
DATA MINING ON BIG DATA – Twitter Sentiment Analysis on Cryptocurrencies

Narayana Darapaneni *

Anwesh Reddy

David

Sri Karan H C

Ajith KS

Bhushan Tayade

Israel Richards Paul Moses S

Nikita Khadke

Nidhi Goel

Shiven Mehru

Suyash Pandey

1. Abstract

Crypto currency such as Bitcoin is getting more popular these days among investors. It is known that the crypto market is unpredictable, and it is driven by many factors including current affairs, trends, inflation rate, economics, and natural calamities. Many traders believe in and use Twitter tweets to guide their daily cryptocurrency trading. In the proposed work, it is studied to analyse the public opinion regarding the Crypto in Twitter. This solution will collect all the tweets related to crypto using the hashtags, clean the data and make a dataset with it. Exploiting the dataset machine learning module is introduced for analysing the public sentiment of the tweet. The aim of this work is to derive the opinion of the public towards crypto is positive, negative, or neutral. Solution results can be used for reference for potential investors in crypto market.

2. Introduction

Since January 2022, there has been a downtrend in the valuation of the major crypto exchanges, and we have witnessed major crashes among them. Combined losses of the investors across the globe were reported to be more than 1 billion USD and even rendering many firms bankrupt. As an investor who wishes to explore the crypto domain to invest into and diversify their profile, would the current market be a good starting point for the investment? As part of the literature review over multiple research papers available online it was found that Crypto is posted about on social media every 3 seconds.

Cryptocurrencies are said to be driven by speculation, we can assume that this is somewhat true as tweets of certain people have led to increase or decrease in their market value. In this paper, we try to analyse the sentiment of people towards cryptocurrencies based on their tweets.

3. Materials and Methods

We used Flume to scrape tweets from Twitter using the Twitter API. Data consists of following parameters:

- Datetime of the tweet
- Username
- Tweet

To prepare the data for analysis, we take the following steps:

1. Removed @mentions from tweets
2. Removed # hashtags
3. Removed RT from retweets
4. Removed hyperlinks

We used Textblob to get subjectivity and polarity of the tweets. Using the values of polarity, tweets were categorised as positive, negative or neutral.



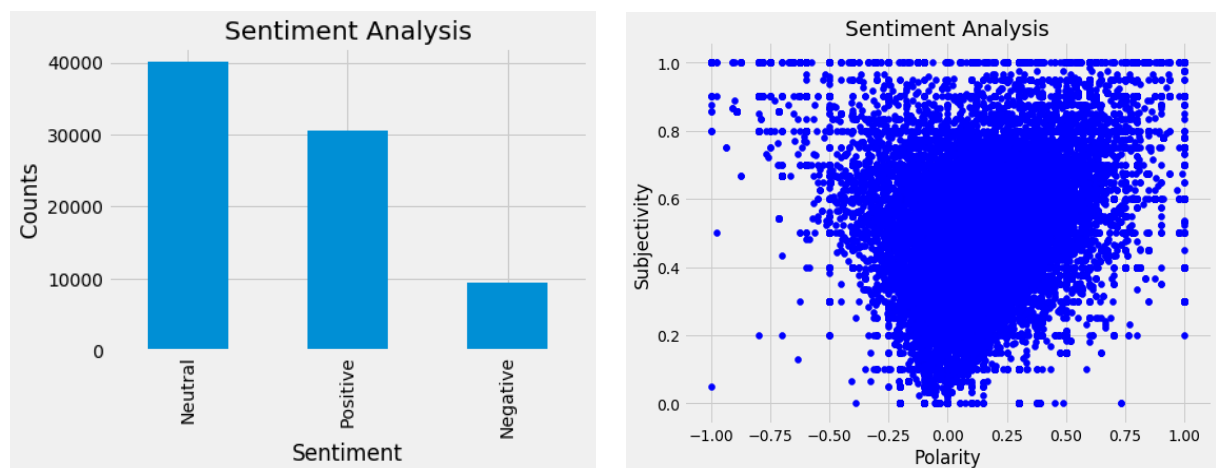
Tools and methodology used:

- Flume
- Twitter API
- Jupyter Notebook
- Spark
- ML Lib (Python)

4. Results

We cleaned the data and got the sentiment of the tweets by using polarity. It was seen that 38.1% of the tweets were positive and 11.8% were negative.

Even though the crypto market has been crashing, people still seem to have positive forecast for this market. Like other markets, this too might be a season and things may turn around but that is yet to be seen.



5. Conclusion

The sentiment analysis provided many important insights over the cryptocurrency sentiment based on the tweets that were fetched.

While most of the tweets were classified as Neutral and that'd be due to the bot generated traffic over the microblogging website consisting of giveaways and shout outs, it was interesting to the volume of positive tweets even during this bear market which consisted of major crashes across some prominent crypto exchanges and many major firms filing of bankruptcy.

As the most experienced investors say that patience is what makes someone successful in this cut throat industry and especially at a time when the whole market is going down and people are losing millions of dollars in their investments, the major sentiment across the website is positive taking this market status as a token of opportunity of the years to come when Cryptocurrency is expected to be at the top of its potential again.

This would be an indicator for already invested individuals or firms to hold on to their portfolio and make slow but steady investments over established coins like Bitcoin and Ethereum if they're willing to take the risk.

This would also be an indication for a new investor who would be looking to diversify their portfolio to consider cryptocurrency as another point of investment, given the current market is not an ideal place to invest in but bear markets do generate the greatest of returns provided the investments are done sensibly and responsibly and one does their part of the due diligence.

Acknowledgement

PGP DE-March 2022 Group 2 would firstly like to thank Great Learning for undertaking this course and providing opportunity to the enrolled working professions to broaden their horizons and step into the world of Data Engineering though this PG Diploma program.

This work would not have been possible without the leadership of Dr. Narayana Darapaneni, the SME's conducting weekly sessions for the course curriculum and our TA's who helped us during each phase of this course with our doubts and for undertaking the weekly mentorship sessions.

We would especially like to thank Gaurav Pandey and Sahil Chawla who helped us with our queries and provided resolutions on the last-minute technical capstone blockers and guided us on the career preparation and expectations. They have shown us by their example, what a good Data Engineer one should aim to be.

References

1. "Bitcoin and Beyond: A Technical Survey on Decentralized Digital Currencies" by Christian Decker, Roger Wattenhofer.
2. "Ethereum Casper Explained" by Vitalik Buterin
3. "A brief survey of Cryptocurrency systems" by Ujan Mukhopadhyay; Anthony Skjellum; Oluwakemi Hambolu; Jon Oakley; Lu Yu