



# How can we predict the impact of the social media messages on the value of cryptocurrency? Insights from big data analytics



Chahat Tandon<sup>a</sup>, Sanjana Revankar<sup>a</sup>, Hemant Palivela Dr., Ph.D, Head of AI and Machine Learning<sup>b,b,\*</sup>, Sidharth Singh Parihar<sup>c</sup>

<sup>a</sup> Department of Computer Engineering, Bapuji Institute of Engineering and Technology, Davangere, India

<sup>b</sup> Head of Artificial Intelligence, eClerx Services Limited, Airoli, Navi Mumbai, Maharashtra 400708, India

<sup>c</sup> Data Scientist at Dynacare, Brampton L6T 5M3, Canada

## ARTICLE INFO

### Keywords:

Bitcoin  
Dogecoin  
ARIMA  
Blockchain  
Price prediction  
Cryptocurrency  
Social media analytics  
Text mining  
Time series analysis

## ABSTRACT

Cryptocurrency and blockchain are one of the most beautiful digital transformations occurring around the world. They have changed the orthodox meaning and working of currency as we know it. It is interesting to note how it excites and worries some. The main reason for the popularity of cryptocurrencies is tremendous returns in very little time. Social media platforms like twitter, provide a safe-place where individuals' can share their thoughts as well as mindsets, which then can be heard and be reciprocated by others. This paper aims to draw a correlation between the hyped tweets and the prices of cryptocurrencies like Bitcoin - The Crypto King and Dogecoin - The Memecoin during those times. We also aim to predict the future price values of Bitcoin using its past values. By using cryptocurrencies' financial data, twitter data, RAPIDS and cuml, a fine line can be drawn between the amount of impact tweets have on people as well as on the market. The tweets on cryptocurrency were segregated and price forecasting was done using augmented dickey fuller test and ARIMA models, 10 future values of bitcoin were predicted with 96% accuracy and 0.0395 average error. Besides, from the investigations above of the authentic cost of BTC, it is perfectly clear that there have been way more steep falls in the history of Cryptocurrencies even before Elon started tweeting about it. Thus, it can clearly be stated that no one person can control the utter volatile world of cryptocurrencies! And the decentralized system ledger of cryptocurrency remains unharmed.

## 1. Introduction

Cryptocurrency is a very broadly classified topic. It is a virtual currency. A currency that could be considered relevant and real in the form of "virtual coins". Halaburda (2018) Cryptocurrency is a concept that works on a decentralized ledger. Some of the most famous cryptocurrencies are Bitcoin, ADA, Dogecoin, Ethereum, etc. The development and interest were basically brought about by news stories which revealed the exceptional returns of digital currencies, which in this manner pulled in a kind of dash for unheard of wealth. All the while, current worldwide guidelines on digital forms of money are exceptionally restricted, as digital currencies are not yet recognized as a developed resource class. The blockchain community has grown into many fields such as medically extracting coronavirus by lung images Mittal, Gupta, Chaturvedi, Chansarkar, & Gupta (2021), Jun, Yoo, & Choi (2018). Block chain Yli-Huumo, Ko, Choi, Park, & Smolander (2016) innovation is progressively making footing changes in different regions like money, Mettler (2016) production network the executives, and cloud

administrations due to its capacity to give hearty online protection natural in its arrangement of having decentralized information stockpiling Chauhan, Palivela, & Tiwari, 2021. Pilkington (2016) The rising intricacy in the engineering of well known block- chain stages make obstructions to address selection of the innovation. With the advent in technology such as Tandon et al. (2021) Natural Language Inference or NLI which adds extraordinary advantages to already existing Natural Language Processing use cases available as well as the effective ways of dealing with problems caused by the pandemic Bongale, Tandon, Palivela, & Nirmala (2021); the blockchain becomes a necessary addition to today's world. Cachin et al. (2016) It becomes basic that educational instruments are accepted in the blockchain environment to address this apparent or genuine obstructions for the take-up of the innovation. By using its characteristic properties of decentralization and complete encryption, Kshetri (2017) Blockchain can ensure unapproved information access while its on the way and being utilized in applications like Social Media Analytic Grover, Kar, & Janssen, 2019. This paper mainly concentrates on the trade-offs between the two most pop-

\* Corresponding author.

E-mail addresses: [chahat.7874@gmail.com](mailto:chahat.7874@gmail.com) (C. Tandon), [SanjanaRRDVG@gmail.com](mailto:SanjanaRRDVG@gmail.com) (S. Revankar), [hemant.datascience@outlook.com](mailto:hemant.datascience@outlook.com) (H. Palivela Dr.).

ular cryptocurrencies of 2021, Bitcoin and Dogecoin [Nofsinger \(2005\)](#). Twitter is widely used as a big source of information sharing and ideologies sharing platform. The tweets from people showcase [Nakamoto \(2009\)](#) the sentiments towards any specific topic. Similarly, by mining tweets from famous personalities like Elon Musk, a comparison can be put forward so as to understand the effect of centralized image building through the use of social media such as Twitter onto the world of “decentralized” cryptocurrencies. Bitcoin is more of a puzzle. Bitcoin is one of the most popular and leads the market in terms of capitalization as well as the user base of the world. [Dinh et al. \(2018\)](#) Bitcoin started a revolution that is being carried forward by many others. The concept of bitcoin can be explained very easily. Let’s treat it as a puzzle. [Kshetri & Voas \(2018\)](#) All the computers in the world are adamant to solve that puzzle. The puzzle needs to be solved in exactly ten minutes. There is no central ledger in the whole process. Bitcoin works on a distributed ledger. Bitcoin makes use of [Di Pierro \(2017\)](#) Blockchain to carry out its functions. Every block in the blockchain has a certain history of transactions done so far. Every new transaction or function needs to be placed in the blocks of the blockchain. These blocks are then added to the blockchain. A single block contains at least twenty five hundred new functions. Given such a high number, validation of each block is necessary. This validation is done by the crypto miners. To validate the block, each one has the answer to the last block. Whoever solves the puzzle first, gets a reward of twelve and a half new bitcoins. Then the whole process starts again. Every ten minutes a new puzzle is released. It is made sure that it takes exactly ten minutes to solve the problem. The bitcoin algorithm adjusts the difficulty of the problem at hand based on how many computers are on the hunt for it [Tasatanattakool & Techanupreeda \(2018\)](#).

More computing power equals more security. As computational energy goes up, so does the complexity of the whole process. You need Graphics processing unit also known as a GPU or an application-specific integrated circuit (ASIC) to be able to set up a mining rig. Only 27 million bitcoins are available in nature. As soon as that number is reached, which has been predicted by 2041; the bitcoin mining will stop. Past that, the field of cryptographic forms of money has extended drastically since Bitcoin was dispatched longer than 10 years prior, and the following incredible advanced token might be delivered tomorrow. In [Chohan \(2017a\)](#), Chohan, U.W. addressed the historical evolution of Bitcoin With the perspective of investors. [Cocco, Concas, & Marchesi \(2017\)](#) proposed a model that was able to entail many statistical properties of the actual returns proposed by the market. The model proved to be a starting point to analyse the cryptocurrency market. Crypto Economics can also be an application of crypto-democracy as discussed by [Aste, Tasca, & Di Matteo \(2017\)](#). A study on Bitcoin and Ethereum all the while exploring features in their network that explain their price hikes was done by [Saad, Choi, Nyang, Kim, & Mo-haisen \(2020\)](#).

Bitcoin Cash (BCH) holds a significant spot in the historical backdrop of altcoins right now. In the light of the fact that it is one of the earliest and best hard forms of the first Bitcoin. Because of the decentralized idea of advanced monetary forms, discount changes to the code hiding the token or coin within reach should be made secure. In 2008, the father of Bitcoin [Nakamoto \(2009\)](#) had published his proposal for the vision of electric cash flow, that we now know as cryptocurrencies. The use of digital signatures was encouraged as a third party authentication solution. The use of blockchain as well as how the longest chain was obtained just by the use of the most powerful CPUs in the world was showcased. Eventually, the longest proof-of-work chain was accepted as a proof. Bitcoin officially came into existence in the year 2013–2014. The now world-wide famous cryptocurrency became almost 5 times the US dollar by the late 2013 and has been a part and parcel of the crypto world ever since. In “Is Bitcoin a real currency? An economic appraisal” has put forward an extensive approach to the excessive volatility of bitcoin. He extensively put forward views on the risk management transactions and how there is no proper way to limit the growth of bitcoin against other

currencies. At last, bitcoin faces a drawn out underlying financial issue identified with the total breaking point of 21 million units. Thus, the future can be economically fathomable. Digital forms of money need the two components, what makes the market be very unpredictable and presently make digital forms of money inadmissible as a dependable stockpiling of significant worth or a mode of trade [Baur, Hong, & Lee \(2018\)](#). Cryptocurrencies like Bitcoin can’t be afforded by all potential investors, so they tend to be drawn towards alternative coins like dogecoin, ethereum, cardano etc.

Doge coin was first created by Billy Markus, a programmer from Portland, Orlando. He introduced it as a joke cryptocurrency and really made it for fun. Those days i.e. in 2013, a popular meme on shiba inu was in trend which led to creation of fun coin [Chohan \(2017b\)](#). Later on, Jackson Palmer, a software developer got inspired by a tweet from a student and set out to develop dogecoin to make it a reality. The first dogecoin block was mined on 6th December 2013. In the starting years, the time taken to mine it was about 4 h but as time passed on, it can be mined in 1 min as of now. Technically, dogecoin is based on proof of work of lucky coin which is in-turn a fork of litecoin. So the miners can mine doge using the same method as mining lite, there is no original block structure for doge. Reward awarded per block of doge block mined was ranged between 0 to maxB were insane. At times, people started blaming founders that they might keep the highest rewards for themselves. So to prevent it, the founders came up with the conclusion of rewarding 10,000 coins per doge block. Doge was started with an original supply of about 100 billion coins. Now the dogecoin limit is made unlimited. And every year about 5.2 million coins are mined since the creation rate is 1 min per block. This implies the inflation rate of doge is 5% which is almost double of litecoin with 2.4%. Concerned about security, the dogecoin network is as secure as lite because it uses the same proof of work. But there was a time as said by [Baur & Dimpfl, 2021](#), the doge network was attacked by hackers and stole about millions of coins from investor accounts. The Doge community overcame this disaster by helping the investors who lost their coins. Must say, dogecoin has an active and loyal community!. There are some instances to prove it - In 2014 doge community raised about 30,000 USD for Jamaican bobsleigh team to the Sochi Winter Olympics in Russian Federation and also they raised funds of about \$50,000 for Kenya water crisis. Doge coin was meant to be used for exchange/ trading such as Doge/LTC (litecoin).

## 2. Related work

Today many exchange communities, softwares are including doge for trading. The first exchange happened in the year 2014 by AltQuick.co. Doge can be legally traded in countries including South Korea, UK, HongKong and so forth. During the year 2017–2018 doge peaked to \$0.017 per coin which generated about 2 billion total market capita. Later on in 2020, tik tok started a trend to make doge worth 1\$. It might be a global currency for Gen Z. In Feb 2021, the total capita leveraged to 10.2 B and is still rising. Another thought that always strikes our mind when we speak about doge is Elon Musk. He has been called “Godfather of Doge” due to his tweets, memes and concern about doge. He also announced that SpaceX will accept doge in future in his recent tweets and will take it to the moon. The Doge community uses this phrase to show their love towards doge. People get influenced way too easily by Multinational Corporations in the name of faith and might even end up gambling away their life savings. Having said that the market volatility of bitcoin and ethereum is extreme and almost 10–15 times the volatility than any other stock market entity, [Baur & Dimpfl, 2021](#) put forward an approach on how and what effects can be observed on personal portfolios. Bitcoin has not arrived at the mass market selection rates that would be important to give alternative worth to enormous holders of the money [Chohan \(2017b\)](#). The same goes for Dogecoin as well. But is there a way for any individual to influence the market in such a way as to derive a way out of this safely? High-Profile Losses Raise Fear. Are

the long shots trying to leverage that fear and use it to dominate the cryptocurrency world?

In paper titled, [Chu, Chan, Nadarajah, & Osterrieder \(2017\)](#); currencies like Bitcoin, Dash, Dogecoin, Litecoin, ADA, Monero and Ripple were fit on twelve diverse GARCH-type models in order to track paces of each of these digital currencies. This helped calculate the exchange rate; maximum likelihood as well as the value at risk. Diffusion of blockchain in insurance industry also became a topic of discussion at that time [Kar & Navin \(2021\)](#). A study showcasing short-term prices based on Twitter and Google Trends was done by [Wok \(2020\)](#). A later study using long short-term memory (LSTM) to predict the Bitcoin exchange was done without considering the previous exchange rate [Chen, Xu, Jia, & Gao \(2021\)](#). The CoinDesk Releases Q2 2018 State of Blockchain Report stated that HODLing could actually result in very good Return On Investment rates if done as per the market. Various fiat-to-crypto volumes were talked upon along with trade pairs which proved to be about 26% down from the crypto-to-crypto volumes. The sentimental survey on the investments made were also taken into account. Ethereum along with Bitcoin proved to have good RoI future predictions. Used tweets from twitter to help predict the box-office revenues for movies. They also showcased how sentiment analysis leads to better forecasting of human behaviour as well as the probable revenue.

### 2.1. Limitations of current literature

With the coming of block chain and digital currencies being however new and promising as it could be, anticipating the six-year extended worth of Bitcoin, Ethereum and Litecoin requires various variables to be thought of. [Nakamoto \(2009\)](#) Bitcoin can use its enormous over the top clientele base and try to predict use cases that indicate further development in the future skyline. Ethereum's wide scope of results, both positive as well as drastic, demonstrates that it ought to be remembered for the speculation portfolio to exploit this reality. With Dogecoin showing lesser variety than Ethereum, can be liked as a new alternative. Seeing the volatility of the market, it becomes very difficult to actually predict how any currency would perform. No one seems to understand the influence of CEOs / COOs of Multinational Corporations on the psychology of cryptocurrency trade offs.

### 2.2. Hypothesis/statement of purpose

Given the high volume tradeoffs and cryptocurrency hypes, specially in the year 2020; we hypothesised a research to find out the correlation between digital currencies and the influence of socialites on it. Elon Musk made \$272 million this year on the profit on Bitcoin alone. Tesla paid \$171 million for Bitcoin it sold for \$101 million more, clinching a 59% profit. Along with Tesla accepting Bitcoin as a payment method and Elon tweeting major updates and investing plans on Twitter; he proved to be a compelling use case for our study. We plan to investigate two main cryptocurrencies for this study :

- Bitcoin
- Dogecoin
- Are cryptocurrency prices and tradeoffs affected by the fluctuation in Elon Musk's tweets?
- Can the price of bitcoin be forecasted?
- Can Dogecoin or Memecoin actually become a currency of the future?

## 3. Materials and methods

### 3.1. Dataset

The problem statement of this research is mainly focused on the impact of tweets on pricing of cryptocurrency. Thus, Elon Musk's tweets from 2009 till 2021 were considered. These tweets were scraped using

this cool python library called twint which stands for Twitter Intelligent Tool. Twint is capable of scraping tweets of any certain user, hashtags, topics and trends. This does not use any API so the fuss can be avoided. This is a very useful tool. The scraped data contains a total of 38 parameters including email, tweet, id's, time etc. The final shape of the tweet dataset we considered was 12,562 samples with 15 parameters.

Bitcoin is a well-known and trending crypto-currencies of all time. Trading data on bitcoin is available all over and can be scraped using API's. We considered historical bitcoin market data of 1-min interval bitcoins' market trading. The scraped data contains the time duration from January 2012 to March 2021 with updates for each minute on high, low, open, close prices at the time window, volume, indicated currency, weighted bitcoin price and also time stamp. If there is no considerable activity or jumps in the price then that value of that sample is taken as NaN. All efforts were made to deduplicate the data and make the dataset stable for use. This is one huge dataset with total samples 4,857,377 with 8 parameters.

Another crypto-currency here is Dogecoin which is favoured by Shiba Inus. Dogecoin data is almost identical to Bitcoin data in terms of parameters. This dataset contains historical stock prices in USD starting from September 2014 with total samples 2418 and 7 parameters.

### 3.2. Data pre-processing and feature selection

To carry on with further analysis in any problem, quality data plays a significant role. Twitter data is one of the highly unstructured data and contains high noise. So, extensive data pre-processing became a mandatory step. An array of about 18 unicode regular expression matching were considered for preprocessing. Unicodes for emoticons, symbols and pictographs, transport and map symbols, flags, chinese characters, dingbats etc. These patterns were matched and replaced with a blank space. Furthermore, the cleantweets function is used for cleaning the remaining characters in text. For instance, removal of "/n", "/t", "/xa0", "pic.twitter", "http://" + "https://" links, mentions and hashtags, punctuations, making lowercase and stripping the whitespaces. To illustrate, consider the following tweet - "Let us make Dogecoin as our Future currency.@Dogecoin", after omitting the hashtags, mentions and emojis we will be left with - "Let us make Dogecoins as our Future currency." as a processed statement.

Parameters are nothing but features of the dataset. We have come across many features in the dataset but all of those are not required to build our model. It is desirable to reduce features for both reducing computational power and to improve the efficiency of the model. We considered some few important features in Elon Musk's tweet dataset like - id, date, time, username, tweet, mentions, urls, photos, hashtag, replies\_count, retweets\_count, likes\_count and link. These features were considered in order to visualise his (Elons') popularity based on the tweet count, likes, retweets etc. Furthermore, we split the date of the tweet in terms of its year. Bitcoin historical data had some missing values, the best way to handle them was to just drop those samples. Thereafter, we proceeded with visualization.

### 3.3. Intro to baseline models/materials

#### 3.3.1. RAPIDS

The dataset we scrapped contains more than 4 million samples (bitcoin data) and it is not easy to handle such huge data, so we made use of a library called RAPIDS, that provides GPUs for easy handling of data reading, transforming, validation etc. It is well known that GPUs are much faster than CPUs. That is why we can save time and money in turn increasing the model's accuracy.

#### 3.3.2. Augmented dickey fuller test

Since our bitcoin data is a time series data, it is important to check whether the data has any kind of stationarity. To check this we considered ADF which is an acronym for Augmented Dickey Fuller test, is

unit root test. These unit roots are the reason for causing unpredictable results in time series data analysis. This ADF is a significance test, so a null and alternative hypothesis comes into play by that, test statistics are calculated and  $p$ -value is reported. Based on  $p$ -values, the stationarity of data is determined. Basically, ADF determines the trend of data and determines how strongly or weakly the time series is defined by a trend. Stationarity or stationary data means that data's statistical properties do not depend on time. And if suppose the given data is non-stationary then we need to change it to stationary by applying certain methods including differencing, applying log etc.

### 3.3.3. ARIMA

Model we choose to use is ARIMA, short for AutoRegressive Integrated Moving Average. ARIMA is a part of linear regression models, mainly used for predicting future values based on the past behaviour of the target. It is said that history doesn't repeat itself but it surely has and had its own rhythm. The beauty of ARIMA models is that these do not use any exogenous values imposed on them but rather, are completely dependent on the past target values for prediction. ARIMA can be broken as AR, I and MA. As mentioned above, AR stands for Auto-Regressive, it works on the idea of regressing target on its past variable which is nothing but lagging on itself. Eq. (1) indicates a value  $Y$  is a linear function of its past  $n$  values. These  $n$  values can be chosen and  $B_0, B_1$  are beta values used during fitting of the model. This equation can be made to forecast future values by simply making the following changes as in Eq. (2).

$$Y = B_0 + B_1 * Y_{lag1} + B_2 * Y_{lag2} + \dots + B_n * Y_{lagn} \quad (1)$$

$$Y_{forward1} = B_0 + B_1 * Y + B_2 * Y_{lag1} + \dots + B_n * Y_{lag(n-1)} \quad (2)$$

The Integrated part in ARIMA deals with stationarizing of data. Here, the differencing is applied on data as shown in Eq. (3), it indicates the future values of  $Y$  is some linear function of its past changes. Reason for differencing is that the time series data is not stable and the values of  $Y$  should have mean variance stationary.

$$Y_{forward1} - Y = B_0 + B_1 * (Y - Y_{lag1}) + B_2 * (Y_{lag1} - Y_{lag2}) + \dots \quad (3)$$

Moving Average is all summarized in Eq. (4), somewhat similar to AN equation with lags.  $E$  indicates the error in the data is nothing but the residual derivations between the model and target value.

$$Y = B_0 + B_1 * E_{lag1} + B_2 * E_{lag2} + \dots + B_n * E_{lagn} \quad (4)$$

ARIMA( $p, d, q$ ) is the standard notation to represent ARIMA models. These parameters can be replaced with integer values to specify the type of the model used. Parameter ' $p$ ' is referred as the lag order of AR that's the number of lags in  $Y$  to be included in the model, ' $d$ ' is the order of differencing required for making data stationary and ' $q$ ' is referred as order of MA that's number of lagged forecast errors.

### 3.3.4. Criteria

There are several different models (i.e. based on the orders like AR and ARMA model if certain order or different order) for building time series. Lower the value obtained using these criteria, better suitable will be the model for our time series data. Parameters used in these criteria include - Log likelihood ( $L$ ), signifies how strong the model is in fitting the data. Generally, it is considered that the more complicated the model is, the better it fits the data. It is true though in fitting but also leverages the concept of overfitting (i.e. model fits the training data better but loses its ability of generalizing on test data). To prevent that, incorporate the number of predictors  $k$  (i.e. number of lags (fixed amount of passing time)) plus a constant. Another parameter to consider here is  $T$ , number of samples or observations used for estimation.

Below listed are the criteria used in this experiment for selecting best model-

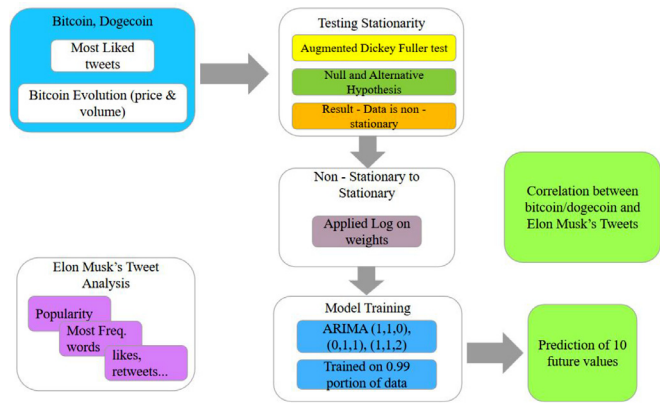


Fig. 1. Block diagram of the experiment.

- Akaike's Information Criterion (AIC) - AIC is used to determine the order of an ARIMA model and also it can be useful for selecting the predictors of the regression model. AIC can be calculated using the formula given below,

$$AIC = -2 \log(L) + 2(p + q + k + 1) \quad (5)$$

In cases with small values of  $T$ , AIC may tend to predict too many predictors so in order to prevent that bias corrected version of AIC i.e. AICc should be considered.

$$AICc = \frac{AIC + 2(p + q + k + 1)}{(p + q + k + 2)(T - p - q - k - 2)} \quad (6)$$

where,  $L$  is likelihood value,  $p$  is order of AR model,  $q$  is order of MA model,  $k$  is number of predictors and  $T$  is number of observations which is used for estimation as mentioned above. In order to get the best model, we need to consider the model with low AIC value. That means, the value of  $k$  should be less and the value of  $L$  should be maximum, illustrating that the model will be simple as  $k$  is small and fitting the data well with max  $L$ .

- Bayesian Information Criterion (BIC) - BIC also known as Schwarz information criterion is used for model selection on score obtained.

$$BIC = AIC + [\log(T) - 2](p + q + k + 1) \quad (7)$$

Here also minimum value should be taken into consideration. BIC with small value illustrates that the model is simple with relatively less number of  $k$ , best fits the model and it is trained on a few number of observations.

## 4. Methodology

The methodology shown in (Fig. 1) consists of the flow used for the study. Information such as most liked tweets and price/volume data was collected about both Bitcoin as well as Dogecoin cryptocurrencies. After this the future prediction of the pricing was done, which has been explained in detail in the (Section 4.2). The final step was Tweet analysis and drawing the conclusion to the question "Does social media influence by well known individuals correlate to actual tradeoffs in the crypto world?"

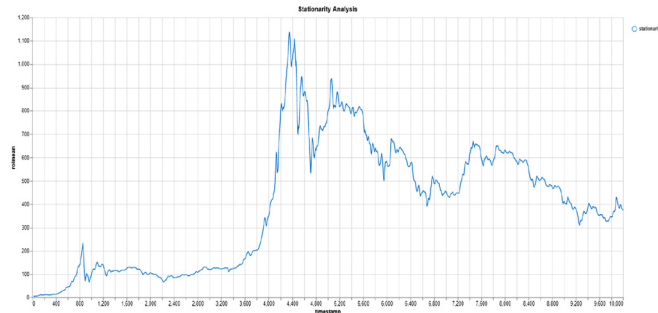
### 4.1. Popularity of Elon Musk

Elon Musk is founder of well known organizations SpaceX and Tesla. Now-a-days, his tweets are making a huge impact on not only his company but also the market whether it's the stock or share market. We analyzed the reason behind his popularity over these years by visualizing his tweets based on tweet count, likes, retweeted tweets, most used words etc. We also mainly considered his tweets on Bitcoin and Dogecoin for analysing his influence on the prices on BTC and Doge.

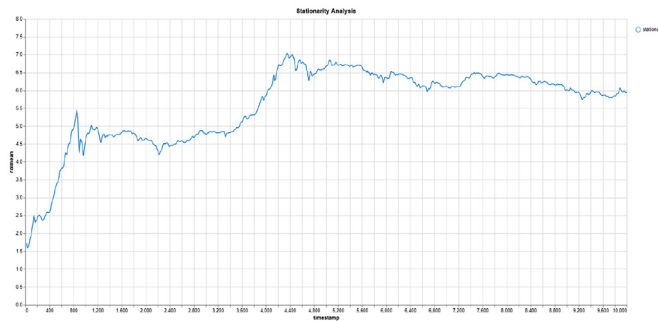


**Table 1**  
Comparing stats values of stationary and nonstationary data.

	ADF statistics	p-value
<b>Stats of Non-Stationary Data</b>	4.695348	1.000000
<b>Stats of Stationary Data</b>	-3.434271	0.009845



**Fig. 2.** Trend analysis of non-stationary data.



**Fig. 3.** Trend analysis of stationary data.

#### 4.2. Predicting future of bitcoin pricing

Considering historical stock data of bitcoin, we first analyzed the price and volume evolution over years. Since stock data is a time series data, it's necessary to check this stationarity for model building.

##### 4.2.1. Stationarity testing

We used the ADF test (Section 3.3.2) for determining the property of our time series variable. To do so, we considered null and alternative hypotheses as- The representation of test data can be done using unit root or is not - stationarity, the test data is stationary respectively (Table 1).

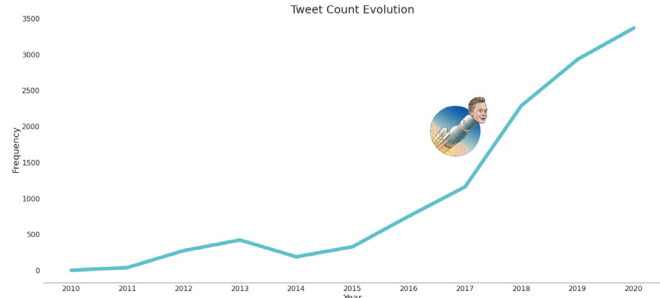
ADF statistics, p-value, critical values, rolling mean and standard deviation of time series variables were calculated. The result was not - stationary that is the data shows some trend. To make it stationary and remove the trend we simply applied natural logarithm on weighted price values of bitcoin. The values of time series before and after removing trends can be found in the table. Figs. 2 and 3 illustrate the non-stationary data and stationary datas' rolling mean v/s timestamp line plot respectively.

##### 4.2.2. Model training

As given in Section 3.3.3, ARIMA model predicts future values based on past behaviour. We considered certain criterion's (Section 3.3.4) to know the condition of our model after training and the loss of information during training and to select the best model. Minimum loss indicates better training. We considered 2 models (ARI and MA) at a time to tweak these criterias and get the best ARIMA model for training. For the first type, we considered  $p = 1$  and  $d = 1$  (Section 3.3.4) and calculated

**Table 2**  
Criterion values for ARIMA model orders .

	Log-likelihood	AICc	BIC
<b>Auto Regressive(1,1,0)</b>	-230,366.49	460,738.98	460,764.46
<b>Moving Average(0,1,2)</b>	-230,338.28	460,694.57	460,738.57
<b>ARIMA(1,1,2)</b>	-230,338.2	460,686.41	460,728.88



**Fig. 4.** Tweet count evolution from 2009 to 2021.

low-likelihood, AICc and BIC values. Then for the second type (MA), we considered  $d = 1$  and  $q = 2$  and criterion values of both the models can be found in the Table 2. Both the models together form ARIMA, so we choose  $p = 1$ ,  $d = 1$  and  $q = 2$  that is ARIMA(1,1,2) model for training since this gives minimum criterion values. We choose about 0.99% of total bitcoin stock data for training which is 36,148 samples in which we considered to predict 10 future values of bitcoin pricing. We then calculated error (Eq. (8)) for each predicted future value based on forecast and actual values. Average error of prediction was 0.03957.

$$Error = \frac{abs(forecasted\ value - testprice)}{testprice} * 100 \quad (8)$$

## 5. Results

### 5.1. Elon Musk tweeting over the years!

Tesla's Powerwall battery was the advent of Elon Musk's Ante (2021) popularity on Twitter along with his biography coming out in the year 2015 (Fig. 4). It appears to be that his notoriety developed with the quantity of tweets. He expanded progressively in likes with the top in 2020 (and will likely proceed in 2021).

### 5.2. The impact of tweets on the common man's perspective

"We are what we read", having said this; it is clear how the perspective of common man can easily be influenced by Tweets or any social media influence. Having said this, people might get motivated to invest even more than their capacities without realizing the volatility of the market. This could result in people losing all their life savings and even more. To record how prominent this effect is, this study was carried out by segregating the tweets. The use of keywords like Tesla, rocket, Mars, starship, launch, bitcoin, cryptocurrency, hodl was accounted for segregating the tweets based on cryptocurrency. The charts found were baffling! (Fig. 5)

### 5.3. The dramatic evolution of bitcoin

The Supply of Bitcoin Is Limited to 21 Million. Figs. 6 and 7 depict the price evolution as well as the volume produced and trade offs from the year 2012 to 2021. As every bitcoin problem keeps getting solved every 10 min, the production volume has been increasing rapidly. As the data is extremely large (more than 4 million observations), RAPIDS distribution was used for this experiment. The volume of bitcoin generated with respect to time has been showcased. The price seems to be fluctuating on large scales as well. The bitcoin that started at \$0.0008 to \$0.08 per coin in July 2010 is now worth \$39,241.40. After making

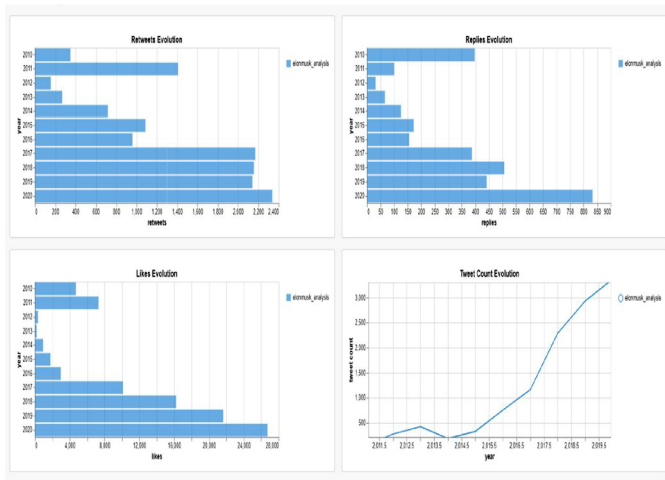


Fig. 5. Frequency of retweets, likes and replies on the tweets.

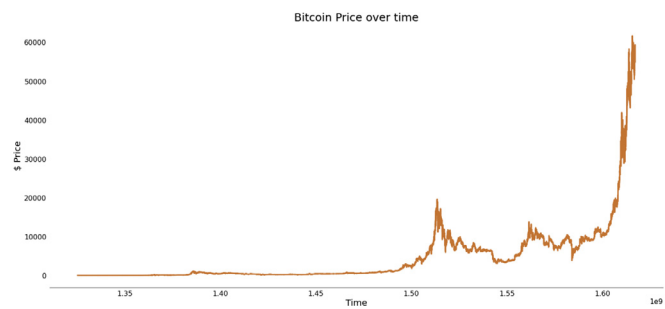


Fig. 6. Bitcoin price evolution.

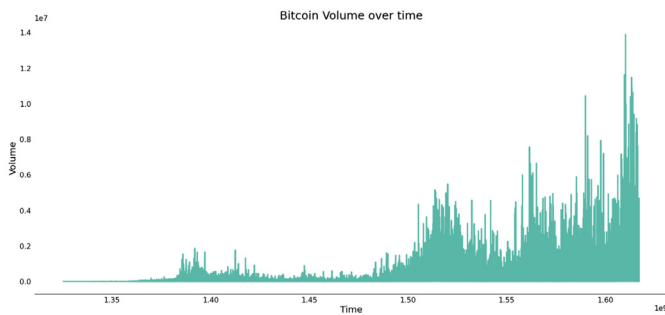


Fig. 7. Bitcoin volume evolution.

the data stationary enough to use, an ARIMA model was used to predict the bitcoin price.

#### 5.4. Predicting bitcoin price

Auto Regressive Integrated Moving Average with order (1,1,2) was used to predict the future values. The model has been trained on 0.99% of the bitcoins' financial data (after making it stationary). The testing was done on random sets of data to predict the next values of price details. Final accuracy of the model was 96% and average error of all the 10 predictions was about 0.0395. One of 10 samples tested is as follows:

- Actual value: 58696.198
- Predicted value : 58694.629
- Error: 0.002674

The actual vs. Predicted Views Forecasting has been showcased below in (Fig. 8).

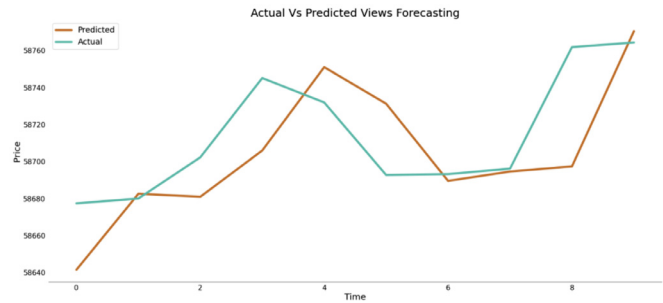


Fig. 8. Actual price vs. Predicted price of the bitcoin.

1. pay by bitcoin capability available outside us later this year
2. tesla is using only internal amp open source software amp operates bitcoin nodes directly bitcoin paid to tesla will be retained as bitcoin not converted to fiat currency
3. you can now buy a tesla with bitcoin

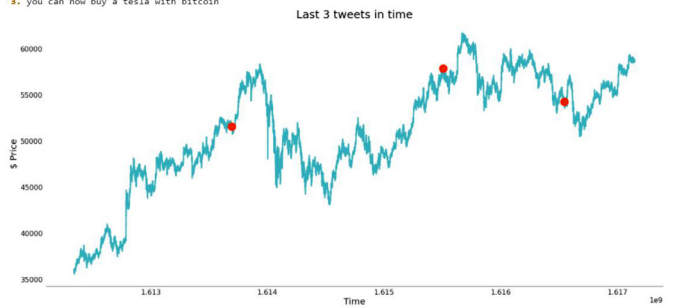


Fig. 9. Correlation of Musks' tweets and bitcoin pricing.

1. spacex is going to put a literal dogecoin on the literal moon
2. origin of doge day afternoon the ancient romans sacrificed a dogecoin at the beginning of the doge days to appease rage of sirius believing that the star was the cause of the hot sultry weather
3. 2022 dogecoin is dumb
4. if major dogecoin holders sell most of their coins it will get my full support too much concentration is the only issue imo
5. bought some dogecoin for lil x so he can be a toddler hodler
6. dogecoin is the people's crypto



Fig. 10. Correlation of Musks' tweets and dogecoin pricing.

## 6. Correlation

### 6.1. Bitcoin and Musk

Fig. 9 showcases the correlation between the tweets and pricing. The last three tweets were considered and time frames of the tweets were compared with the price of bitcoin for the next few days! It is evident that slight peaks were found for 1–4 days at a stretch right after the tweets were posted. And spiraled down with time based on the liquidity of the assets.

### 6.2. Dogecoin and Musk

Whereas, the same can't be stated for Dogecoin. Nothing drastic or worth noting can be found from the last 8 tweets, here. Thus, the claims can be termed irrelevant and illogical (Fig. 10).

## 7. Discussion

Elon Musk with his advent knowledge of memes and sarcastic jokes seems to connect with people on common grounds Urban (2015). The

% of tweets about Bitcoin: 0.111%  
 % of tweets about Dogecoin: 0.0796%

#### Most liked BITCOIN tweets:

1. you can now buy a tesla with bitcoin
2. btc bitcoin is an anagram of the boring company what a coincidence
3. bitcoin is my safe word
4. tesla is using only internal amp open source software amp operates bitcoin n retained as bitcoin not converted to fiat currency
5. pay by bitcoin capability available outside us later this year
6. bitcoin is almost as bs as fiat money
7. bitcoin is not my safe word

#### Most liked DOGECOIN tweets:

1. dogecoin is the people's crypto
2. bought some dogecoin for lil x so he can be a toddler hodler
3. spacex is going to put a literal dogecoin on the literal moon
4. if major dogecoin holders sell most of their coins it will get my full support issue imo
5. dogecoin rulz
6. origin of doge day afternoon the ancient romans sacrificed a dogecoin at the rage of sirius believing that the star was the cause of the hot sultry weather
7. dogecoin value may vary

Fig. 11. Top bitcoin and dogecoin tweets.



Fig. 12. SNL v/s dogecoin price.

retweets, likes and replies prove how popular his opinions have been. Words like “yes, yeah, good, thank, people, sure”, indicate how positive and enthusiastic his tweets usually are. Coming to cryptocurrencies; some of the famous tweets obtained have been displayed below to give an idea about the nature of the tweets Fig. 11 :

The model was trained on 0.99% of the Bitcoin data collected (after making it stationary). The Tweets were segregated and outliers (Tweets not specific to the study) were discarded. A comparison was made with the time series specific correlations and results were noted. We also predicted the values with the average error of about 0.0395; proving the research to be 96% accurate and market specific.

#### Is there correlation between bitcoin and Elon Musk's tweets?

In 2014, Elon initially referenced bitcoin to be ‘probably a good thing’. Soon enough the rumors that he could be “Satoshi Nakamoto” pseudonym of the Bitcoin founder) spread(which were then found to be untrue). The study showcased the impact of a single tweet was able to impact the market for a little while but eventually proved to be volatile and kept fluctuating.

SpaceX Vance (2015) claims to want to put dogecoin on the moon and tweets from Elon stating the same can be found in the dataset. Dogecoin, being a meme coin has been the talk of the tech industry and rumors of Doge going upto \$1 got pretty common in the mid of May-2021. With the SNL episode featuring Elon Musk, it was evident that Dogecoin (Fig. 12); if actually controlled by the popularity/influence of Elon, will reach a dollar at least. But none of the claims proved true. In fact, the price dipped and has been fluctuating ever since.

#### 8. Conclusion

Are Elon's tweets exclusively answerable for the crypto instability? The above account appears to paint an image that Elon has the ability to move the crypto markets with his tweets. Nonetheless, a more profound investigation of the working of crypto as a venture proposes something

#### BTC Price March - May 2021



Fig. 13. BTC price March–May 2021.

else. The value pattern of any resource can be divided into four stages - Accumulation, Markup, Distribution and Markdown. This normal cycle that any venture goes through is fundamental for its development and sustenance. Tesla recently stated that the use of Bitcoin has been suspended attributable to the high energy utilization of Bitcoin in the mining interactions. Cryptocurrencies went into a downward spiral with the news. This normal cycle that any venture goes through is essential for its development over the long haul.

By the end of 2020, crypto market went through the Accumulation stage, where numerous financial investors entered the market. Towards mid-February 2021, the cash increased and settled at an unsurpassed high of \$60,000. The Distribution and Markdown stage started when the RSI marker showed that the resource was overbought, gradually setting off a markdown. The sentimental value of the currency with the tweets by Elon indeed took a toll on him. Be that as it may, the fall was not unexpected in light of the fact that Elon was tweeting about it.

Besides, from the investigations above of the authentic cost of BTC, it is perfectly clear that there have been way more steep falls in the history of Cryptocurrencies Nofsinger (2005) even before Tesla made any declarations. This can be seen in (Fig. 13). Thus, it can clearly be stated that no one person can control the utter volatile world of cryptocurrencies!

#### Declaration of Competing Interest

Authors declare that they have no conflict of interest.

#### References

- Ante, L. (2021). How Elon Musk's twitter activity moves cryptocurrency markets. Available at SSRN 3778844.
- Aste, T., Tasca, P., & Di Matteo, T. (2017). Blockchain technologies: The foreseeable impact on society and industry. *Computer*, 50(9), 18–28.
- Baur, D. G., & Dimpfl, T. (2021). The volatility of bitcoin and its role as a medium of exchange and a store of value. *Empirical Economics*, 1, 1–21.
- Baur, D. G., Hong, K., & Lee, A. D. (2018). Bitcoin: Medium of exchange or speculative assets? *Journal of International Financial Markets, Institutions and Money*, 54, 177–189.
- Bongale, P., Tandon, C., Palivela, D. H., Nirmala, C. R. et al. (2021). Effect of the pandemic on students' learning habits in India.
- Cachin, C., et al. (2016). Architecture of the hyperledger blockchain fabric. *Workshop on distributed cryptocurrencies and consensus ledgers, Chicago, IL*: 310.
- Chauhan, T., Palivela, H., & Tiwari, S. (2021). Optimization and fine-tuning of densenet model for classification of COVID-19 cases in medical imaging. *International Journal of Information Management Data Insights*, 100000–100020.
- Chen, W., Xu, H., Jia, L., & Gao, Y. (2021). Machine learning model for bitcoin exchange rate prediction using economic and technology determinants. *International Journal of Forecasting*, 37(1), 28–43. [10.1016/j.ijforecast.2020](https://doi.org/10.1016/j.ijforecast.2020).
- Chohan, U. W. (2017a). Cryptocurrencies: A brief thematic review. SSRN.
- Chohan, U. W. (2017b). A history of dogecoin. Discussion series: Notes on the 21st century.
- Chu, J., Chan, S., Nadarajah, S., & Osterrieder, J. (2017). Garch modelling of cryptocurrencies. *Journal of Risk and Financial Management*, 10(4), 17.

- Cocco, L., Concas, G., & Marchesi, M. (2017). Using an artificial financial market for studying a cryptocurrency market. *Journal of Economic Interaction and Coordination*, 12(2), 345–365.
- Di Piero, M. (2017). What is the blockchain? *Computing in Science and Engineering*, 19(5), 92–95.
- Dinh, T. T. A., Liu, R., Zhang, M., Chen, G., Ooi, B. C., & Wang, J. (2018). Untangling blockchain: A data processing view of blockchain systems. *IEEE Transactions on Knowledge and Data Engineering*, 30(7), 1366–1385.
- Grover, P., Kar, A. K., & Janssen, M. (2019). Diffusion of blockchain technology. *Journal of Enterprise Information Management*, 32(5), 735–757.
- Halaburda, H. (2018). Blockchain revolution without the blockchain? *Communications of the ACM*, 61(7), 27–29.
- Jun, S.-P., Yoo, H. S., & Choi, S. (2018). Ten years of research change using google trends: From the perspective of big data utilizations and applications. *Technological Forecasting and Social Change*, 130, 69–87. [10.1016/j.techfore.2017.11.009](https://www.sciencedirect.com/science/article/pii/S0040162517315536). <https://www.sciencedirect.com/science/article/pii/S0040162517315536>
- Kar, A. K., & Navin, L. (2021). Diffusion of blockchain in insurance industry: An analysis through the review of academic and trade literature. *Telematics and Informatics*, 58, 101532. [10.1016/j.tele.2020.101532](https://www.sciencedirect.com/science/article/pii/S073658532030191X). <https://www.sciencedirect.com/science/article/pii/S073658532030191X>
- Kshetri, N. (2017). Can blockchain strengthen the internet of things? *IT Professional*, 19(4), 68–72.
- Kshetri, N., & Voas, J. (2018). Blockchain in developing countries. *IT Professional*, 20(2), 11–14.
- Mettler, M. (2016). Blockchain technology in healthcare: The revolution starts here. In *2016 IEEE 18th international conference on e-health networking, applications and services (healthcom)* (pp. 1–3). [10.1109/HealthCom.2016.7749510](https://doi.org/10.1109/HealthCom.2016.7749510).
- Mittal, A., Gupta, M. P., Chaturvedi, M., Chansarkar, S. R., & Gupta, S. (2021). Cybersecurity enhancement through blockchain training (CEBT)—A serious game approach. *International Journal of Information Management Data Insights*, 1(1), 100001.
- Nakamoto, S. (2009). Bitcoin: A peer-to-peer electronic cash system, may 2009.
- Nofsinger, J. R. (2005). Social mood and financial economics. *The Journal of Behavioral Finance*, 6(3), 144–160.
- Pilkington, M. (2016). Blockchain technology: Principles and applications. *Research handbook on digital transformations*. Edward Elgar Publishing.
- Saad, M., Choi, J., Nyang, D., Kim, J., & Mohaisen, A. (2020). Toward characterizing blockchain-based cryptocurrencies for highly accurate predictions. *IEEE Systems Journal*, 14(1), 321–332. [10.1109/JSYST.2019.2927707](https://doi.org/10.1109/JSYST.2019.2927707).
- Tandon, C., Bongale, P., Arpita, T. M., Sanjana, R. R., Palivela, H., & Nirmala, C. R. (2021). Use of natural language inference in optimizing reviews and providing insights to end consumers. In *2021 7th international conference on advanced computing and communication systems (ICACCS)*: 1 (pp. 60–65). [10.1109/ICACCS51430.2021.9442026](https://doi.org/10.1109/ICACCS51430.2021.9442026).
- Tasatanattakool, P., & Techapanupreeda, C. (2018). Blockchain: Challenges and applications. In *2018 international conference on information networking (ICOIN)* (pp. 473–475). IEEE.
- Urban, T. (2015). *Elon Musk: The world's raddest man. Wait But Why*.
- Vance, A. (2015). *Elon Musk: How the billionaire CEO of SpaceX and Tesla is shaping our future*. Random House.
- Wok, K. (2020). Advanced social media sentiment analysis for short-term cryptocurrency price prediction. *Expert Systems*, 37(2), e12493.
- Yli-Huumo, J., Ko, D., Choi, S., Park, S., & Smolander, K. (2016). Where is current research on blockchain technology?—A systematic review. *PLoS One*, 11(10), e0163477.