

KABIR CHATURVEDI

Semester: Spring 2024

Navigating the Sentiment Waves: Analyzing Cryptocurrency Discussions on Reddit

(1496 words)

Introduction:

In the volatile world of cryptocurrency investment, market dynamics are heavily influenced by the sentiments held by participants in online communities, particularly on platforms such as Reddit. These sentiments, while individual in nature, can aggregate and catalyze substantial market movements, as notably seen with the GameStop stock incident in 2021, where collective discourse on social platforms precipitated a dramatic market rally. This study applies advanced computational techniques to dissect and understand the collective sentiment expressed across various cryptocurrency forums. By scrutinizing sentiment trends and their reactions to market events, this research aims to demystify the complex interplay between public conversation and cryptocurrency market movements. Pioneering studies like those of Raheman et al. (2022), which delved into the predictive power of social media sentiment on cryptocurrency markets, and Mondal et al. (2023), who examined the causality between sentiment and price changes in the cryptocurrency realm, provide a critical foundation for this research. This work is set to explore the extent to which shifts in online sentiment are correlated with the financial tides within the cryptocurrency markets.

Research Question:

What are the dominant sentiment trends across key Reddit cryptocurrency forums?

Method:

Data:

For this study, the dataset was carefully curated from select posts on three prominent Reddit subreddits: r/CryptoCurrency, r/Bitcoin, and r/SatoshiStreetBets, chosen for their high level of activity and impactful discussions on cryptocurrency. The selection of posts was based on their potential to generate rich, sentiment-laden discussions. Specifically, two posts were chosen from r/CryptoCurrency; the first for its direct query about the current market sentiment, which likely elicits detailed emotional and analytical responses from the community, and the second for its inquiry presented with a negative nuance, providing a range of sentiment perspectives. From r/Bitcoin, a post was selected from a daily discussion thread on March 5, 2024—a day marked by extreme volatility as Bitcoin hit an all-time high and then plummeted by 7%, capturing a wide array of community reactions. A single post from r/SatoshiStreetBets was included due to its concise yet expressive 31 comments. The "sentiment" keyword, not case-sensitive, was

used to filter and select a post from the r/CryptoCurrency subreddit. Additionally, the keyword "Discussion" was used to identify a relevant post from r/Bitcoin. The remaining posts were chosen based on their prominence during the specified timeframe using the 'top posts' feature, ensuring they contained sufficient data points and offered a diverse range of sentiments for the analysis. This selection method was informed by the frequent use of these terms in discussions reflecting market reactions and emotional responses within these online communities. Data was collected using the Python Reddit API Wrapper (PRAW), which interfaces efficiently with Reddit's API, over a period from March 2022 to March 2024. However, the posts selected for detailed analysis specifically contained comments spanning from November 2023 to March 2024, a timeframe chosen to capture sentiment during significant market events and transitions, thus enriching the study's analysis of market sentiment dynamics.

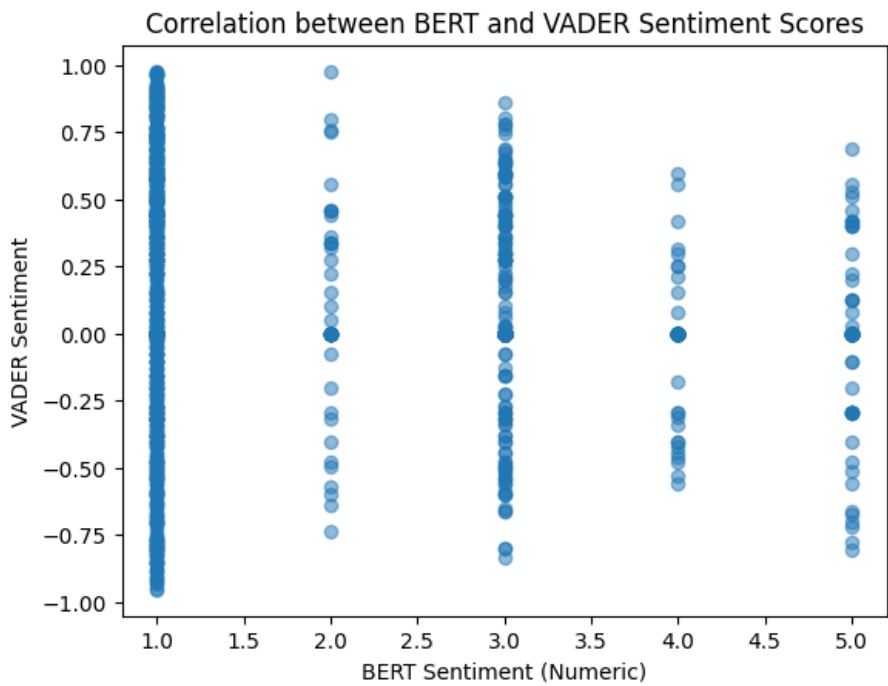
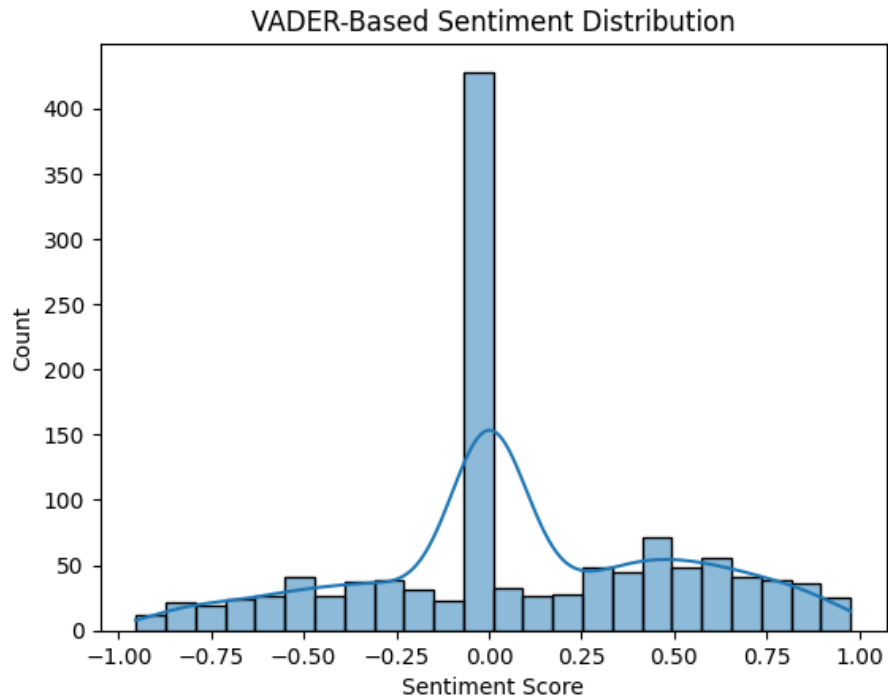
Analysis:

In analyzing the dataset for this study, a thorough data cleaning process was executed using a custom Python function to standardize the text by stripping non-word characters, converting to lowercase, removing single-letter words, and consolidating spaces. This step was crucial for normalizing the data for computational sentiment analysis, ensuring accuracy in the subsequent interpretation. The standardized text data was then truncated to 512 tokens to meet the input constraints of the BERT sentiment analysis model, focusing the analysis on the most pertinent parts of the text. This step, along with removing duplicates and missing values, was essential for enhancing the quality of the dataset.

The visualization of the sentiment analysis results utilized various graphs tailored to specific interpretative goals. Histograms displayed the distribution of sentiment scores, clearly depicting the range and frequency of sentiments across categories, which was pivotal in understanding the overall sentiment landscape. A pie chart was chosen for its effectiveness in illustrating the proportions of sentiment categories within each year, allowing immediate visual comparison. A bar graph was utilized to compare and visualize the frequency of key terms, aiding in the identification of prevalent themes within the cryptocurrency discussions.

Furthermore, a scatter plot was crafted to demonstrate the correlation between BERT and VADER sentiment scores, serving as a valuable tool for assessing the agreement between two distinct sentiment analysis methods. This visualization, by highlighting the relationship between the models, provided insight into the consistency of the sentiment analysis. Additionally, box plots conveyed the variation and range of sentiment scores over time, shedding light on trends and potential outliers in sentiment data. Density plots were also employed to compare sentiment distributions year-over-year, effectively illustrating shifts in sentiment polarity and their potential correlations with market events. Collectively, these visualizations offered a multifaceted view of the community's sentiment, contributing to a comprehensive understanding of sentiment dynamics within cryptocurrency discussions on Reddit.

Results:

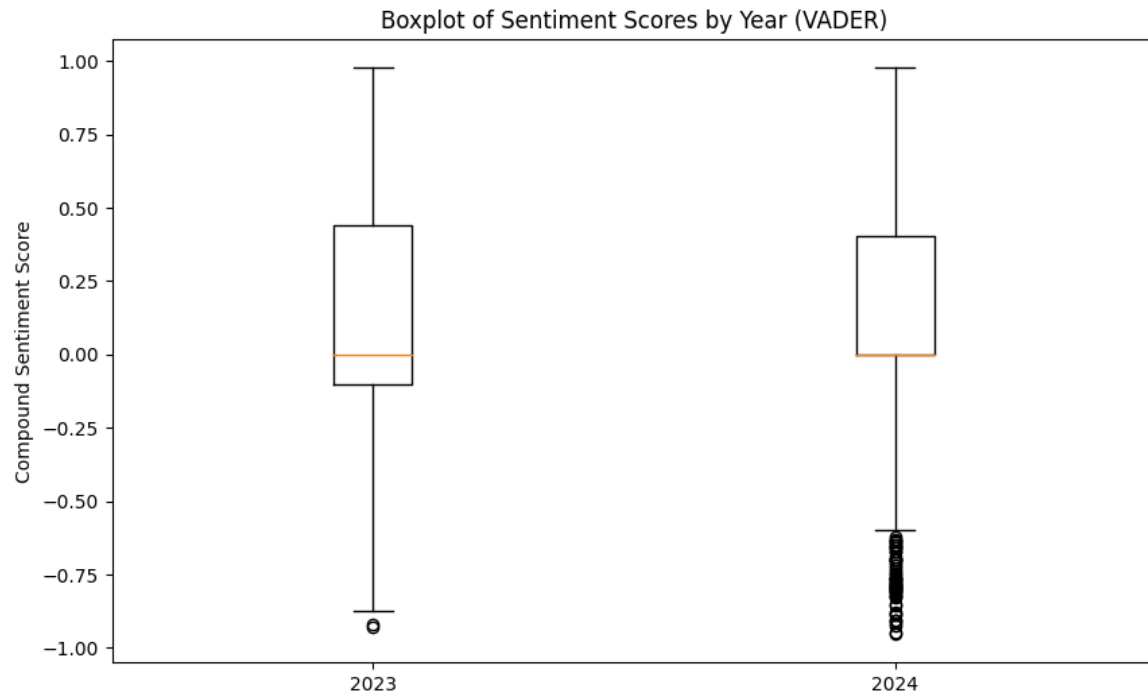


The sentiment analysis results, as visualized in the BERT-based sentiment distribution graph, revealed a notable concentration of negative sentiments, with an abundance of 1-star ratings. This phenomenon suggests either a dataset with predominantly negative discourse or a propensity of the BERT model to classify sentiments as negative. In a stark contrast, the correlation graph comparing BERT and VADER sentiment scores depicted a scattered distribution across moderate sentiment ratings. This distribution

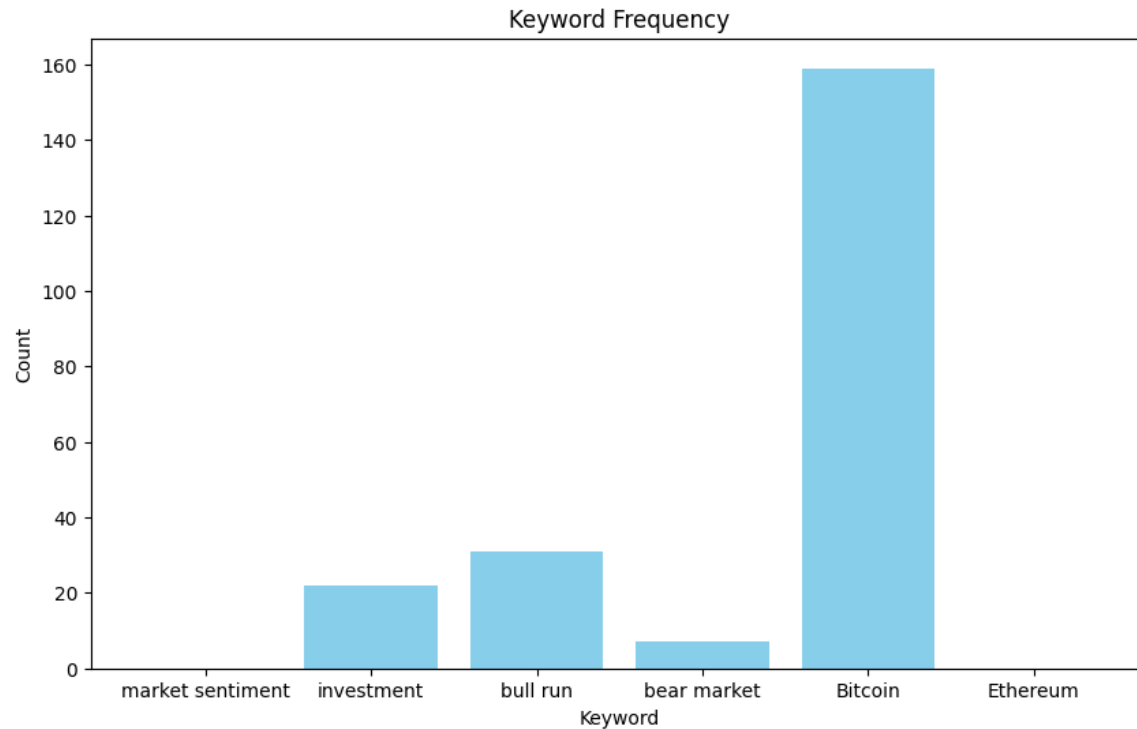
suggests a dissolution of uniformity, prevalent at the extreme ends of the sentiment spectrum, into a variability within the moderate range. The divergence between the models becomes apparent in their treatment of subtler tones: BERT, with its nuanced context understanding, and VADER, with its predefined sentiment lexicon, often vary in their sentiment assessments.

```
📄 If things were shit that would cause a bigger dip
BERT Sentiment: 3 stars
VADER Sentiment: -0.6597
-----
Text: Last chance to buy above 44k
BERT Sentiment: 4 stars
VADER Sentiment: 0.25
-----
Text: Lock in profits
BERT Sentiment: 3 stars
VADER Sentiment: 0.4404
-----
Text: No one knows anything until time passes and we can look at what happened.
BERT Sentiment: 4 stars
VADER Sentiment: -0.296
```

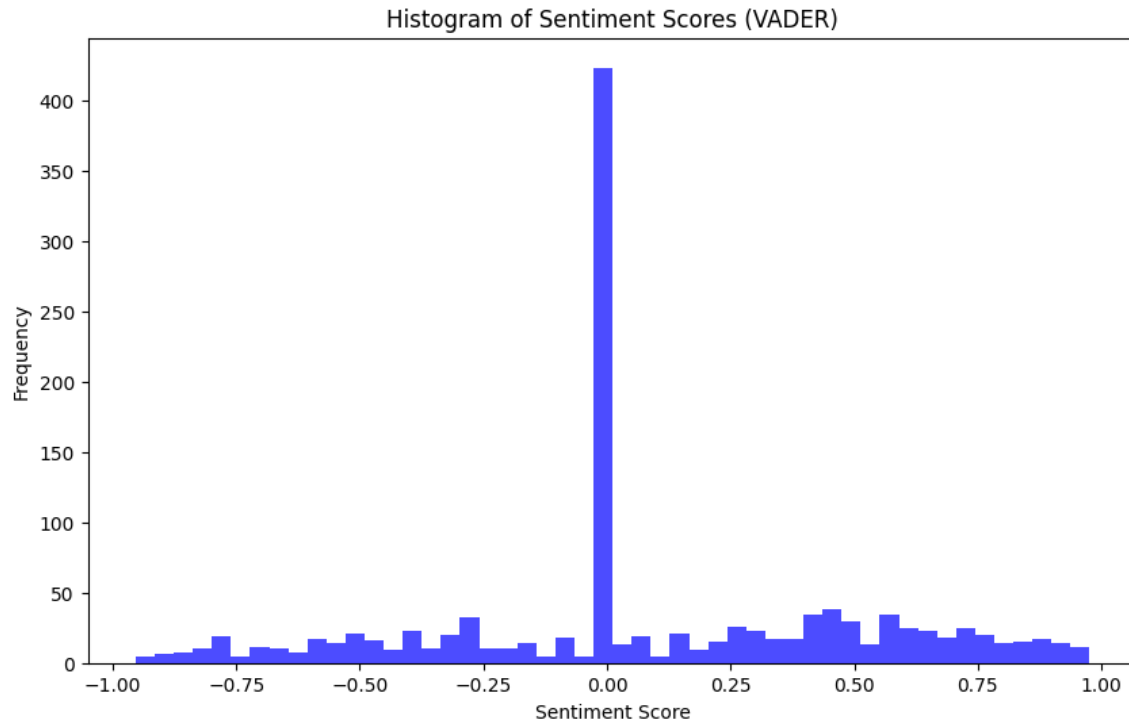
To investigate this further, I wrote code to specifically pinpoint data points where the sentiment assessments of BERT and VADER differed the most. The subsequent analysis indicated that while both models were adept at detecting distinctly positive or negative sentiments, their assessments showed significant variation when confronted with nuanced expressions. This discrepancy underscores the complexity inherent in sentiment analysis, highlighting the necessity to employ multiple analytical perspectives to gain a comprehensive understanding of sentiment dynamics. It is crucial to comprehend these differences, as they shed light on the strengths and limitations of each model.



The boxplot generated from VADER's sentiment analysis for the years 2023 and 2024 indicates a median sentiment score leaning towards neutral, with a slight positive skew in 2023. The distribution of sentiment scores is symmetrically represented by the interquartile range for both years. Notably, the year 2023 has an extreme negative outlier, while 2024 shows a wider variance in sentiment intensities, as seen by the extended whiskers and several negative outliers, signifying particular instances of strong negative sentiment during that year.

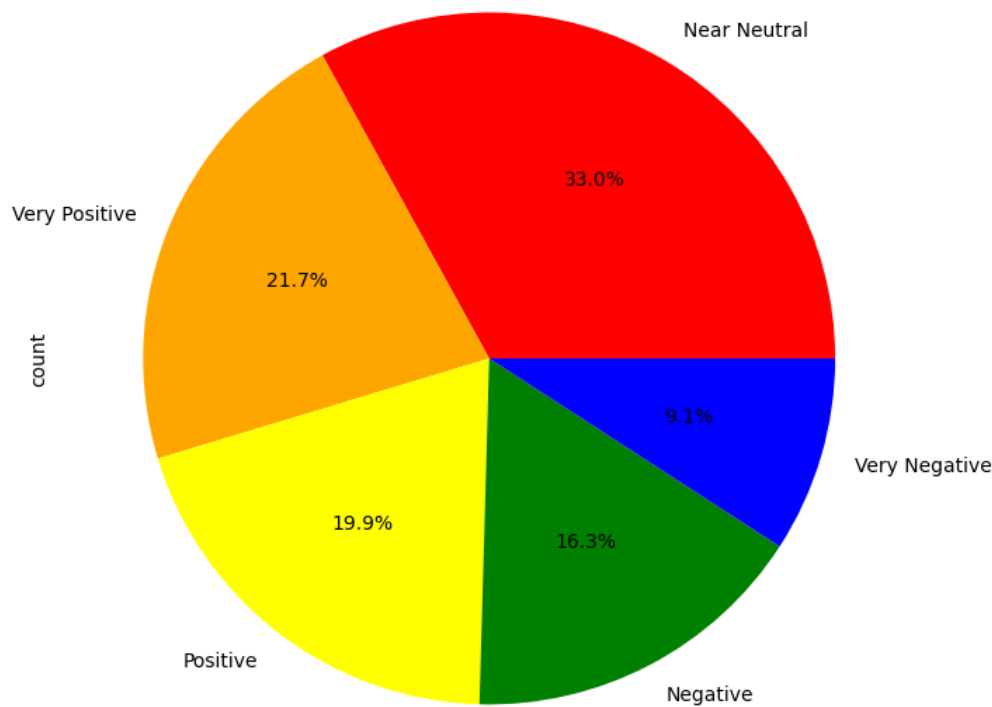


The analysis of keyword frequency illuminates the prevalence of "Bitcoin" and "Ethereum" within the discussions, underlining a significant interest in these specific cryptocurrencies. By contrast, terms such as "market sentiment," "investment," "bull run," and "bear market" are mentioned less frequently, denoting a relative scarcity of broader market terminology in the discourse when compared to the discussions centered around particular cryptocurrencies.



The histogram depicting VADER sentiment scores shows a notable aggregation of data points around a neutral sentiment score, suggesting that the bulk of the sentiments expressed in the dataset are neither overwhelmingly positive nor negative. There are fewer instances of extreme sentiments, as demonstrated by the sparser bars at the ends of the histogram, which reveals a dataset characterized predominantly by neutral sentiment expressions.

Sentiment Category Distribution in 2023



The pie chart for 2023 illustrates the sentiment categories derived from the dataset, with varying proportions representing the prevalence of each sentiment. The 'Very Positive' category, marked in red and constituting 21.7% of the data, indicates a substantial share of highly positive expressions. Close to this, the 'Positive' sentiment in yellow accounts for 19.9%, suggesting that positive feedback is nearly as prevalent as its more emphatic counterpart. Dominating the chart is the 'Near Neutral' segment in orange, occupying a 33.0% share, reflecting a substantial portion of discussions with a balanced or ambivalent sentiment, which could point to measured discussions or a lack of strong sentiment indicators.

On the other side of the spectrum, the 'Negative' sentiments, shown in green, make up 16.3% of the chart, suggesting a discernible but smaller section of the data colored by dissatisfaction or critical views. Lastly, the 'Very Negative' sentiments, represented in blue, account for just 9.1% of the data, signifying the least frequent sentiment category, which indicates that strongly negative expressions are comparatively rare in the dataset. Together, these segments reveal a trend where neutral and positive sentiments are more commonly expressed, with a significant portion of discussions in 2023 gravitating towards these categories. The comparatively smaller fractions of negative and very negative sentiments may imply that the overall sentiment within this context was skewed towards more neutral and positive tones, with fewer instances of pronounced negativity.

Conclusion and Limitations:

The analysis of sentiment dynamics within cryptocurrency-related discussions on Reddit, spanning from November 2023 to March 2024, has yielded insights that align with the fluctuations in cryptocurrency prices during this period. The predominance of neutral and positive sentiments uncovered in the study could correlate with the rising trend in cryptocurrency valuations, as optimistic outlooks generally proliferate in bullish market conditions. Notably, the inclusion of data from a day when Bitcoin reached a new all-time high might have introduced a positive sentiment bias, despite the subsequent 7% price drop, which tempered the sentiment with a dose of negativity. These findings suggest that sentiment in cryptocurrency forums is closely tied to market performance, reflecting the highs and lows of the financial landscape within the discussions.

As for the limitations, the data comes from a limited number of posts and might not capture all the varied sentiments present in the subreddit conversations. Including more posts in future research could provide a fuller picture of the community's feelings and opinions. Moreover, the sentiment analysis methods used here, along with the intricate nature of language, might affect how accurately sentiments are identified. This points to an opportunity for improving how sentiment analysis is conducted in future studies.

References:

- Raheman, A., Kolonin, A., Fridkins, I., Ansari, I., & Vishwas, M. (2022, April 19). *Social Media Sentiment Analysis for Cryptocurrency Market Prediction*. arXiv.org. <https://arxiv.org/pdf/2204.10185>
- Mondal, L., Raj, U., S, A., S, B. G., P, S., & Chandra, A. (2023, June 9). *Causality between sentiment and cryptocurrency prices*. arXiv.org. <https://arxiv.org/abs/2306.05803>