remain favorable (when the tweets had any sentiment at all). This could be because people who continue to tweet about cryptocurrencies while prices are decreasing are interested in them for reasons other than their value, such as their anonymity. However, this is a technological fact, not a fluctuating price. This suggests that tweet volume, rather than emotion, is a preferable indicator to use because the number of individuals talking about cryptocurrencies on Twitter may shift with market prices. The Bitcoin price is shown in Figure 8 together with the volume of tweets by day. The graph reveals a strong association, much like Google Trends. It's particularly encouraging that the relationship holds true whether prices rise or fall. With a p-value of 0.000, Pearson R was 0.841.

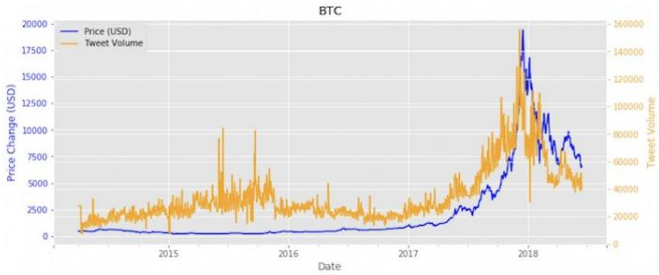

Fig. 8. Bitcoin tweet volumes *(yellow line)* by day. Bitcoin price shown by the blue line. The plot was created in Python using matplotlib.

5. RESULTS

Three model inputs were taken into account. The sentiment of tweets was found to be unreliable as a predictor of dropping cryptocurrency prices, thus it was not included. Price was substantially associated with both Google Trends and tweet volume. Furthermore, the fact that the connection remained during both rising and falling prices suggests that the link is resistant to high variance and non-linearity. A linear model was employed to perform a direct one-to-one comparison because the input variables and the result shared the same non-linear tendencies. Furthermore, considering short-term trends was made possible by using variables for input from the last 15 days as independent inputs.

The model was built using standard machine learning principles, and the entire dataset was split in two, with 80% to train the model and the rest to test it. Although cross validation was excluded in the model, greater research into its use in upcoming projects could help decide whether a more general model emerges. Multiple linear regression was used as the best modelling strategy because of the excellent correlation metrics.

Figure 9 shows model residuals for predicting Bitcoin closing daily price in a linear regression approach utilizing a 15-day time frame of Google Trends data and tweet volume. On the x-axis, actual value is depicted, while on y-axis, model price estimates are plotted. A blue line indicates the exact site of perfect prediction, with olive and cherry colored markers showing whether the point was in the training dataset or testing dataset, respectively. The identical training and testing estimates are shown in Figure 10 on a time series with the definite daily day end price on y-axis. The estimates move up and down the non-linear parts of the time series, as illustrated in the starting input variable correlation graphs.

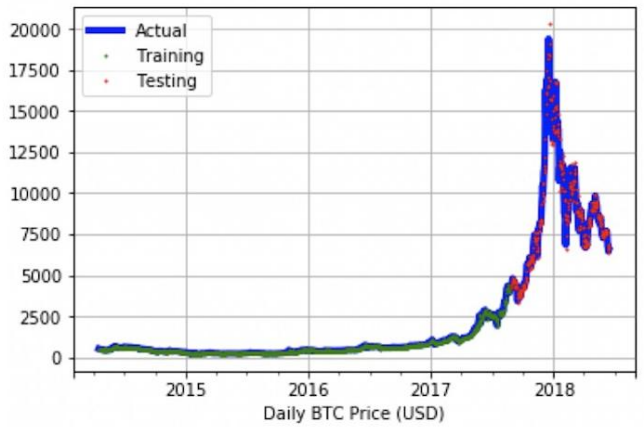

Fig. 9. Model fit shows as actual price *(blue line)* for Bitcoin, training data shown by green dots, and test results as red dots. The plot was created in Python using matplotlib.

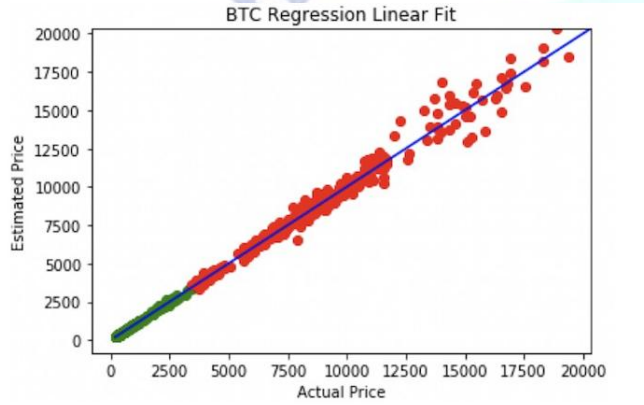

Fig. 10. Bitcoin regression fit shown as estimated price on the y-axis and actual price on the x-axis. Green dots are training data. Red dots are testing results. The plot was created in Python using matplotlib.

ETHICS

Using Twitter data for work raises ethical considerations for the authors' privacy, as well as the ethical need to respect the privacy of individuals who entrust you with their information. The author makes all of the tweets gathered publicly viewable through Twitter. The tweets are also accessible via Twitter's API (or "application programming interface"). Collected data was never saved in areas where they may be accessed by the general public, such as GitHub. Additionally, usernames (also known as "Twitter handles") and any other personally identifiable information were removed when giving examples of tweets for this study.

The cryptocurrencies themselves raise further ethical considerations. They are currently unregulated, which is one of the things that attracts many individuals to them. On the dark side, this allows uncontrolled cash to empower individuals with nefarious motives. The facelessness of cryptos made it appealing as a tender of trade for illicit goods or services[15] on the "dark web" (basically, websites that aren't indexed and can only be accessed with specialized software) [16]. Another part of this is that, because to the lack of regulation, their value may be regarded as less predictable. The availability of a model to forecast crypto price fluctuations may lead to a false sense of trust in the model's effectiveness. As a result, people may be more inclined to store wealth using cryptocurrency.

Finally, the model allows for manipulation of others to a level that it is beneficial in anticipating upcoming crypto price fluctuations. This could be accomplished by persuading others to put money by same strategy as you have because of the model, or by releasing misleading information to improve the model's advantage. Any data given should be as accurate as feasible. Furthermore, the caveats and danger of the information should be communicated to the greatest extent possible and appropriate so that every individual may make educated choices.

6. Conclusions and Future Work

Prior attempts to forecast bitcoin variations used Twitter sentiment research as a substitute for upcoming crypto demand, resulting in rising or falling prices. We've proven that these conclusions were influenced by the fact that the study took place during a period when cryptocurrency values were steadily rising. Furthermore, regardless of potential price changes, Twitter attitude around cryptocurrency is generally positive. People that tweet about cryptocurrencies, even when their prices fall, are interested in them for reasons other than investment, thus the tweets are skewed toward the positive. An extra robust model would include a volume-based measure of cumulative interest. The use of proxies for broad interest, such as Google Trends or tweet volumes, is recommended in this paper. We've established that the search volume index, as well as tweet volumes, are substantially connected with cryptocurrency prices, both when they rise and when they decline. Future research should look into whether these findings hold true in a variety of pricing scenarios.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

[1] Nakamoto, S.:Bitcoin: A peer-to-peer electronic cash system.In: CryptographyMailinglistathttps://metzdowd.com.(032009)

[2] Kristoufek, L.:What are the main drivers of the bitcoin price? Evidence fromwaveletcoherenceanalysis.PLOSONE10(4)(042015)1–15

[3] HYUNYOUNG,C.,HAL,V.:Predictingthepresentwithgoogletrends .EconomicRecord88(s1)2–9

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

[5] Miraz, M.H., Ali, M.:Applications of blockchain technology beyond cryptocurrency.CoRRabs/1801.03528(2018)

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

[7] Kahneman, D., Tversky, A.:Prospect theory: An analysis of decision under risk.Econometrica47(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.:Customersasadvisors:Theroleo fsocialmediainfinancialmarkets(2011)

[11] Tetlock, P.C.: Giving content to invsotry sentiment: The role of media in the stockmarket.TheJournalofFinance(2007)

[12] Gartner:Gartnersaysmajorityofconsumersrelyonsocialnetworkstog uidepruchasedecisions(2010)

[13] Kouloumpis, E., Wilson, T., Moore, J. In: Twitter Sentiment Analysis: The GoodtheBadandtheOMG!AAAIPress(2011)538–541

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

[15] O'Connor,B.,Balasubramanyan,R.,Routledge,B.R.,Smith,N.A.:Fro mtweetsto polls: Linking text sentiment to public opinion time series.In: Proceedings oftheFourthInternationalAAAIConferenceonWEblogsandSocialM edia.(2010)

[16] Bordino,I.,Battiston,S.,Caldarelli,G.,Cristelli,M.,Ukkonen,A.,Webe r,I.:Websearchqueriescanpredictstockmarketvolumes.PLOSONE7(7)(072012)1–17

[17] Asur, S., Huberman, B.A.:Predicting the future with social media.In: 2010IEEE/WIC/ACM International Conference on Web Intelligence and IntelligentAgentTechnology.Volume1.(Aug2010)492–499

[18] Sul, H., Dennis, A.R., Yuan, L.I.:Trading on twitter: The financial informationcontentofemotioninsocialmedia.201447thHawaiiIntern ationalConferenceonSystemSciences(2014)806–815

[19] deJong,P.,Elfayoumy,S.,Schnusenberg,O.:Fromreturnstotweetsand back:An investigation of the stocks in the dow jones industrial average.Journal ofBehavioralFinance18(1)(2017)54–64

[20] Bollen, J., Mao, H., Zeng, X.: Twitter mood predicts the stock market. Journal ofComputationalScience2(1)(2011)1–8

[21] Stenqvist,E.,L¨onn¨o,J.:Predictingbitcoinpricefluctuationwithtwitte rsentiment analysis.KTH Royal Institute of Technology School of Computer Science andCommunication(2017)3–28

[22] Lamon, C., Nielsen, E., Redondo, E.: Cryptocurrency price prediction using newsandsocialmediasentiment.Master'sthesis,Standord(2015)

[23] Colianni, S., Rosales, S.M., Signorotti, M.:Algorithmic trading of cryptocurrencybasedontwittersentimentanalysis(2015)

[24] Shah, D., Zhang, K.: Bayesian regression and bitcoin. 2014 53nd Annual AllertonConferenceonCommunication,Control,andComputing(Al lerton)(2014)409–414

[25] Kimoto, T., Asakawa, K., Yoda, M., Takeoka, M.: Stock market prediction systemwith modular neural networks. In: 1990 IJCNN International Joint Conference onNeuralNetworks.(June1990)1–6

[26] Guresen,E.,Kayakutlu,G.,Daim,T.U.:Usingartificialneuralnetwork modelsin stock market index prediction. Expert Systems with Applications 38(8) (2011)10389–10397

[27] Zhu, X., Wang, H., Xu, L., Li, H.:Predicting stock index increments by neuralnetworks: The role of trading volume under different horizons.Expert SystemswithApplications34(4)(2008)3043–3054