# Bitcoin Sentiment Analysis and Market Predictions

**Doug Marcum**

**March 6, 2021**

## Abstract

Ponzi scheme? The currency of the future? A hedge against the stock market? After a decade since the first decentralized digital currency, Bitcoin, began circulating, defining what cryptocurrencies are is still debatable. What can no longer be debated is their impact on all markets today. With the current global cryptocurrency market capitalization being approximately $1.45 trillion[1], multinational corporations (Tesla, MicroStrategy, Mass Mutual, Square) and hedge funds dominate the surge. These internet coins are no longer a novelty and thus should be reviewed in the same manner as one would view Amazon's stock or the US Dollar. As with those assets, determining the value and knowing when to buy or sell is extremely important. Traders have utilized statistics and machine learning to determine optimizations, so incorporating those systems to process cryptocurrencies is necessary. This project looks to review Twitter data to assess whether sentiment analysis can illustrate a correlation in crypto market swings caused by influencers. Second, analyzing historical information, in conjunction with the Twitter data if necessary, the effectiveness of predictive models optimizing one's gains and minizine losses will be reviewed.

## Data

For this project, obtaining the necessary data was not an issue. The data utilized for sentiment analysis was obtained from Twitter via the Tweepy library. Tweepy provides access to the entire Twitter RESTful API and returns a Tweepy model class instance. The Twitter API wrapper, connected to the search_full_archive method, allows for a search of tweets beginning in March 2006. Keywords and influencers supplied for the query can be found in Appendix A. Only tweets posted from January 1, 2020, to February 15, 2021, were scraped from Twitter. The Twitter data provides a robust dataset with layers of metadata and 5473 tweets for review. More information about Tweepy can be found at https://docs.tweepy.org.

For Bitcoin price analysis data, this information was available for download from Yahoo Finance. Financial information from January 1, 2016, to February 15, 2021, was downloaded. This information can be accessed here: https://finance.yahoo.com/quote/BTC-USD/history?p=BTC-USD.

## Sentiment Analysis

Sentiment analysis is a technique through which you can analyze a piece of text to determine its sentiment by combining machine learning and natural language processing (NLP). A program can decide if the text's sentiment is positive, negative, or neutral using fundamental sentiment analysis.[2] With the explosion of social media, the ability to share one's opinions and thoughts have essentially established a medium that is more than mere information sharing but rather a platform for expression. Social media has created a means for individuals and brands to interact and influence ideas and directions quickly. With hundreds of platforms and thousands voicing opinions every second, the ability to differentiate the positive from the opposing views is vital, thus the need for sentiment analysis.

As previously stated, the data was obtained via Twitter[3] for the sentiment analysis review. Initially, the idea of exploring daily sentiment related to Bitcoin and cryptocurrency was considered. However, research illustrated that while hundreds of thousands of tweets are being shared on the platform related to the topic, many are

---

[1] CoinMarketCap. (2021, February 14). Global Cryptocurrency Charts. https://coinmarketcap.com/charts/.
[2] Shivanandhan, M. (2020, September 30). What is Sentiment Analysis? A Complete Guide for Beginners. https://www.freecodecamp.org/news/what-is-sentiment-analysis-a-complete-guide-to-for-beginners/.
[3] Steps on how the Twitter data was obtained can be found here: https://github.com/MarcumDoug/Bitcoin_Sentiment_Analysis_Market_Predictions/blob/main/EDA_Code/Twitter_Data.ipynb.

retweets or filled with little value. Understanding groupthink is necessary, yet analyzing the influence of powerful voices and Bitcoin influencers was determined as a better avenue of exploration.

For purposes of this exercise, twenty Bitcoin influencers were selected for Twitter sentiment analysis. Little regard towards the number of followers or number of tweets was considered when determining the influencers, but instead, they were set on perceived market influence. This was established by reviewing chat channels and posts illustrating their impact from a pro or con standing relating to Bitcoin. The influencers selected are listed in Appendix A.

Data preparation consisted of removing unnecessary items (e.g., RT, @, HTTPS) that can impact interpreting words for processing. Additionally, the period for analysis was reduced to approximately the previous six months (8/1/2020 to 2/15/2021). This was done to balance the distribution of tweets from highly engaged influencers with those with a lower engagement. The final tweet tallies are shown below and in Appendix B.

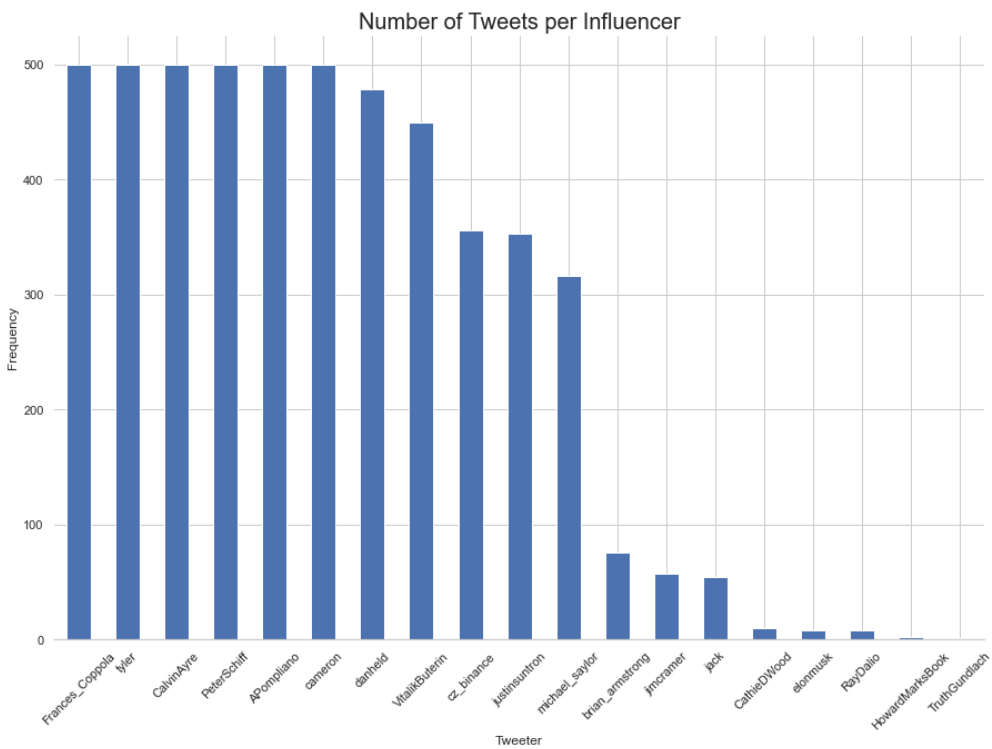



Figure 1: Tweet Frequency by Influencer

Sentiment analysis, at this time, will not deliver results with 100% accuracy, as research has shown that humans only agree 79% of the time on the sentiment.[4] To provide balance in analysis, it was determined to analyze the results of two sentiment analysis tools, VADER and TextBlob.

VADER[5] (Valence Aware Dictionary and sEntiment Reasoner) is a lexicon and rule-based sentiment analysis tool tuned explicitly to sentiments expressed in social media. VADER uses a combination of A sentiment lexicon is a list of lexical features (e.g., words) which are generally labeled according to their semantic orientation as either positive or negative. VADER not only determines the Positivity and Negativity score, but it also illustrates the

[4] Ognieva, M. (2010, April 19). How Companies Can Use Sentiment Analysis to Improve Their Business. https://mashable.com/2010/04/19/sentiment-analysis/.
[5] Additional information relating to VADER can be found here: https://github.com/cjhutto/vaderSentiment.

degree a sentiment is positive or negative.[6] VADER is also able to translate utf-8 encoded emojis, an essential tool. An example from the tweets dataset is below.

| | User | Tweet | Date | Positive | Negative | Neutral | Compound | Rating |
|---|---|---|---|---|---|---|---|---|
| 0 | elonmusk | Any crypto wallet that won't give you your pr... | 2021-02-10 | 0.000 | 0.138 | 0.862 | -0.3400 | Negative |
| 1 | elonmusk | Doge appears to be inflationary, but is not m... | 2021-02-08 | 0.000 | 0.000 | 1.000 | 0.0000 | Neutral |
| 2 | elonmusk | It's the most fun crypto! | 2021-02-07 | 0.493 | 0.000 | 0.507 | 0.5974 | Positive |
| 3 | elonmusk | Dogecoin is the people's crypto | 2021-02-04 | 0.000 | 0.000 | 1.000 | 0.0000 | Neutral |
| 4 | elonmusk | Bitcoin is almost as bs as fiat money | 2020-12-20 | 0.000 | 0.000 | 1.000 | 0.0000 | Neutral |

Figure 2: VADER analysis of Elon Musk tweets

TextBlob is a Python library for processing textual data. It provides a simple API for diving into standard natural language processing (NLP) tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation, and more.[7] The sentiment function of TextBlob returns two properties, polarity and subjectivity. Polarity is a float that lies in the range of [-1,1], where 1 means a positive statement and -1 means a negative statement. Subjective sentences generally refer to personal opinion, emotion, or judgment, whereas objective refers to factual information. Subjectivity is also a float in the range of [0,1]. An example from the tweets dataset is below.

| | User | Tweet | Date | Score | Rating |
|---|---|---|---|---|---|
| 0 | elonmusk | Any crypto wallet that won't give you your pr... | 2021-02-10 | 0.0000 | Neutral |
| 1 | elonmusk | Doge appears to be inflationary, but is not m... | 2021-02-08 | -0.0500 | Negative |
| 2 | elonmusk | It's the most fun crypto! | 2021-02-07 | 0.4375 | Positive |
| 3 | elonmusk | Dogecoin is the people's crypto | 2021-02-04 | 0.0000 | Neutral |
| 4 | elonmusk | Bitcoin is almost as bs as fiat money | 2020-12-20 | 0.0000 | Neutral |

Figure 3: TextBlob analysis of Elon Musk tweets

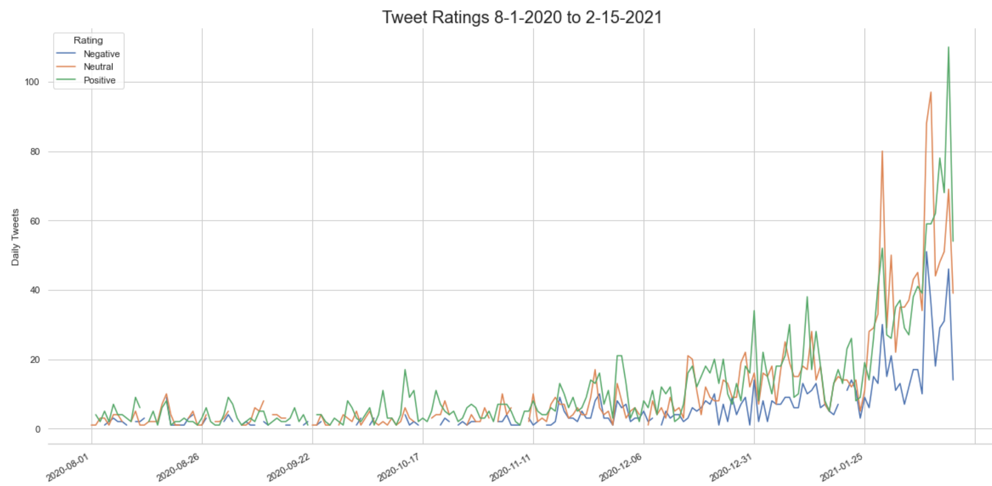
After being processed via VADER and TextBlob, the results of each tweet's sentiment illustrate the differences in human understanding of sentiment and that of a machine. VADER appeared to differentiate sentiment at a better rate than that of TextBlob. This is primarily due to VADER's ability to process emojis and emoticons. Users have limited time and space (280 characters on Twitter) to express their thoughts. The utilization of emojis is often done to substitute words or even an entire sentiment. TextBlob's ability to capture the tweet's sentiment is limited by not considering those factors.

| Rating | VADER | TextBlob |
|---|---|---|
| **Positive** | 2287 | 2247 |
| **Neutral** | 1942 | 2190 |
| **Negative** | 939 | 731 |

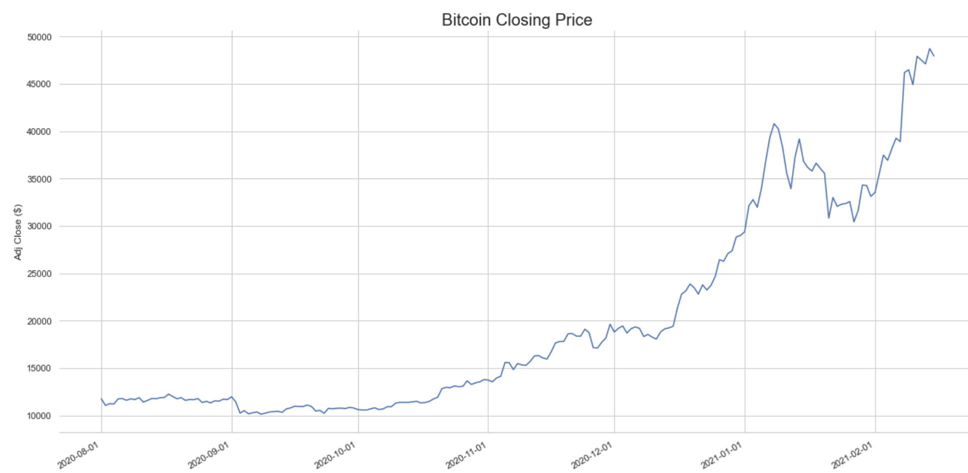
While the tweets' sentiment analysis did shed some light on these influencers' attitudes relating to cryptocurrency tweets, determining the significance and impact these same tweets have on the Bitcoin market is not as clear. Posted below are three charts, and they are as follows:

[6] Hutto, C.J. & Gilbert, E.E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014.
[7] Additional information relating to TextBlob can be found here: https://textblob.readthedocs.io/en/dev/.

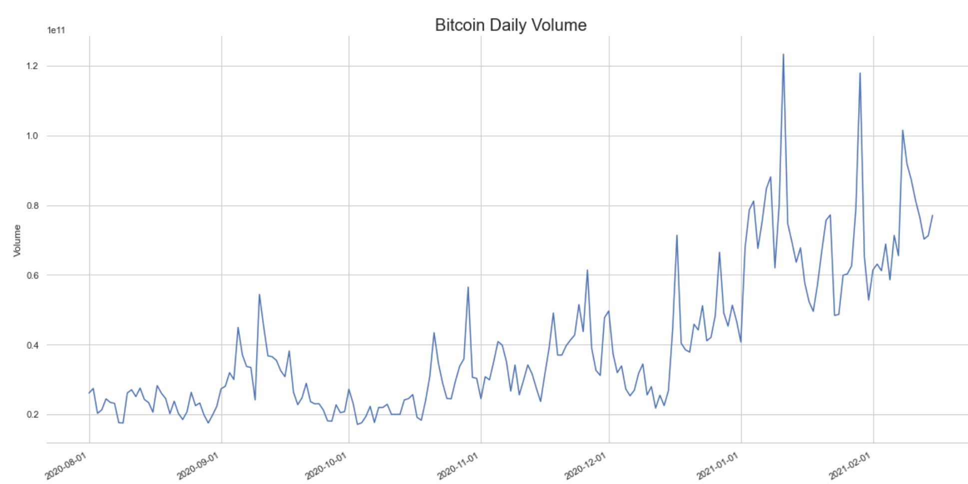
1. Tweet sentiments and daily frequency occurring from 8/1/2020 to 2/15/2021 via VADER




2. Bitcoin's closing price from 8/1/2020 to 2/15/2021




3. Bitcoin daily trading volume from 8/1/2020 to 2/15/2021



By overlaying each of the charts, the Bitcoin market's growth is easier to visualize in relation to the increase in Twitter activity. During this six-month period, the Bitcoin market grew tremendously in price and daily volume. The influencers' volume of tweets increased, but the variance between positive and negative tweets remained relatively the same. Additionally, when reflecting on an individual's specific tweet, there may be instances of correlation, but directly attributing any of the tweets to direct causation in the market's growth is suspect. While sentiment analysis works well for cracking text, much of social media is driven by images beyond emojis, and they are not easily dissected to determine sentiment.

The factor that may come more into play is the influencer's perceived sentiment and influence over their followers, for instance, Elon Musk and his estimated 48.5 million Twitter followers. Elon's tweets are somewhat all over the place regarding his positive and negative outlook on cryptocurrencies. On 12-20-2020, he tweeted first, "Bitcoin is my safe word.", which is determined to be positive, but yet on the same day, he later tweeted, "Bitcoin is almost as bs as fiat money." That tweet would lead one down an opposing viewpoint toward Bitcoin. So, which of the two same-day tweets by Elon mustered a significant impact on the Bitcoin market and his followers? When reviewing notoriously pessimistic Bitcoin skeptic Peter Schiff's tweets, we can see that his tweets are a mix of 182 negative, 129 neutral, and 189 positive sentiments relating to Bitcoin and cryptocurrencies. The results leave room for a much more thorough examination on this topic.

### Bitcoin Pricing Predictions

With Twitter sentiment analysis providing limited correlation and causation results, predicting Bitcoin pricing is limited to market history. Unlike the stock market, the cryptocurrency market is decentralized, relatively unregulated, and traditional fundamentals utilized in evaluating an asset's worth are not as helpful in this space. Unfortunately, this has caused confusion, skepticism, and pessimism. Even as Bitcoin's market cap has hovered around $1 Trillion, global acceptance remains low. This is changing as many multinational corporations invest cash holdings into Bitcoin and accept Bitcoin as a form of payment. With cryptocurrencies not being a traditional currency or stock-like asset, predicting how one like Bitcoin will perform is difficult but not impossible.

Before the model construct could begin, the historical data was reviewed. Since this information is readily available, only four fields had NaN values, and they were forward filled. Unlike the stock market, the crypto market never closes; with that consideration, a 'closing adjusted price' was utilized in predicting prices. The closing price is the last price of that given day.

For prediction comparisons, two models were constructed, one utilizing Prophet and the other a Long Short Term Memory (LSTM), a modified artificial recurrent neural network (RNN). Prophet is a procedure for forecasting time series data based on an additive model where non-linear trends are fit with yearly, weekly, and daily seasonality, plus holiday effects. It works best with time series with strong seasonal effects and several historical data seasons. Prophet is robust to missing data and shifts in the trend and typically handles outliers well.[8]

Long Short-Term Memory networks can be considered extensions of RNNs, once more applying the concept of preserving the context of inputs. However, LSTMs have been modified in several important ways to interpret past data with superior methods. The alterations made to LSTMs deal with the vanishing gradient problem and enable LSTMs to consider much longer input sequences.[9]

---

[8] Facebook Prophet. https://facebook.github.io/prophet/ & https://facebook.github.io/prophet/docs/quick_start.html#python-api.
[9] Nelson, D. (2020, August 3). What are RNNs and LSTMs in Deep Learning? https://www.unite.ai/what-are-rnns-and-lstms-in-deep-learning/.

The results from the LSTM model[10] are interesting and are illustrated below.

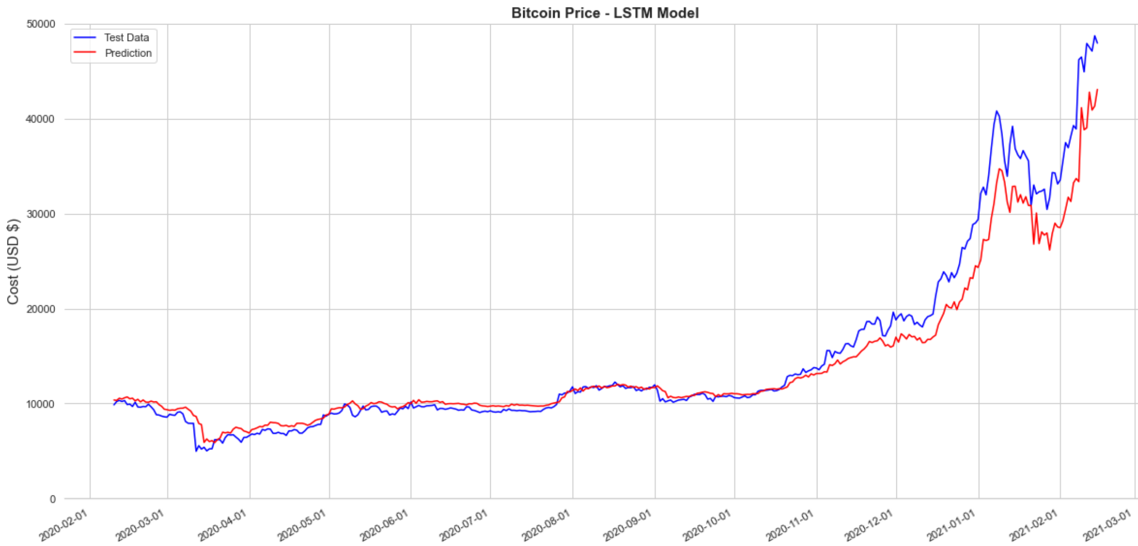



Figure 4 - LSTM Bitcoin Pricing Model

The model's initial view is a relatively accurate prediction based on the trailing three days Bitcoin closing price. However, the model does not handle volatility well, which is a significant Bitcoin component. Beginning 11/1/2020, the model starts to underpredict the price of Bitcoin continuously. The margin of error is substantial, especially in a trader's eyes. This leaves the model with limited appeal in accurately predicting a price.

The Facebook Prophet model[11] is interesting as well. This model is geared toward forecasting markets with strong seasonal and holiday effects. The outcome of the model is listed below.


Figure 5- Facebook Prophet Bitcoin Pricing Model

---

[10] Details on the LSTM model creation can be found here: https://github.com/MarcumDoug/Bitcoin_Sentiment_Analysis_Market_Predictions/blob/main/Models/LSTM_Model.ipynb

[11] Details on the Prophet model creation can be found here: https://github.com/MarcumDoug/Bitcoin_Sentiment_Analysis_Market_Predictions/blob/main/Models/Prophet_Model.ipynb.

The Prophet model is easy to construct and requires only a minor modification to the data to begin. Much like the LSTM model, volatility is a variable that the Prophet model has a difficult time capturing. The flexibility in forecasting is simple and easy to navigate, allowing quick input and modifications.

## Conclusion

Overall, the cryptocurrency market is extremely speculative. Drawing conclusions based on sentiment analysis is difficult, particularly when pulling source data from one medium with little overlap or other media analysis in conjunction. However, sentiment analysis does provide a vital role in quickly determining how an idea, product, person, and so on are being received or judged. For influencers, the sentiment itself might not always be clear, but delivering a message to a large following can lead to a chain reaction of perceived opinion in the group. This could be the case when speculating the influential powers of Elon Musk in relationship to Bitcoin or Dogecoin markets.

Regarding predicting Bitcoin pricing, this speculative, combustible market is challenging to predict. With little data other than perceived value, the asset's volatility has swings that are not commonly seen in more traditional markets. At this point, no one is certain as to what the future holds for cryptocurrencies, but as they grow, more data will be created and consumed in hopes of catching lightning in a bottle.

**References**


Abraham, J., Higdon, D., Nelson, J., and Ibarra, J. (2018). Cryptocurrency Price Prediction Using Tweet Volumes and Sentiment Analysis. *SMU Data Science Review:* Vol. 1: No. 3, Article 1. Retrieved from https://scholar.smu.edu/datasciencereview/vol1/iss3/1.

Akyildirim, E., Göncü, A., & Sensoy, A. (2018). Prediction of Cryptocurrency Returns using Machine Learning. Retrieved from https://www.researchgate.net/publication/329322600_Prediction_of_Cryptocurrency_Returns_using_Machine_Learning.

Ashayer, Alireza. (2019). *Modeling and Prediction of Cryptocurrency Prices Using Machine Learning Techniques.* Retrieved from http://hdl.handle.net/10342/7280.

Ganetti, E. (2020, November 19). How to Build a Bitcoin Sentiment Analysis Using Python and Twitter. Retrieved from https://medium.com/analytics-vidhya/how-to-build-a-bitcoin-sentiment-analysis-using-python-and-twitter-beb89e6ce0c8.

Lertlumprasert, C. (2018, January 6). Cryptocurrency Price Prediction with Python. Retrieved from https://towardsdatascience.com/demystifying-cryptocurrency-price-prediction-5fb2b504a110.

Lielacher, A. (2020, June 27). 3 Bitcoin Sentiment Analysis Tools: How They Work & How to Use Them. Retrieved from https://cryptonews.com/exclusives/top-3-btc-sentiment-analysis-tools-how-they-work-how-to-use-6889.htm.

Munson, E. "Sentiment Analysis of Twitter Data" (2018). *Theses and Dissertations* - 1853. Retrieved from https://scholar.afit.edu/etd/1853.

Valencia F, Gómez-Espinosa A, Valdés-Aguirre B. Price Movement Prediction of Cryptocurrencies Using Sentiment Analysis and Machine Learning. *Entropy*. 2019. https://doi.org/10.3390/e21060589.

Yildrim, S. (2020, April 15). Cryptocurrency Prediction with LSTM. Retrieved from https://towardsdatascience.com/cryptocurrency-prediction-with-lstm-4cc369c43d1b.

Zhao, D., Rinaldo, A., & Brookins, C. (2019). Cryptocurrency Price Prediction and Trading Strategies Using Support Vector Machines. Retrieved from https://export.arxiv.org/pdf/1911.11819.

## Appendix A

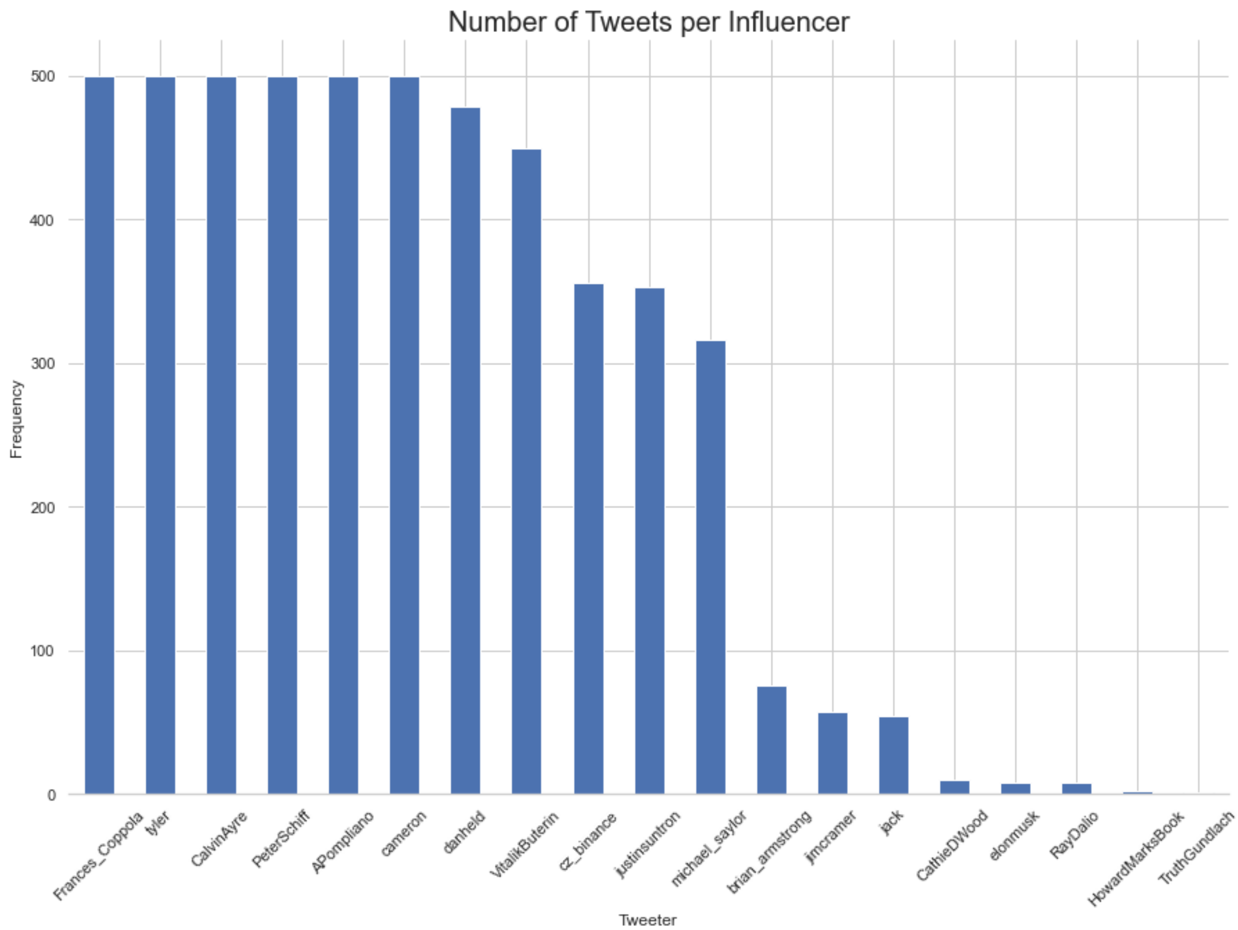
### Keywords

Bitcoin, BTC, crypto, cryptocurrency, #bitcoin, #BTC, ETH, Ethereum, #ETH, #ethereum

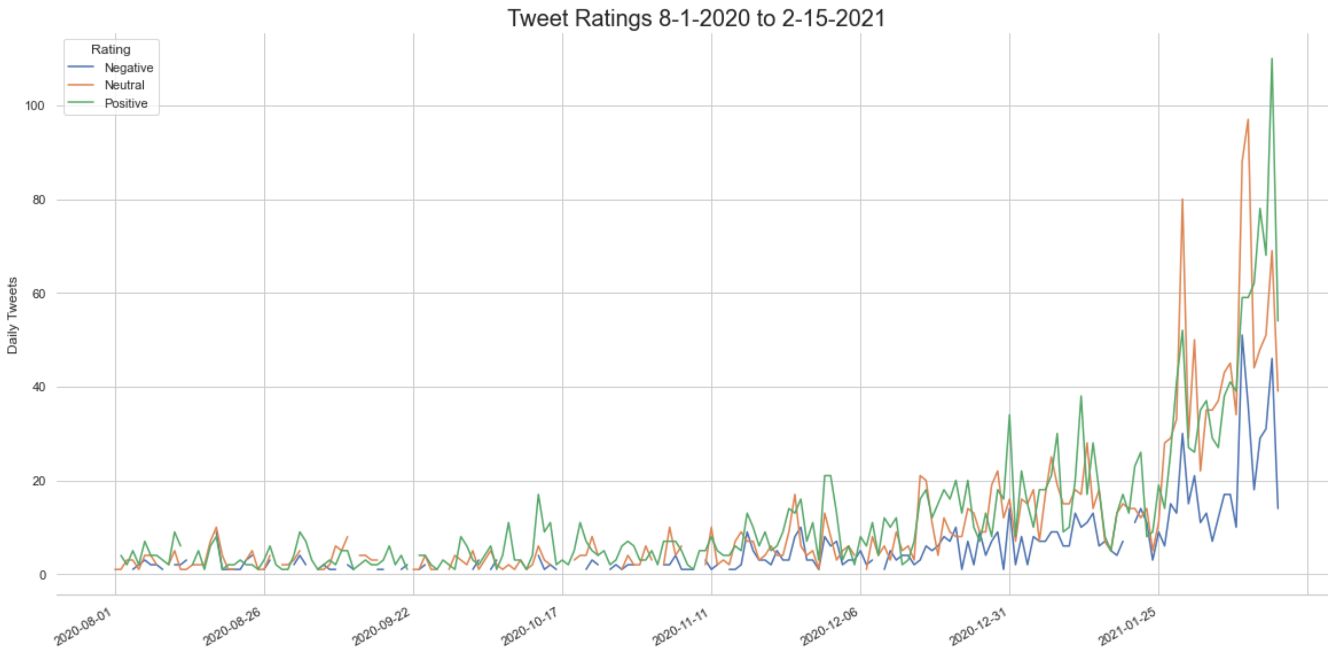
### Influencers

| User | Twitter Handle | Followers | Bio |
|---|---|---|---|
| Anthony Pompliano | @APompliano | 611.1K | Investor |
| Bill Ackman | @BillAckman | 282.4K | CEO of Pershing Square Capital Management |
| Brian Armstrong | @brian_armstrong | 492.9K | CEO/CO-Founder Coinbase |
| Calvin Ayre | @CalvinAyre | 28.7K | Founder Ayre Group and CoinGeek.com |
| Cameron Winklevoss | @cameron | 401.6K | Investor |
| Cathie Wood | @CathieDWood | 659.2K | Founder, CEO, and CIO of ARK Invest |
| Changpeng Zhao | @cz_binance | 1M | CEO of Binance |
| Dan Held | @danheld | 146.3K | Growth Lead at Kraken |
| Elon Musk | @elonmusk | 48.1M | CEO Tesla, SpaceX, and many other things |
| Frances Coppola | @Frances_Coppola | 59.4K | Freelance Writer and Speaker |
| Howard Marks | @HowardMarksBook | 86.5K | Co-Chairman & Co-Founder of Oaktree Capital |
| Jack Dorsey | @jack | 5.2M | Founder/CEO Twitter, Founder/CEO Square |
| Jeffrey Gundlach | @TruthGundlach | 174.2K | Founder of Double Line Capital LP |
| Jim Cramer | @jimcramer | 1.6M | Founder of The Street, Host Mad Money CNBC |
| Justin Sun | @justinsuntron | 2.2M | Founder of Tron, CEO Rainberry |
| Michael Saylor | @michael_saylor | 605.9K | Founder, CEO MicroStrategy |
| Peter Schiff | @peterschiff | 419.7K | Chief Market Strategist and Sr Economist |
| Ray Dalio | @RayDalio | 637.1K | Founder of Bridgewater Associates |
| Tyler Winklevoss | @tyler | 556.1K | Investor |
| Vitalik Buterin | @VitalikButerin | 1.2M | Co-founder of Ethereum |

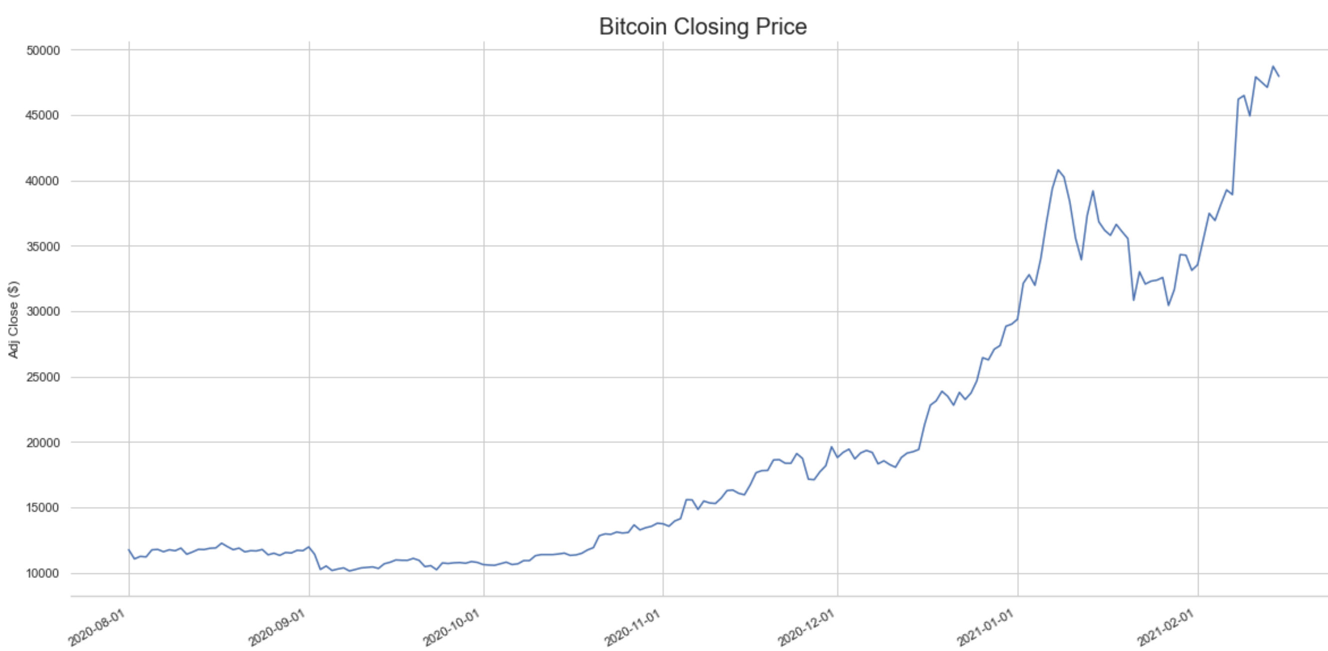
**Appendix B**


Number of Tweets per Influencer
Frequency
0
100
200
300
400
500
Frances_Coppola
tyler
CalvinAyre
PeterSchiff
APompliano
cameron
danheld
VitalikButerin
cz_binance
justinsuntron
michael_saylor
brian_armstrong
jimcramer
jack
CathieDWood
elonmusk
RayDalio
HowardMarksBook
TruthGundlach
Tweeter

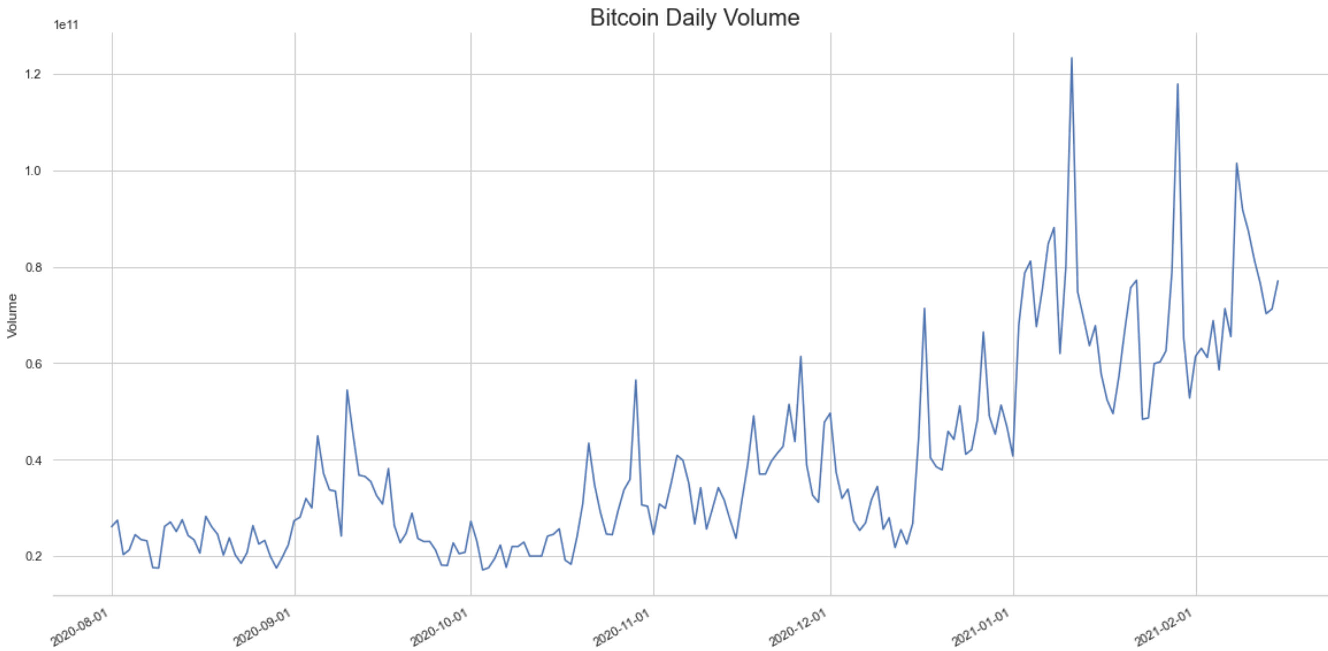
# Appendix C



Tweet Ratings 8-1-2020 to 2-15-2021
Rating
Negative
Neutral
Positive
Daily Tweets
0
20
40
60
80
100
2020-08-01
2020-08-26
2020-09-22
2020-10-17
2020-11-11
2020-12-06
2020-12-31
2021-01-25

Bitcoin Closing Price
Adj Close ($)
10000
15000
20000
25000
30000
35000
40000
45000
50000
2020-08-01
2020-09-01
2020-10-01
2020-11-01
2020-12-01
2021-01-01
2021-02-01



Bitcoin Daily Volume
1e11
1.2
1.0
0.8
0.6
0.4
0.2
Volume
2020-08-01
2020-09-01
2020-10-01
2020-11-01
2020-12-01
2021-01-01
2021-02-01

# Appendix D

Bitcoin Price - LSTM Model
Test Data
Prediction
Cost (USD $)
50000
40000
30000
20000
10000
0
2020-02-01
2020-03-01
2020-04-01
2020-05-01
2020-06-01
2020-07-01
2020-08-01
2020-09-01
2020-10-01
2020-11-01
2020-12-01
2021-01-01
2021-02-01
2021-03-01


Bitcoin Closing Price
Adj Close ($)
50000
40000
30000
20000
10000
0
2015-11-01
2016-03-01
2016-07-01
2016-11-01
2017-03-01
2017-07-01
2017-11-01
2018-03-01
2018-07-01
2018-11-01
2019-03-01
2019-07-01
2019-11-01
2020-03-01
2020-07-01
2020-11-01
2021-03-01
2021-07-01
2021-11-01