Sentiment Analysis of Cryptocurrencies using Tweets (1498 words)

Himani Anil Deshpande

Data Science, Indiana University Bloomington

March 27, 2022

1 Research Question

The objective of this project is to find how the emotions/sentiments of the US population changed when Elon Musk changed his Twitter Bio to Bitcoin, to show his support on 29th January 2021. [1] So, I am looking at tweets which were made about Bitcoins; before, during and after the event. I have performed sentiment analysis using VADER (Valence Aware Dictionary and sentiment Reasoner) lexical analysis [2] and NRC sentiment and Emotion Lexicons(EmoLex)[3]. Also, analyze the engagement that the tweets have received using the number of likes and retweets.

2 Introduction

Bitcoins is one of the oldest cryptocurrency based on Blockchain technology, which was created to make a decentralized system for issuing new coins and recording transactions. The decentralized ideals are the main selling point as they are not owned by any government or authority and the crypto market is open 24/7.[4] However, Studies has shown a strong co-movement between the Bitcoin/cryptocurrency price movements and the sentiments of investors using social media.[1,5,6,7] No one knows what moves the 1.3 trillion Dollar crypto market however they are affected by the tweets made by Elon Musk from January to February 2021 which is against the decentralized ideal of crypto market.[5] Studies have tried to analyze the effect influencer like Elon Musk has on sentiments of netizens.[1,9]

3 Methods

I have used Twitter Developer API v2 to gather at 819 random Original English tweets which have USA geographical data and mention another users. I have gathered data about Bitcoin from 26th Jan 2021 to 3rd Feb 2021 with

Hashtags: “#bitcoin”, “#bitcoins”, “bitcoin”, “bitcoins”, “BTC”, “XBT”, and “#satoshi”

Bitcoin’s ticker symbols: “XBT”, “$XBT”,“#BTC”, and “$BTC”. [8,9]

As I will be looking at the sentiments change I have collected data from 26th Jan 2021 to 28th Jan 2021 which is **Before** the Update by Elon Musk, 29th Jan 2021 to 31st Jan 2020 for **during** the event and 1st Feb 2021 to 3rd Feb 2021 for **after** the update.

I have collected, parsed and stored data about the Tweet Id, Author ID(user ID), Tweet content, country, place name, public metrics like reply, likes, quotes and retweet counts, source and timestamp of tweet using Tweepy package. The only pre-processing I have done is removed the URL’s and Usernames from the tweet text and removing stopwords from tweet content which is specific for EmoLex

3.1 Analysis

There are two different methods I have used for analyzing the tweets.

3.2 VADER

VADER is faster than any Machine Learning algorithms and would process the tweets. It one of the best tools for handling social media text data as it can handle unlabelled data as well as understands slang’s, punctuation’s, emoticon’s and importance of capitalized words in textual data which is why a lot of pre-processing wasn’t required. [2,8,11] It categorizes the data into posi-

Clean Final Class	Manual Check (Manual Check)	Count of Manual Check (Manual C..
negative	Negative	8
	Neutral	2
	positive	5
neutral	Negative	2
	Neutral	12
	positive	11
positive	Negative	4
	Neutral	4
	positive	12

Figure 1: Count of Correct and Incorrect Classifications by VADER

tive, negative and neutral scores which are then normalized in a compound score, in the range of -1 to 1.

Referencing this paper [8], I classified the sentiment using the compound polarity score. The tweet was classified as negative if the compound score was below -0.05 and if it is above 0.05 the tweet is classified as positive otherwise it is neutral.

3.3 EmoLex

Second Method of analysing the tweets with NRC Sentiment and Emotion Lexicons [3], which classifies a list of words belonging to categories of eight emotions (i.e. anger, anticipation, disgust, fear, joy, sadness, surprise and trust) and two sentiments (i.e. negative and positive). Using this I have analyzed the change in emotions that the population shows before, during and after the Bio update made by Elon Musk.

3.4 Manual Analysis

I manually analyzed 60 tweets to check if the classification based on VADER scores were correctly identifying statements. Figure 1 shows the count of False positives, However, I have classified outright support of Bitcoins as positives and direct criticism as negative, neutral if the content is off topic.

I have considered the simple tweet text "Bitcoins" as positive which was classified as Neutral by VADER.

Positive: VADER Neutral : Bitcoin

Negative : VADER Positive : Glad I never bought into bitcoins...

Neutral: VADER Neutral : @elonmusk come to our Bitcoin meet up in Anaheim, .

4 Results

4.1 Distribution of Sentiments from VADER over The Timeline

Figure 2 shows a Bar chart with the distribution of the sentiments that the 819 tweets were identified with using VADER. We can see the constant trend of positive sentiments throughout. 77.17% of my data is from January out of which 41.76% belongs to 29th Jan 2021.

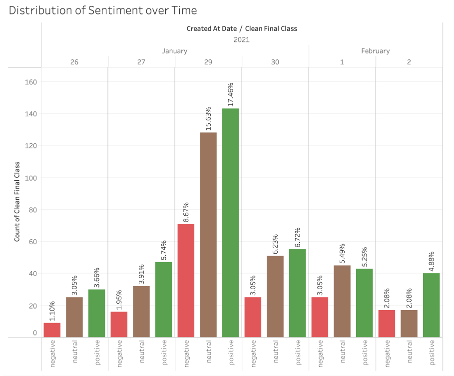

Figure 2: Distribution of Sentiments over Time

Figure 3: Wordcloud of Tweet content with occurrence above 50

4.2 WordCloud of Tweet Content

I have created a wordcloud of cleaned tweet content with word count above 50. From Figure 3 we can see that Bitcoin, Btc, Crypto, Blockchain,money, buy, dodgecoin, etherium, have been repeatedly used.

We can see positive and neutral recurring words like stocks, get, take, buying, new, robinhood, trading, investment, coinbase, people, one time, link and great.

4.3 Top 10 States in USA with Twitter posts on Bitcoins

Figure 4 shows a Bar chart with top 10 US states out of 47 that are in my dataset. 36.01% of my tweets originate from California, USA out of which 19.14% are positive tweets. Followed by New Jersey, Florida and Texas with around 12-13% of tweets.

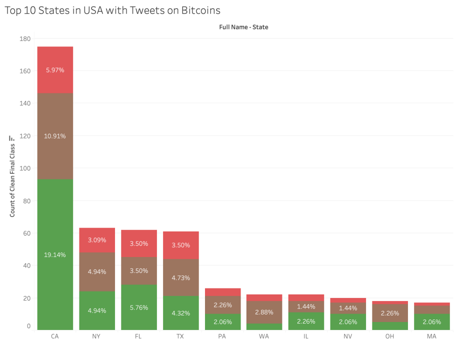

Figure 4: Top 10 States with high Twitter posts

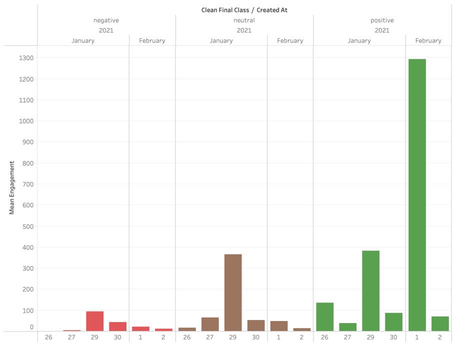

Figure 5: Trend of Tweets by Mean Engagements

4.4 Trend of Mean Engagement and Retweets by Sentiments and Time

Figure 5 shows a Bar chart with the trend of sentiments over Time periods using Mean engagement, which is average of retweets and likes. Similarly, Figure 6 shows the Trend of sentiments using Retweet counts. From both the figures, we can see that the overall trend for neutral and negative tweets is peaking at 29th January 2021 and then decreasing in **before** and **after** the event. The positive tweets peaks **After** the event.

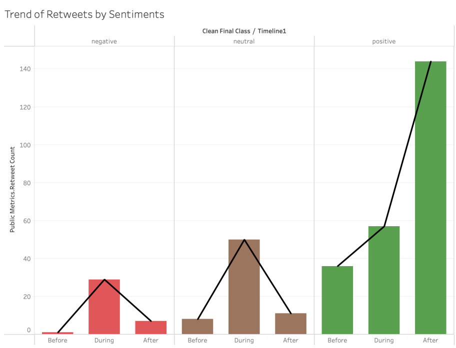

Figure 6: Trend of Retweets by Sentiments

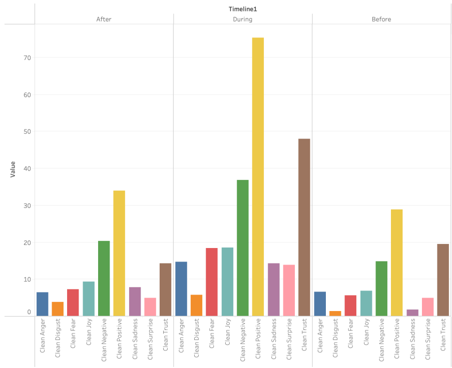

Figure 7: Bar Chart of Emotions and Sentiments by NRC sentiment Analysis

VADER		NRC	
negative	-72.9	Clean Negative	72.10
neutral	0.2	Clean Positive	138.17
positive	192.3		

Figure 8: Scores of Sentiments generated by NRC and VADER

4.5 Sentiments and Emotions by NRC

Figure 7 shows the Bar Chart of all the emotions and sentiments that EmoLex produced. Similar to VADER we can see that NRC shows highest score for positive sentiment. We can see an overall peak **during** the event. **Trust** has increased in **after** the event when compared to the **before** the event. On the other hand, emotions of Fear, joy, sadness and disgust has decreased in **after** period compared to **before** period. Figure 8 shows the overall scores generated by NRC and VADER for the sentiments they identify.

4.6 Frequent Platform used for posting

Figure 9 shows a Bar chart with top platforms used for posting the tweets. We see that iPhone users are most active followed by Android users.

5 Conclusions

My goal is to find how the sentiments of US population has changed after the update that Elon Musk did on his Twitter Bio showing is support towards Bitcoins. The sentiments by VADER as seen from Figure 2 shows that the positive sentiments were high **during** the event. Similarly for EmoLex, Figure 7 shows that the positive scores/sentiments were high **during** the

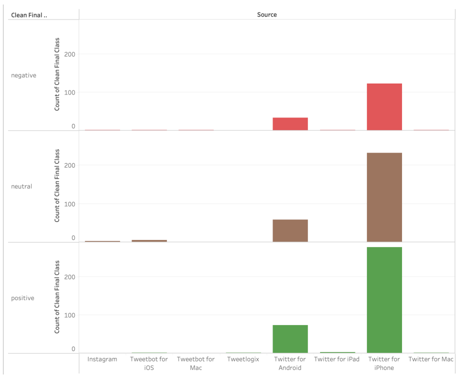

Figure 9: Platforms for Twitter content

event. And both the sentiment analysis method shows that the positive sentiments remains almost same **before** and **after** the event.

The overall negative and neutral sentiments have increased **afterwards** compared to **before** the change as seen in Figure 2.

EmoLex gives more insight into the emotions that the tweets evoke and we see a wide range of emotions which we could not have seen when using VADER. Unlike VADER, EmoLex shows a decrease in negative sentiment **after** the change. But, emotions like disgust, fear and sadness see a decrease. However, its pertinent that these emotions are not always negative, they can also be positive in different context.[3] Notably, **Trust** shows an increase when comparing the **before** and **after** period.

We can see that influencer like Elon Musk does have a huge effect on the Volume of the twitter post about the topic he has tweeted. And with the small dataset I can observe that he also has a slight sway over the sentiments of the crowd.

6 Limitations and Future Works

The dataset of 819 tweets is small and skewed towards a date range of 29th to 31st Jan 2021. Also, I have focused on US population with English language tweets which limits the research as people all over the world invests into cryptocurrency.

Sentiment analysis using VADER's compound polarity score instead of using the individual polarity scores of negative, positive and neutral sentiments could lead to much different results.[9] The pre-processing of removing URL's and usernames from the original tweet doesn't affect the compound score, but it does slightly affect the individual polarity scores, which could lead to different results.

The codebook for Manual Analysis can be more sophisticated. However, the comparison between manual and VADER analysis shows that most of the tweets were correctly classified but it is difficult for VADER to classify without the context of the research. For example I consider "Bitcoin" as positive however it is appropriately classified as neutral without any context by VADER.

I have also focused on a positive event like Elon Musk supporting the Bitcoin. To reaffirm that Elon musk can sway the sentiment of netizen's would require an extensive research on the different tweets that he has posted about Bitcoin and other cryptocurrency.

References

1. Ante, L., (2021, February 3) How Elon Musk's Twitter Activity Moves Cryptocurrency Markets. Blockchain Research Labs BRL Working Paper Series No. 16 Social Science Research Network. Available at SSRN: https://ssrn.com/abstract=3778844 or http://dx.doi.org/10.2139/ssrn.3778844
2. Hutto, C. J. Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text.. In E. Adar, P. Resnick, M. D. Choudhury, B. Hogan A. H. Oh (eds.), ICWSM, : The AAAI Press. ISBN: 978-1-57735-659-2
3. Mohammad, Saif Turney, Peter. (2010). Emotions Evoked by Common Words and Phrases: Using Mechanical Turk to Create an Emotion Lexicon. Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text.
4. Molla, R., (2021, June 14) When Elon Musk tweets, crypto prices move How Elon Musk affects bitcoin prices, in one chart. Vox. URL
5. Parker, E., (2021, May 26) Opinion: Cryptocurrency has an Elon Musk problem. The Washington Post. URL
6. Abdali, S., Hoskins, B., (2021 Spring) Twitter Sentiment Alanysis for bitcoin Price prediction. CS229: Machine Learning Standford University Reports. URL
7. Hassan, M Hudaefi, F Caraka, R. (2021). Mining netizen's opinion on cryptocurrency: sentiment analysis of Twitter data. Studies in Economics and Finance. 38. 10.1108/SEF-06-2021-0237.
8. Pano, T., Kashef, R. (2020). A Complete VADER-Based Sentiment Analysis of Bitcoin (BTC) Tweets during the Era of COVID-19. Big Data and Cognitive Computing, 4(4), 33. https://doi.org/10.3390/bdcc4040033
9. Zaman, S., Yaqub, U. and Saleem, T. (2022), "Analysis of Bitcoin's price spike in context of Elon Musk's Twitter activity", Global Knowledge, Memory and Communication, Vol. ahead-of-print No. ahead-of-print. https://doi-org.proxyiub.uits.iu.edu/10.1108/GKMC-09-2021-0154
10. Beri, B., (2020, May 27) Sentimental Analysis Using VADER. TowradsDataScience. URL
11. Emblem, H., (2021, Jul 18) Are you Scared, VADER? Understanding how NLP Pre-Processing Impacts VADER Scoring. TowradsDataScience. URL