The Cryptokeeper: Elon Musk's Influence on the Crypto Market

Kelly Yang, Chris Meggs, and Anh Dao

University of Chicago

Author Note

Correspondence concerning this paper should be addressed to Kelly Yang (kyy@uchicago.edu), Chris Meggs (cmeggs0@chicagobooth.edu), and Anh Dao (adao0@chicagobooth.edu).

Abstract

With the rise of cryptocurrencies built upon the implementation of blockchain technology, we have ultimately transformed the modern understanding of currency, ledgers, and financial systems. In particular, the need for a decentralized global currency that is not tied to any governing entity has presented itself as an attractive way for people to conduct transactions, store value, and even generate wealth. Fueled by the global Covid-19 pandemic and the subsequent inflationary pressures due to stimulus spending, cryptocurrencies have seen increased mainstream and institutional adoption driving continual, volatile, and rapid market expansion. This meteoric rise corresponds closely with increased social media discussion and positive sentiment regarding the burgeoning technology. This study will be conducted by examining past market history and trying to understand the relationship between social patterns and price movements. We will build upon existing research around the use of machine learning and Bayesian regression on Twitter sentiment to improve price prediction models. It will do so by focusing on the impact that a particular influential voice (Elon Musk) may have in the crypto market.

The Cryptokeeper: Elon Musk's Influence on the Crypto Market

Section 1. Introduction

1.1 Sentiment Analysis and Twitter

Sentiment analysis of social media has long been used as a proxy for public sentiment. In short, sentiment analysis uses natural language processing to analyze words to vectorize them into positive, negative, and neutral scoring (at its simplest form, it is possible to do multidimensional vectoring regarding sentiment). It has been utilized in a range of functions from marketing, voice of customer research, political sentiment regarding candidates, to economic and market trends[8]. This technique has been particularly powerful with the rise of "microblogging" platform Twitter, where public users post seemingly instantaneous reactions to the events and happenings in the world around them. Twitter provides a powerful source of data for extrapolating the sentiment of the public.

Bollen et. al. found that short term news had immediate and measurable effects on public sentiment, whereas long-term economic indicators had more gradual and cumulative effects on public sentiment in their landmark paper, "Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena"[2]. Since then, there has been a large body of research tying public sentiment to particular markets and exploring whether that relationship was causal and could provide predictive value to forecasting models. Indeed when Pagolu et. al. found strong correlation in Twitter sentiment with stock performance the following day [13].

1.2 Crypto Markets and Price Prediction

With the exponential growth of the overall crypto market in 2017 (the overall market cap increased from $18.5 billion to $696 billion), far more attention has been paid to crypto markets by the general public and researchers alike. Cryptocurrencies have emerged as a viable alternative to conduct transactions, act as an immutable ledger, and store and grow value when compared to traditional asset classes. During the Covid-19 pandemic, there has been a similar explosion in the amount of investment into this evolving market. The total market cap currently sits at roughly 2,335 billion, down from its all-time high of 2,948 billion, but still far above the $200 billion value at the outset of the pandemic. This meteoric rise corresponds closely with increased social media discussion and positive sentiment regarding the burgeoning technology.

As noted above, this explosion has similarly been present in the academic literature surrounding cryptocurrency. Guegan and Renault explored StockTwits (a social media platform for investors) as a novel proxy for investor attention, extrapolating investor sentiment from the tone of the messages, and how they play a role in the Bitcoin market. They found that there was a correlation between investor attention and pricing, but only for a period of up to 15 minutes [10]. Li et. al. conducted a similar study, but looking at realized volatility as well as pricing, and using Google trend data as a proxy for investor attention. They not only established significant bi-directional casualties between cryptocurrencies and online attention, but also that a combination of investor attention proxies (social media and search engine intensity) proved more causal than one metric alone [11]. While Abraham et. al. would argue that Tweet volume is a more significant factor than Tweet sentiment, there seems to be high levels of evidence that sentiment analysis is useful in improving cryptocurrency pricing models [1]. For more information regarding the current state of the academic literature, Fang et. al. have provided an excellent survey paper, "Cryptocurrency Trading: A Comprehensive Survey" [7].

1.3 Bitcoin, Dogecoin and Elon Musk

By now Bitcoin (BTC) has become a household name as the most established and longest running cryptocurrency, with a current value of just north of $50,000 and a market cap of $930 billion. BTC accounts for roughly 40% of the overall cryptocurrency market. Dogecoin (DOGE) is a relative upstart in the evolving crypto ecosystem; it is currently valued at $0.176 with a market cap of $23.5 billion, accounting for around 1% of the market, and falling just outside of the top 10 cryptocurrencies in terms of market cap. This difference in spot price to market cap ratio is due in large part to the much larger, and ever increasing, supply of DOGE relative to BTC (132 billion to 18.8 million)[4].

Cryptocurrencies have acted as currencies to some, investments to others, long term stores of value (in spite of their volatility). What makes DOGE inherently different from BTC and most major cryptocurrencies is that it is what is referred to in the crypto community as a "Meme coin," that is a coin that was never intended to be used in the aforementioned manners. Instead they are coins that are created as part of an internet joke. What makes DOGE even more remarkable is that a year ago, it was relatively worthless and trading at fractions of a cent. It underwent a massive surge, increasing in value over 12,000% in 2021, peaking at $0.7442 with a market cap of around $90 billion. This massive surge and crash has been attributed in large part to Elon Musk, the self-proclaimed "Dogefather," and his introduction of DOGE to the masses, and the ensuing social media storm surrounding his proclamations[6]. We will look to understand the measurable impact that Elon had on DOGE and BTC pricing, and whether the influence of a singular individual can create irrational behavior in a seemingly rational market[7].

Section 2. Data & Methodology

2.1 Data & Processing

For our analysis, we obtained Elon Musk's tweets ("Elon Musk Tweets (2010 - 2021)") and historical prices of DOGE and BTC ("G-Research Crypto Forecasting") from Kaggle. The tweets datasets consist of approximately 45,000 tweets related to DOGE and BTC by Elon Musk himself between 2010 and 2021. For each tweet, we have information on the text of the tweet, the time it was posted, the conversation it belonged to, the language it was in, and whether it contains photos or videos. The DOGE and BTC pricing data includes minutely prices over 2011 - 2021, which was then filtered from 2019 - 2021, including open, high, low and close prices. Volume and percent change from the previous day was also provided.

Due to time constraints, we chose to focus on Elon Musk's tweets since he's one of the most prolific and influential Twitter celebrities when it comes to cryptocurrency and specifically DOGE and BTC. Our model can be easily extended to cover other media personalities, crypto assets or social media platforms due to the similar nature of available information. To account for the social buzz around DOGE and BTC outside of Twitter, we supplemented these data with monthly search trends from Google (Google Trends).

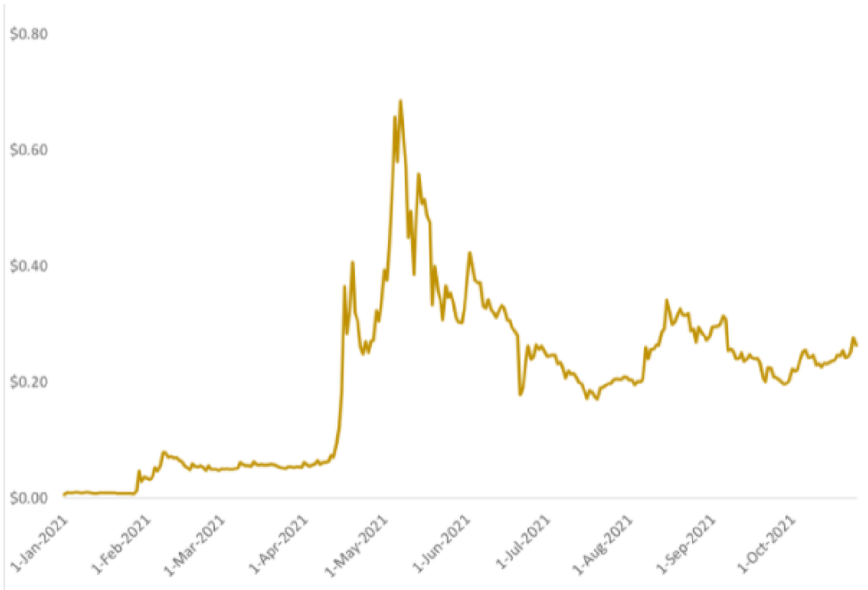
To prepare our datasets, we first restricted DOGE's and BTC's prices to the period from January 1, 2019 to October 25, 2021 since both cryptocurrencies experienced the most price movements during this time (DOGE's price was close to zero before then, while BTC's just bottomed at $3,800 after the height of $19,700 in 2018). We then computed percent change and price change based on close and open, and used the percent price change in our model Afterwards, we limited Elon's tweets to those mentioning DOGE and BTC over that time span. His first tweet was on April 2, 2019 while his last tweet was on March 18, 2021. Finally, we gathered Google Trends data for the search term "Elon doge" of the months that Elon's tweets were published, which ran from April 2019 to March 2021.

2.2 Summary Statistics

2.2.1 DOGE

To get a general sense of prices, we plot the price series for DOGE since the start of 2021 when it began to gather momentum and gained national recognition (DOGE's price remained under 1 cent throughout 2019 and 2020).

Daily Closing Price of DOGE from 1/1/2021 to 10/25/2021



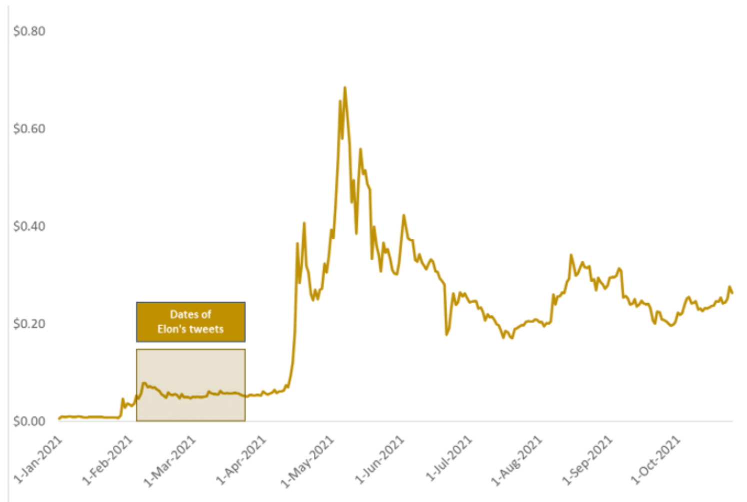
DOGE's price began to pick up in February, rising from under 1 cent in January to 7 cent at the start of February. However, it remained relatively flat until mid-April when it started a huge ascent to the year's maximum of 68 cent around mid-May. Then, it went through a rough correction path to 18 cent in mid-June, before slowly recovering to 26 cent by the end of October.

Word Cloud of Elon's Tweets on DOGE in 2021

When looking at the words that Elon used when talking about DOGE, we see that the most viral ones are also those that are more playful and often associated with the meme-ing content that usually flood the Twitter-verse. Each word is then weighted by the virality of the tweet where it appeared in, which is defined as the number of likes, replies, and retweets that that tweet gets. One could think that the more serious tweets should have a bigger influence on price movements, but it might not be the case given the fact that a large portion of early DOGE adopters are meme-savvy or born in the age of memes and thus would react more strongly towards this type of content.

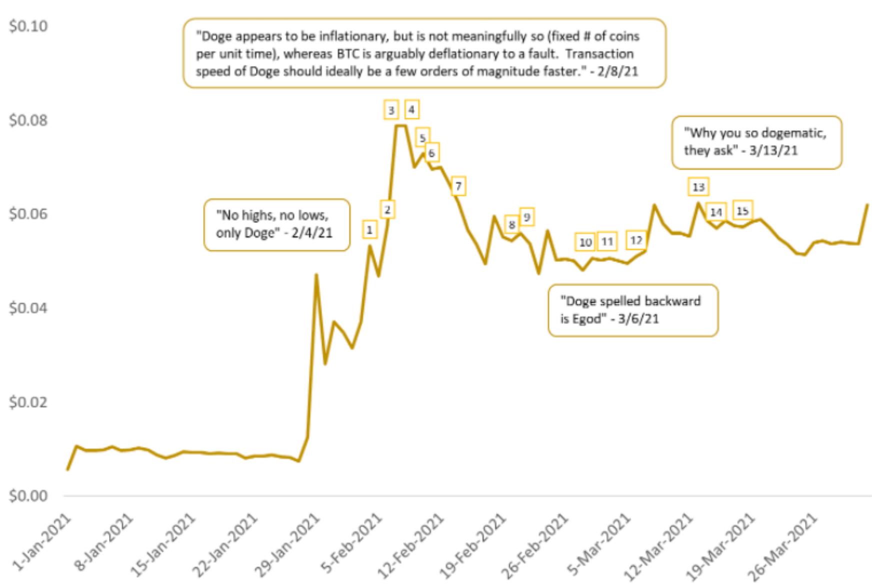
Musk also had some more traditional discussion on his account such as DOGE's relationship with inflation, its comparison to BTC, his opinion on whether DOGE is fairly valued or his actual trading activity, even though they do not get as much interactions:

Elon's Tweets Had Limited Impact on DOGE's Prices in 2021



In 2021, Elon first tweeted about DOGE on February 4 and made his last remark on March 18. Relative to its 2021 evolution, DOGE's price remained quite flat between these dates. This is a time when DOGE hadn't garnered national attention and even though Musk was a vocal supporter on Twitter, the hype he created didn't move the needle all that much especially compared to what happened during the summer. Interestingly, he chose to remain silent for the rest of 2021 when the most trading activities occurred so we won't be able to see if his impact would have been larger when DOGE is in the national spotlight.

Elon's Tweets Had Most Impact on DOGE's Prices in 2021Q1



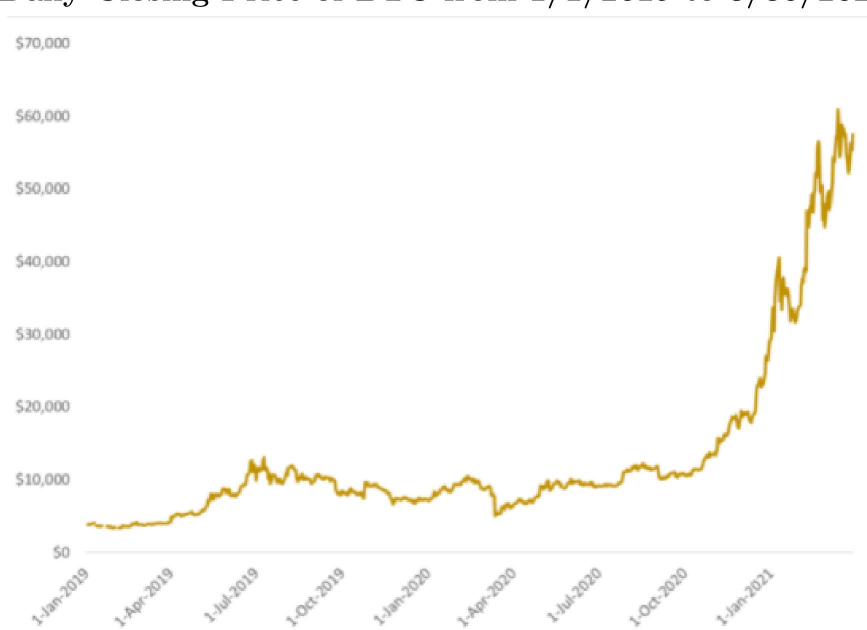
If we zero in on the time that Elon tweeted, i.e. the first quarter of 2021, we can see more pronounced price movements after he made a statement even though the impact can sometimes be ambiguous:

- On February 4, he voiced his support for Doge and its price picked up from around 5 cent to around 8 cent in four days.
- On February 8, he admitted that DOGE has an inflationary supply (i.e. its value will decrease automatically over time by design), which is followed by a swift correction of DOGE's price back to around 5 cent.
- On March 6 and March 13, he tweeted his support for DOGE in a more amusing manner and DOGE's price moved in either direction afterwards (increased from 5 cent to around 6.5 cent, before dropping back to just under 6 cent.

2.2.1 BTC

Below we plot the price series for BTC since the start of 2019 to get a sense of its meteoric rise after falling off from its peak in 2018.

Daily Closing Price of BTC from 1/1/2019 to 3/30/2021



BTC's price began to pick up in April 2019, rising from around \$4,000 to \$13,000 at the start of July 2019. It actually remained relatively flat for more than a year until October 2020 when it started its longest and steepest ascent to date, reaching one high after the other and haven't yet showed signs of slowing down. BTC's price reached \$39,000 in early January 2021, \$56,500 in mid-February 2021 and \$60,000 by the time March 2021 rolled around.

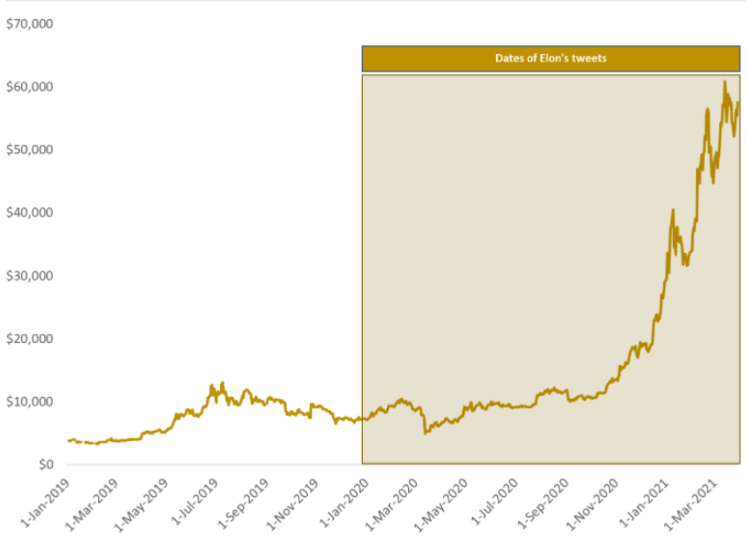
Word Cloud of Elon's Tweets on BTC in 2021

When looking at the words that Elon Musk used to talk about BTC, we see that the most viral ones focus around topics such BTC's role as a virtual alternative to fiat money and how its importance as a keeper of value has increased given the negative real interest rates that central banks have been maintaining to pump up the economy. Again, each word is then weighted by the virality of the tweet where it appeared in, which is defined as the number of likes, replies, and retweets that that tweet gets. He justified his decision for Tesla to accept payments in BTC as a way to counter the deflationary pressures during the pandemic:

Elon Musk was also active in engaging with other Twitter celebrities when BTC gained mainstream traction and they started to learn more about it. He threw his weight behind the cryptocurrency, explaining the reasoning behind its increase to JK Rowling and encouraging Maisie Williams to hold in for the long term:

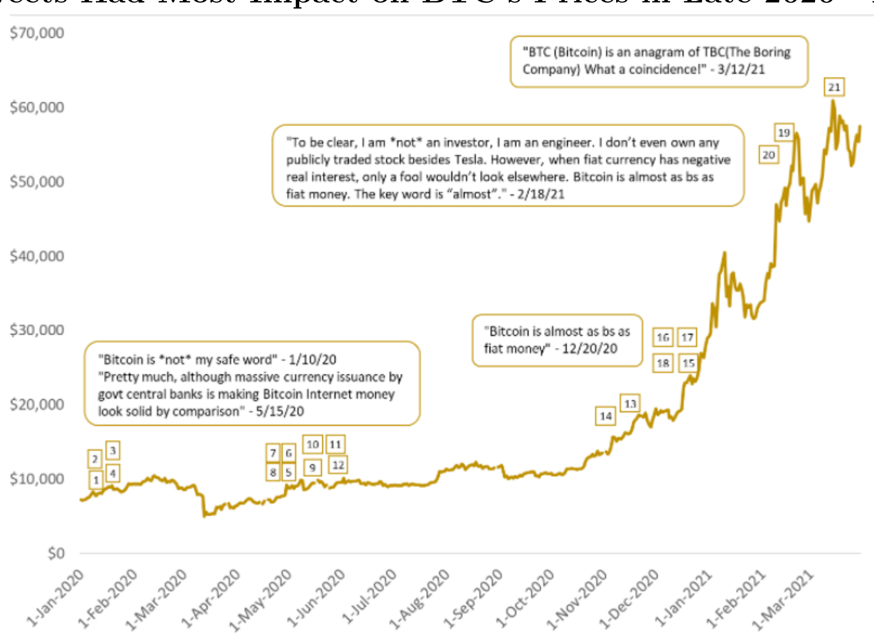
When indeed he only holds a minuscule amount of BTC:

Elon's Tweets Possibly Had Significant Impact on BTC's Prices in 2020-2021



In our data, Elon first tweeted about BTC on Januaray 10, 2020 and made his last remark on March 12, 2021. This timeframe coincided with the hottest streak of BTC prices in its history, so it's interesting to see if Elon Musk's opinions swayed the public perception of the cryptocurrency in one way or another. Given his vocal support of BTC in general and especially his decision to allow Tesla to accept BTC payments, it's very likely that those actions gave BTC the legitimacy and confidence that it needed to get off the ground.

Elon's Tweets Had Most Impact on BTC's Prices in Late 2020 - Early 2021



If we zero in on the times that Elon tweeted about BTC since 2020, we can see some significant price movements after he made a statement even though the impact can sometimes be ambiguous:

- On January 10, 2020, he tweeted out a cryptic message about BTC not being his safe word. We don't know what he meant but at that time it was a big deal if any influential businessman mentioned BTC at all. BTC's price remained flat after, however.
- On May 15, 2020, he made his support for BTC much clearer through explaining to JK Rowling why BTC had been gaining in popularity. BTC's price still remained flat afterwards.
- On December 20, 2020, he once again threw his weight behind the currency amidst doubts about its legitimacy, claiming that BTC is "almost as bs as fiat money." This seemed to spark the meteoric raise all the way to March 2021.
- On February 18, 2021, he explained his reasoning for allowing Tesla to accept BTC payments and reiterated his preference of BTC over fiat money. This is around the peak of BTC price, and surely it helped reinforce BTC's position.

2.3 Modeling

2.3.1 Approach

Our goal is to predict daily prices of DOGE and BTC through a combination of time series and social media variables. The dependent variable is percent change in daily closing prices from the previous day. Time series variables include daily prices (open, high, low, close) and volume. Social media variables include sentiment scores (explained in detail below) as well as the virality measures (number of likes, replies and retweets) for each of Elon's tweets. Finally, we also added interest scores from Google Trends to account for platforms outside of Twitter. For a given time period, a search term is scored on a scale of

0 and 100 - with 100 being the time when it's most popular (so a score of 50 means the term is half as popular) and 0 being there's not enough data to score that term.

2.3.2 Model Selection

We wanted to find a reliable way to predict price changes in blockchain assets based on Elon's tweets and general interest in Elon relating to the asset. Doge was used for our model because Elon tweets about this more frequently than other cryptocurrencies. Features that we wanted to use were a number of daily google searches containing both 'Elon' and 'Doge', whether or not Elon tweeted about Doge that day, the subjectivity of his tweet, the polarity of the tweet, and the volume of Doge trading. These variables are all continuous, with the exception of whether or not Elon tweeted about Doge that day, which is binary. The outcome we want to predict is the price change.

A number of models were considered. This included linear regression, logistic regression, decision trees, linear SVC, and random forests. Initially, logistic regression was heavily considered. However, in terms of predicting the outcome, we would prefer to predict price changes as a percentage variable, which would be continuous. Logistic regression would need a binary outcome variable, such as whether or not the price went up or down. This did not feel as explanatory as predicting the price change as a percent. For this reason, we ultimately decided not to go with a logistic regression. Linear SVC was considered as a way to build a model that could classify the data. However, somewhat similar to logistic regression, linear SVC would require the outcome to be a categorical variable. Decision trees and random forests were also considered, but since too many of our features were continuous, this turned out to be difficult. Therefore, a multivariate linear regression seemed like the best choice given the classification of our features, scope of our project, and the intent of our predicted model.

2.3.2 Sentiment Analysis

Two of our features consisted of trying to decipher the sentiment behind Elon's tweets, and how that affected Doge prices. Within natural language processing, there are many different ways to conduct a sentiment analysis on given text. Due to time constraints and lack of computing power, we decided to use a popular Python Library called textblob that has a built-in functionality and API for conducting sentiment analysis. Textblob is reliable and provides a consistent API that allows us to dive into common natural language processing.

Using textblob, we could run the library to process our tweets in text format. Two measures were conducted on each of Elon's tweets relating to Doge. The first consisted of a sentiment score, which would return a subjectivity score, which would return a float from 0 to 1, with 0 being the highest level of objectivity, and 1 being the highest level of subjectivity. The second measure conducted was to return the polarity score, which calculated a float from -1 to 1. The more negative the polarity, the more negative the sentiment, and the more positive the polarity, the more positive the sentiment.

One important thing to note is that this sentiment analysis is unable to detect hints of sarcasm in speech, despite humans being able to pick up on satirical statements. For instance, Elon semi-frequently tweets sardonic comments when Doge and other cryptos go up, such as "I only sell Doge!". This is a sarcastic response to the price increase and likely directed to people who denounce cryptocurrencies. However, the sentiment analysis conducted can not pick this up, and gives this a negative polarity score corresponding to a negative sentiment, when it really correlates with price increase. Additionally, Elon often tweets in response to large shifts in prices, so when using this feature, it is important to remember that his tweets could very well be a response to price shifts and not the other way around. These biases need to be taken into careful consideration when using a similar model of trying to utilize sentiment analysis

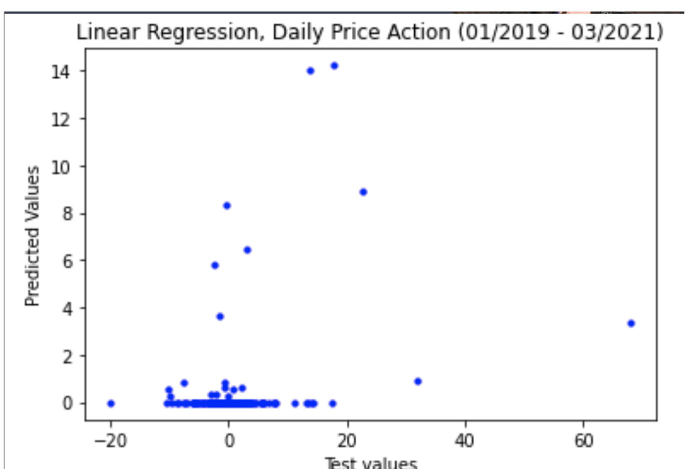
2.3.3 Multivariate Linear Regression

As stated above, we decided to use a multivariate linear regression model due to the scope of our project, the limitations, and the continuous percent change variable we are trying to predict. We decided to create two models within the context of the dogecoin blockchain asset. This is because Elon does not tweet about Doge every day. Therefore, one dataset used to create the model will utilize daily prices with two features: google searches of 'elon doge' and whether or not Elon tweeted that day. The second dataset to create the second model uses prices only on days Elon tweeted, with four features: google searches of 'elon doge', subjectivity score of his tweet, polarity score of his tweet, and the daily trade volume.

Both of these models are then used to try and predict percent changes in price on a daily scale with the respective features. For our models, we will assume that there is some form of linear relationship between the features and the model.

Section 3. Results & Discussion

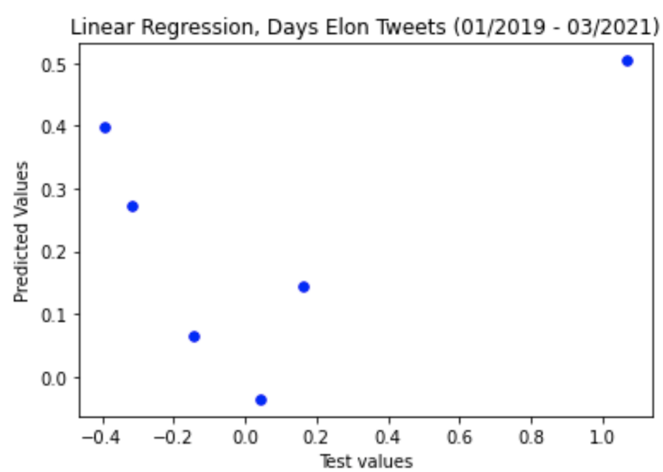
The first figure below shows the model that was run on daily price data with the full dataset, each day from the start of 2019 to the end of March, 2021.




```
The MAE is 2.957795954755966
The MSE is 37.373165657730546
The R^2 value is 0.10141930004259025
Feature: elon doge: (United States), Score: 1.23457
Feature: elon_tweet, Score: 0.93189
```

Although the data is relatively scattered, we can observe a slightly linear relationship. Due to the inherently subjective nature of determining human nature around a volatile blockchain asset in Dogecoin, as well as our limited datasets, our model performs relatively okay. The MAE represents the mean-average-error, which represents the average absolute error between predicted and actual values when the model is fitted to the test set. The MSE is is the average of squared errors and the R^2 value represents goodness of fit of the line. With this model, we were able to compute the above feature coefficients. These coefficients represent the degrees of change correlated with our features and our outcome. The interpretation of this would be that a one-unit increase in google searches of ‘elon’ and ‘doge’ results in an approximate 1.2% price increase in dogecoin that day. Similarly, if elon tweeted that day, we would expect an approximate 0.9% price increase in dogecoin that day.

The second figure below shows the model that was run on days only that Elon tweeted about Doge. It is important to note that this dataset is much smaller, so a overfitting could possibly be something that might cause error.




```
The MAE is 0.3746579731515183
The MSE is 0.22331101192197467
The R^2 value is 0.04979824500723029
Feature: elon doge: (United States), Score: 0.00177
Feature: subjectivity, Score: 0.07460
Feature: polarity, Score: 0.19417
Feature: Volume, Score: 0.00000
```

Here we can observe a more linear relationship, with low error scores in both MAE and MSE. However, the R^2 value is lower than the other model, indicating that there is less of a goodness of fit but likely some merit to our predictions. As mentioned before, due to the much smaller dataset, overfitting must be taken into consideration. Based on our model, we can see that all our features have a positive correlation with Dogecoin prices, except for Volume which has no correlation at all. The correlation for google searches of 'elon' and 'doge' is extremely low and negligible, but we can see a 0.194 coefficient for polarity score. This indicates that positive sentiments of Elon's tweets surrounding dodge lead to higher percent increases on average.

When considering these models, we must remember that this is trying to predict human behavior, which is inherently very subjective and difficult to measure. Therefore, we will likely see higher error values when using the model to make predictions. Additionally, since this is confined solely to Elon and Dogecoin, these models would likely not be able to be used on any other blockchain asset, although the same methodology can be used to achieve different results. Finally, we want to consider all the possible biases when trying to utilize this technology. Our project scope was very small, but on a larger scale it would be important to evaluate sentiment around Elon's tweets and other users as well. And take into account that Elon's tweets might be reactionary to price changes, rather than the catalyst. Specific price details by the minute or hour would be needed to gauge these changes.

Section 4. Areas for Further Research

We would have liked to include more data sources, such as the surrounding references to Elon and doge on twitter by other users, but given the time intensive nature of webscraping through Twitter's API, it was not feasible given the time constraints of this study. We also do think that our results are indicative of user influence. Adding feature based on user influence utilizing follower count, retweets, and general interactions could improve on models built by Shah and others to improve pricing prediction performance.

Finally, it would be great to see how these data sets perform in different models, be that random forests, neural networks, etc.

References

[1] Abraham, Jethin, et al. "Cryptocurrency Price Prediction Using Tweet Volumes and Sentiment Analysis." SMU Data Science Review, vol. 1, no. 3, ser. 1, 25 Aug. 2018. 1, https://doi.org/https://scholar.smu.edu/cgi/viewcontent.cgi?article=1039amp; context=datasciencereview.

[2] Bollen, Johan, et al. "Modeling Public Mood and Emotion: Twitter Sentiment and Socio-Economic Phenomena." Proceedings of the International AAAI Conference on Web and Social Media, vol. 5, no. 1, 2011, https://doi.org/https://arxiv.org/abs/0911.1583. Accessed 13 Nov. 2021.

[3] Colianni, Stuart, et al. "Algorithmic Trading of Cryptocurrency Based on Twitter Sentiment Analysis." Stanford CS, 2015, http://cs229.stanford.edu/proj2015/029$_r eport.pdf.Accessed 4Nov.$2021.

[4] "Cryptocurrency Prices, Charts and Market Capitalizations." CoinMarketCap, https://coinmarketcap.com/.

[5] Dalila, Ayhm. "Elon Musk Tweets (2010 - 2021)." Kaggle, 20 June 2021, https://www.kaggle.com/ayhmrba/elon-musk-tweets-2010-2021.

[6] "Dogecoin: Stage 3 of a Massive Bubble (Cryptocurrency:Doge-USD)." Seeking Alpha, 19 May 2021, https://seekingalpha.com/article/4430039-dogecoin-stage-three-of-a-massive-bubble.

[7] Fang, Fan, et al. ArXiv, vol. 2003, no. 11352, ser. 4, 25 Oct. 2021. ArXiv, https://doi.org/https://arxiv.org/pdf/2003.11352.pdf.

[8] Gentzkow, Matthew, et al. "Text as Data." Journal of Economic Literature, vol. 57, no. 3, 2019, pp. 535–574., https://doi.org/10.1257/jel.20181020.

[9] Google Trends, Google, https://trends.google.com/trends/?geo=US.

[10] Guégan, Dominique, and Thomas Renault. “Does Investor Sentiment on Social Media Provide Robust Information for Bitcoin Returns Predictability?” Finance Research Letters, vol. 38, 2021, p. 101494., https://doi.org/10.1016/j.frl.2020.101494.

[11] Li, Yue, et al. “Comparing Search-Engine and Social-Media Attentions in Finance Research: Evidence from

[12] Cryptocurrencies.” International Review of Economics amp; Finance, vol. 75, 2021, pp. 723–746., https://doi.org/10.1016/j.iref.2021.05.003.

[13] Pagolu, Venkata Sasank, et al. “Sentiment Analysis of Twitter Data for Predicting Stock Market Movements.” 2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES), 2016, https://doi.org/10.1109/scopes.2016.7955659.

[14] “Research Crypto Forecasting: Kaggle.” G, https://www.kaggle.com/c/g-research-crypto-forecasting/overview.

[15] Schnaubelt, Matthias. “Deep Reinforcement Learning for the Optimal Placement of Cryptocurrency Limit Orders.” European Journal of Operational Research, vol. 296, no. 3, 2022, pp. 993–1006., https://doi.org/10.1016/j.ejor.2021.04.050.

[16] Shah, Devavrat, and Kang Zhang. “Bayesian Regression and Bitcoin.” ArXiv, vol. 1410, no. 1231, ser. 1, 6 Oct. 2014. 1, https://doi.org/https://arxiv.org/abs/1410.1231.

[17] Zhu, Panpan, et al. “Investor Attention and Cryptocurrency: Evidence from the Bitcoin Market.” PLOS ONE, vol. 16, no. 2, 2021, https://doi.org/10.1371/journal.pone.0246331.