Opinion Dynamics of Bitcoin

The Relationship Between Social Media Sentiment and Market Behavior

Jazlyn Jose [1] , Aanchal Sharma[2], Vansh Saini[3], Unnati Gupta[4], Palak Gupta[5]
Jazlyn Jose (BML Munjal University, (SOET), Kapriwas, (Haryana), India
Aanchal Sharma (BML Munjal University, (SOET), Kapriwas, (Haryana), India
Vansh Saini (BML Munjal University, (SOET), Kapriwas, (Haryana), India
Unnati Gupta (BML Munjal University, (SOET), Kapriwas, (Haryana), India
Palak Gupta (BML Munjal University, (SOET), Kapriwas, (Haryana), India

Abstract— **Background :** This research investigates sentiment trends in Bitcoin-related tweets and their influence on market dynamics.The study consists in investigating the influence that cryptocurrency influencers have on the Ethereum (ETH) and Binance Coin (BNB) values regarding Bitcoin. Also, it compares influential users' sentiments against those of regular users narrated by tweets and examines the correlation between the Bitcoin Forum discussions and market prices. **Methods**: VADER and TextBlob, which are used for sentiment analysis, as well as LDA and NMF for topic modeling, allowed to analyze trends in sentiment for Twitter and Bitcoin Forum data along with topic weights and correlations with Bitcoin prices. **Results**: Research results reveal that tweets about Bitcoin are highly sensitive to crypto influencers when Bitcoin price rises and vice-versa; there might be a possibility New users' influence is not equal to the regular users' one; also the discussion between the people on the Bitcoin forum can affect can affect the market prices. **Conclusion**: The analysis highlights the impact of social media sentiment on cryptocurrency markets as well as the significance of discussions that take place in forums and the inputs of influential users who have the ability to move the prices of bitcoin. On the one hand, it clarifies the mechanism of influence of sentiment on the market perception of digital assets.

Keywords—component; Sentiment Analysis; Topic Modelling; Correlation; Bitcoin

1. Introduction

With the advent of the digital world, cryptocurrencies have become a novel and disruptive method of money and trade that have transformed the financial markets. Among the assortment of them, Bitcoin, which was designed and coded anonymously by a person (2009), is the most prominent one, showing the way for lots of other digital currencies which are based on the blockchain technology. This has constituted a major factor in the transformation of cryptocurrency trading. Nevertheless, it is the volatility of bitcoin price, with sharp changes observed within short intervals, that is considered a drawback which is not successfully addressed by the traditional macroeconomic theories.

Amidst the rising of Bitcoin, a critical question surfaces: Which one is the essential factor of its price? The understanding and appreciation of the factors that govern the price change of the Bitcoin is a theme that must be followed by the investors, the policymakers, and the businesses, as the information can be used for trends forecasting, the making of regulations, and the implementation of digital currencies. In such a situation it is social media content that is the pillar underlying the change, a channel through the spread of info and the moods determines the price. Given the introduction of social media as a reliable communication source where people communicate their opinions so fast and so extensively, this study will be directed towards the influence of social media sentiment on the operation of the Bitcoin markets, paid attention to how the social media tend to predict the future trends and also the different forms of contributions of various users from different platforms.

First of all, in order to go into details and get a deeper understanding of the influence that digital conversations have on Bitcoin's price, our research questions are oriented towards analyzing different aspects of this issue. Our intention is to examine whether the sentiments from the top influencers in the crypto space on Twitter affect Bitcoin price more than the general chatter of the everyday users and the enthusiasts in the media platform. However, the change in sentiment from day to day and the predominance of particular effects during market booms and busts are also investigated in this regard of understanding its capacity to foretell price spikes.

Secondly, to determine the weight of sentiments shared in Bitcoin forums representing diverse fields and those coming from the most renowned social media platforms in the market, our project will explore what channels most affect the market. Finally, this comparison will be a great provider of the things that are to be taken into consideration by investors, policymakers and analysts in the market in order to trace market sentiment and its effects on crypto prices.

This in-depth analysis of social media sentiment and Bitcoin trading dynamics is a major area of research as it prepares communities to develop more sophisticated economic models along with investment plans that take into account the complexities of the digital age. By means of resolution of these crucial issues, we try to make the pathfinder clearer for stakeholders proceeding the wavy cryptocurrencies' market.

Analyzing the factors that drive the prices of cryptocurrencies is essential for having a good opinion of putting in money in crypto and foreseeing the market trends. Twitter with its social media platforms has proved to be another channel for face-to-face communication where what people talk about ranging from how cryptocurrency prices follows through such sentiments effortlessly. The research work in this area in addition will examine the sentiment of twitter messages directed to

Ethereum and Binance Coin and those shared and between the two and the change in Bitcoin price, the leading cryptocurrency in the stock market.

This project is going to facilitate data analytic and sentiment analysis to provide a more all-rounded view on the crypto world. This approach will not only help in prediction of short-term market fluctuations, but also in spotting long-term trends and potential impacts of changes in the society on financial investments and trade. Therefore, knowing sentiment trends becomes a strategic tool which could be used to predict market moves and hence resulting in better management of portfolios.

2. Literature Review

In a comprehensive review of literature, several studies explore the intricate relationship between social media sentiment and Bitcoin metrics. Andrew Burnie and Emine Yilmaz [1] delve into this nexus in their study, "Social media and bitcoin metrics: which words matter," aiming to tie sentiment in tweets and Bitcoin prices, alongside evidence of Granger causality suggesting predictive potential. Similarly, Federico Albanese, Sebastián Pinto, Viktoriya Semeshenko, and Pablo Balenzuela investigate mass media influence on public sentiment during the 2016 US presidential campaign.

Through sentiment analysis and topic modeling on a corpus of over 15,000 articles from prominent news sources, they discern significant relationships between news coverage, public sentiment, and polling variations.identify pivotal words and phrases in social media data that strongly correlate with Bitcoin price fluctuations. Employing sentiment analysis techniques on over 1.2 million Bitcoin-related tweets spanning two years, their findings reveal a robust positive correlation between positive.

Kyeongpil Kang, Jaegul Choo, and YoungBin Kim shift focus to cryptocurrency online communities in their study, "Whose Opinion Matters? Analyzing Relationships Between Bitcoin Prices and User Groups in Online Community." [2]By collecting data from Bitcoin forums and employing network and sentiment analysis, they identify opinion leaders and prevalent topics, highlighting the influence of these factors on community dynamics. Giacomo di Tollo, Joseph Andria, and Gianni Filograsso broaden the scope to include both cryptocurrencies and stock markets in their study, "The Predictive Power of Social Media Sentiment: Evidence from Cryptocurrencies and Stock Markets Using NLP and Stochastic ANNs." They utilize an artificial intelligence system with sentiment analysis on social media data to introduce the crucial part of sentiment analysis for the markets in the short term forecasts.

Extending the discussion, the mentioned research known as "Cryptocurrency Price Prediction Using Social Media Data Mining and Epidemic Modeling" by Sebastian Franz Huppmann, is looking precisely at Twitter sentiment and Bitcoin prices[3]. With the help of innovative statistical models and sentiment analysis process through a large reference database of tweets, Huppmann shows a strong connection between the Twitter mood and Bitcoin's prices, leading to an impressive Forecaster machine.

To the extent, Ali Raheman, Anton Kolonin, Igors Fridkins, Ikram Ansari, and Mukul Vishwas assess multiple sentiment analysis models in the cryptocurrency field appraisingly for AGENTS model to be particularly accurate in revealing the influence of social media sentiment on Bitcoin prices.

On a similar premise, Anitha Reddy, Karnati Reethu, J S Vineesha Raju, and Garneppudi Rishitha carried out another study[4] where, as per their study, they wanted to understand the effect of sentiment in bitcoin tweets and correlated it with bitcoin price trend using the deep learning models for sentiment analysis However, the same as Pavitra Mohanty, Darshan Patel, Parth Patel, and Sudipta Roy, they try out forecasting crypto price fluctuations by combining sentiment analysis and real-time price data, invoking machine learning algorithms into the process to optimize the cost.

Majid Makinayeri studies the investor's triad in social media networks and uses sentiment analysis and ARDL model to look at the association between investors' shared solicitude and Bitcoin price. The results of his research indicate that the social network effect on investor behavior is huge.

An astonishing fact is that the scholars Pavlo Seroyizhko, Zhanel Zhexenova, Muhammad Zohaib Shafiq, Fabio Merizzi, Andrea Galassi, and Federico Ruggeri have developed a dataset of sentiment and emotion annotated from the posts of Reddit which are related to Bitcoin[5], and at the same time that is how Feng Mai, Zhe Shan, Qing These studies as a whole reiterate the fact that the sentiment analysis offers us a lot of insights into the exchange rate patterns of cryptocurrencies, providing evidence of their direct connection with the Bitcoin metrics.

In conclusion, sentimental analysis is mentioned as one of the main factors influencing bitcoin market trend forecasting as revealed by the studied papers[6]. They constantly lead to the establishment that social media sentiment is positively correlated with the Bitcoin prices and determines the usefulness of sentiment analysis models in forecasting. Moreover, the audience of the mass media and social networking investors react to the coverage of Bitcoin and give their feedback which becomes one of the main factors in the price movement of Bitcoin.

With that being the case, sentiment analysis has indeed proven its utility in addressing both the weighing of the market sentiment and the prediction of the near-term price movements. Capitalizing on the enhanced statistical models and machine learning algorithms, empirical researchers point out that sentiment analysis is useful in detecting market trends and as a guiding tool for the trading strategies.

Conclusively, the literature expounded is the base to initiate the understanding of the complex relationship between social media sentiment and cryptocurrency market dynamics. At present time the crypto market environment is evolving and the situation when sentiment analysis integrates into the market analysis and prediction models remains fundamental for the investors' informed choice-making process.

3. Data Overview

3.1 Dataset

3.1.1 Bitcoin Tweets Dataset

The Bitcoin Tweets Dataset covering 2021-2022 comprising around 4m is homogeneous in terms of its word choice, word emotionally and contextual words. 5 million of the tweets analyzed. Kaushiksuresh147 is by Twitter. This wealth of knowledge has allowed us to explore the heated debates surrounding Bitcoin on a variety of topics, with participation from 61 participants. As well as the text of a tweet and the time when it was published, the dataset has lots of very useful metadata which includes information about users (user name, location, description, creation data, followers, friends, favorites, and the team members who have been verified), along with the details about hashtags, sources of the tweets and retweets.

3.1.2 Bitcoin Influencer Tweets Dataset

The Bitcoin Influencer Tweets Dataset, which Chahooki Zare and covers the period 2021 – 2023, offers a specialized angle on the conversations about cryptocurrency from influential figures. Within which 40,000 tweets from people all over the world are extracted, containing not only their posts and timestamps but also delivers engaging metrics like favorite and retweet counts. In addition to this, there is a large amount and a comprehensive view of interactions provided that indicates the comments, date (date of creation), favorites counts, followers counts, listed (followers), media (media pieces and associated multimedia content) counts, screen (screen user name) names and URLs (internet address).

3.1.3 Bitcoin Forum Data

The Bitcoin Forum Data, gathered by Kyeongpil Kang, and covering the period of 2009 - 2018, is a unique source, providing an in-depth look at the history of Bitcoin discussion on the forum. These 36,898 posts uttered by 353,679 users and authored by 32,304 users make up the decade-long dataset, which offers boundless opportunities for analysis, covering the various aspects of the Bitcoin ecosystem such as discussions, debates, and advancements. The dataset may consist of posts, timestamps, (probably some user information) which are helpful for analyzing the language of Bitcoin communities through time.

3.1.4 Ethereum and Binance Influencer Tweet Dataset

The Ethereum and Binance Twitter Influencer Dataset gathered by Ilaria Mazzoli and representing conversations on Ethereum and Binance from the elite opinions in cryptocurrency industry in the period from July 2022 to August 2022, gives us a compressed view of the current situation in the world of cryptocurrency engagement. Dataset with 100K tweet data consists of tweet text, sources, date of tweet, whether the tweet is replied to, retweet, favorite count and also URLs for media tweeted as well as a variety of user information. This data provides very detailed analytics of user IDs, creation dates, follower stats, verification status and others that occurred during this specific time period and measure users' engagement and sentiment. Having measured the number of followers ranged between 30,000 and 40,000 for these influencers, the tweet length is also not transcending 190 characters.

3.2 Data Preprocessing

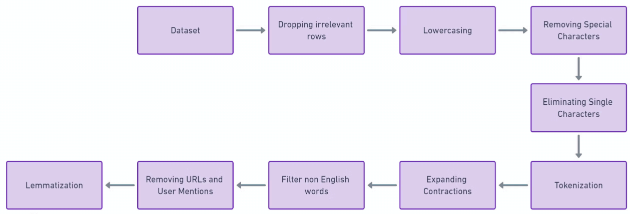


Fig 1. Data Preprocessing flow diagram

Before proceeding with the analysis, data undergoes a necessary pre-processing stage where it is cleansed and prepared, so that the information discovered from the data can be accurate and informative. The steps involved in this research are the deletion of irrelevant rows which means removing NaN-valued rows and those rows whose data is not required for the analysis to be successful. For instance, twitter tweets and posts are packed in rows. When text is missing or when the information is incomplete, the hashing excludes these rows, which ensure that only the most informative and meaningful data is processed. This is an important ingredient of information security as it maintains the data integrity.

Another essential decrease is the lowercasing, which is based on the fact that every text data should be converted to lowercase in order to have the same standards throughout the dataset. The role of letter casing becomes so crucial that the position of letters can affect data processing to the extent that, for example, "Bitcoin" and "bitcoin" would be ruled as different words. Every text analysis software has a case normalization option. There are some reasons why people in their writing process prefer lowercasing. For instance, the practice can enable greater clarity by decreasing repetition and, therefore, aid better matching when analyzing the text.

More often than not, characters such as punctuations, symbols and hashtags do not give information regarding the sentiment and the themes. Hence, they need to be removed. As for the tweeting, sometimes removing hashtags may not be necessary especially where the aim is to examine the significance of hashtags as they primarily represent pertinent topics. Furthermore, removing these characters does affect in most cases the extraction of data from characters. Placing a premium on these can streamline the analysis, with the removal of more trifle words, thereby reducing the noise data.

Tokenization is the base of text analysis, since it means to divide the text into portions called tokens which can be words, phrases or even sentences. In such a process, a long sentence which is composed of characters is broken up into small units that are more convenient for processing. For instance, computation of the word term frequency, which is done in term frequency analysis or by extracting features for machine learning purposes., is an important application of this method. In this respect, eliminating website links and "@" signs make Twitter a universal language for all countries and cultures. g. The only manner for a Twitter platform to provide citation features to the users is putting their username (@username in tweets) because it does not serve the purpose of analysis of sentiments.

Non-English words removal for English-only datasets is a practice that helps to ensure the linguistic structure of the analysis stays consistent. The process of translating English contractions by expanding them has a number of aspects. First of all, it helps in standardization of text and creates a clearer understanding for text analysis tools which in its turn eliminates possible misinterpretation or treating contracted words as different words from the expanded forms. Next, lemmatization includes replacement of words to their base or lemmatized form after considering the contextual information and then putting the word in its meaningful base form. This is of great importance in the classification of the appropriate word forms and gives scope for deeper analysis of the significance of terms as compared to the context of the whole text.

4. Methodology

4.1 Research Questions

4.1.1 Research Question 1 :

How does sentiment expressed in Bitcoin-related tweets evolve over time, and what are the dominant sentiments observed?
The test methodology, the following steps were described: pandas was imported to work with the data, the path to the dataset was defined Bitcoin tweets, and an empty list to save DataFrames was also initialized. The dataset was read in chunks to save memory, and every chunk was added to the data owners` list. The next stage is preprocessing the DataFrame for sentiment analysis.

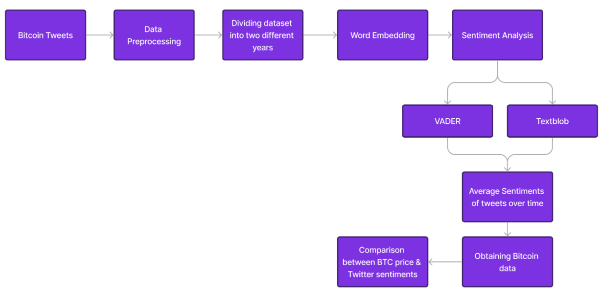


Fig 2. Data Flow Diagram

After testing, it was decided to disperse the dataset of tweets about Bitcoin. Specifically, there are two periods: 2021 and 2022. Converting the column date to data in datetime format but “cooking carefully” errors.

Each year is cut by its start and end dates, and finally, we create ‘ 2021′ and ‘ 2022′ to slice the dataset of ‘2021’ and ‘2022’. The first part of both datasets is shown to have an overview. This approach helps to focus on a two-year period to comparatively analyze the development of sentiment and find trends in the tasks of cryptocurrency discussion. The following are the steps of working with Bitcoin-related tweets: topic modeling and text tokenization, which often depend on special libraries, such as scikit-learn, Matplotlib, Seaborn, pandas, NLTK.

First of all, the tweets were structured using LDA topic modeling from the tool kit scikit-learn which is a simple tool for unsupervised learning in dimensionality. Our 'text' is containing for corpus that first which is then vectorized by CountVectorizer after which is transformed to a document-term matrix and is used for LDA finalization as an input. The model indicates that five topics based on tweets, tweets related to Topic 1 and Topic 5 are mostly used in messages. The areas where students are good are represented in the distribution graphs of the topics and the performances of subjects.

Additionally, NLTK will be used to tokenize. Due to the necessity to manage the memory efficiently, the tweets are processed in chunks using a generator function.

Next, the data cleaning and filtering process is done according to the given criteria followed. After that the filter is applied to extract rows containing particular emoji such as expansion, joy or precaution which relate to conveying the financial sentiment.

The dataframe Result, which is formed with target emojis, screens tweets having those emojis in them, reducing the data to the textual sentiments evoked by Bitcoin. Such data collection is essential in determining whether movements on the stock market are driven by impulsive sentiments and the manner in which investors make their decisions.

But, writing making low-dimensional data forms even more visible is demonstrated. Functionally, it behaves by selecting the ones which may have emojis or alphabetical letters and then removes all the symbols and totally unnecessary characters.

Hence, one aspect is done which is the emoji points get saved that express the financial emotions in the end point.
Thus, the mentioned dataset could be at the foundation of an inquiry regarding the changing dynamics in crypto discussions, including those evolutions that it undergoes over time, such as moods alteration or investors strifes.

The protagonist in this method is Word2Vec, which can likely process text analysis faster.

The next manifests in an extremely clear way, the moment that the DataFrame was loaded with our tokens after the lemmatization and then the text tokenization was in process for the Word2Vec building. Indubitably, data is tokenized and passed to the model to learn with the result that vectors which are to be called word embeddings (carrying the meaning relationships between the words) appear.

Attribute space is visualized with a Principal Component Analysis (PCA) procedure to reflect the original high-dimensional vectors in a simpler two- dimensional form. This PCA scatter plot gives an opportunity to understand word clustering in the semantic space and how far apart word embeddings are distributed. Besides that, the Word2Vec model is also prepared for subsequent utilization by storing it. In this way, the model can be used repeatedly and it can also be adapted to different tasks. Finally, the algorithm does the job of finding the words that are most alike with the given word ('bitcoin') and clustering word vectors using K Means which in turn identifies semantically related word clusters. The way of achieving this purpose is that users using this methodology can clearly analyze and understand the word relationships in text data, which makes it easy to use in such jobs as information retrieval, sentiment analysis, and document clustering. It continues the work of sentiment analysis and visualization using TextBlob. Firstly, the number of occurrences N of each word in the corpus is counted and then a bar graph is used to visualize the most frequent words among the N words that are chosen and depicted.

Text lemmatization takes place after this, while doing sentiment analysis with TextBlob. By calculating sentiment polarity score for every row text, it is visualized that the text is positive or negative. These scores are shown in the form of a plot on all dates, either individually or combined as the average score per date, through the line plot graphing.
The following histogram shows the distribution of the polarities (positive or negative attitudes) of the dataset. A score of valence is calculated for every tweet and polarities are further categorized as either positive, negative or neutral. The frequency of each category is represented in the form of a bar graph.

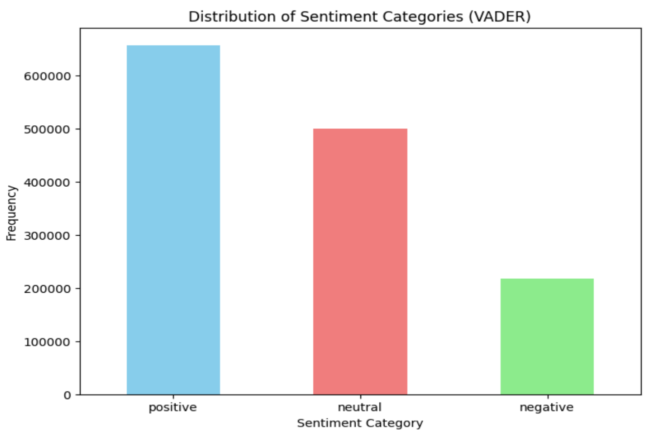
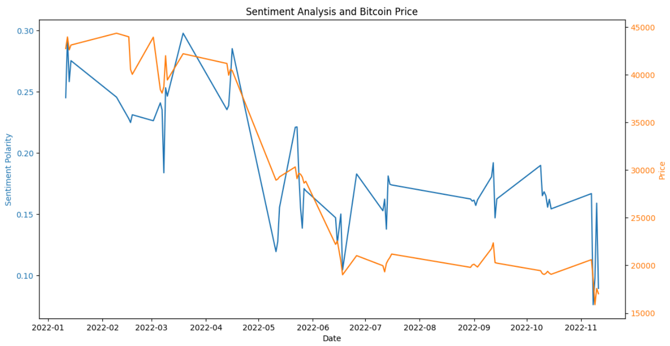


Fig 3. Sentiment Category

This analysis provides a knowledge into core emotions and their movement over time, thus, by means of such study we are able to comprehend and meticulously follow public opinion and emotion dynamics concerning the analyzed subject matter.

The next stage of this process would be to link Bitcoin price data to the code which was used to analyze the sentiment of the news. The two collected datasets allow to view this dependence in sentiment polarity and the Bitcoin average rate changes at different time periods.

The visualization is for displaying the real-time closing prices of Bitcoin, that are adjusted to the sensed period. And moreover, a picture showing the sentiment polarity and Bitcoin price together on the same graph will be provided for the purpose of the identification of any correlations or patterns thereof.



Correlation coefficient between sentiment and Bitcoin price: 0.11837

Fig 4. Price vs Sentiments for year 2022

The step following that is determination of correlation coefficient which reveals the extent and direction of the sentiment polarity and Bitcoin price association. The coefficient of correlation has the information whether there is the statistical significance of the association in terms of the sentiment extracted from textual data and the movement of bitcoin prices.

This survey provides more in-depth awareness of how emotion trends in the Bitcoin messaging can relate or be associated with variations in Bitcoin prices during a given period.

4.1.2 Research Question 2 :

Does sentiment expressed by influential users within the crypto community on Twitter have more influence on Bitcoin prices compared to the sentiment conveyed by regular crypto users on Twitter?

For our study, we sought to determine if the sentiment of users Tweeting in the crypto Twitter community that included influential personalities has a greater association with Bitcoin price than that from regular crypto Twitter users.

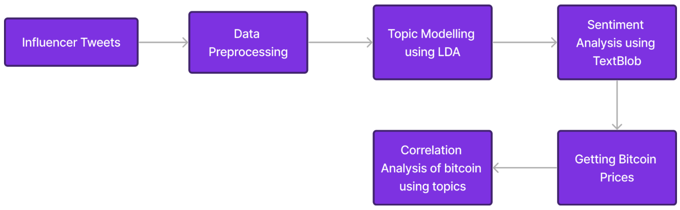


Fig 5. Data Flow Diagram

We gathered a comprehensive dataset containing the tweets of 52 influencers and encompasses tweets around 40 diverse cryptocurrencies.The dataset was obtained from the Apify Twitter API over the period of February 2021 till June 2023.

We preprocessed the dataset meticulously as part of our Data Cleaning. This endeavor involved natural language processing methodologies such as tokenizing the tweets which involves breaking down the text into words or phrases, removing common stop words to focus on more meaningful terms. Next, we utilized stemming which entails reducing words to their main form; and language detection which discarded non-English words and assured only English-tweeted observations. Subsequent to preprocessing, we used TextBlob, a Python library to conduct sentiment analysis of our preprocessed dataset using its popularly viewed methods. TextBlob facilitates API simplicity in performing common text processing tasks inclusive of calculation of the positivity/negativity of a text.

Additionally, to further our analysis, we utilized the Latent Dirichlet Allocation algorithm for topic modeling which is a statistical model that allows examining sets of observations to be interpreted as being created by several groups that explains why some parts of the data are comparative. In our case, we attempted to recognize the major topics that the tweets fall into. Therefore, LDA maps every tweet into one of our k topics.

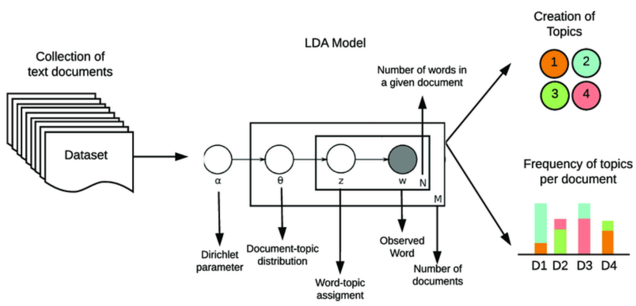


Fig 6. LDA Model

The above diagram illustrates the process of Latent Dirichlet Allocation (LDA). The core of the LDA model is characterized by several key components that interact to analyze the text.

At central there is a Dirichlet parameter (α), this setting indicates how topics spread and are distributed among documents. A higher alpha argument implies that a lower alpha document is probably to have a mixture of almost all topics, compared to that of the single topic. The term content diversity parameter becomes relevant with this function only. On this step the model thinks back to the Document-Topic Distribution (θ), which is a probability distribution that estimates the chance to find a given topic inside a particular document. This distribution is fundamental not only to allow us to gain the knowledge of the allocation of the topics within the entire corpus, as well as to comprehend the weight of each topic in the whole document. In this model, the topic assignment term (z) is a vector which is defined for each word in the document.
Another part of the process that the model utilizes is evaluation of each individual word and their distribution amongst other terms, where the frequency and co-occurrences are taken into account. The words, observed (w) in the processed documents, are analyzed within this structure. Topic model is the one which takes into account the occurring words and their designated topic assignments as already happened. The overall words (N) count and the amount of documents. (M) in the dataset influences the complexity and the depth of the analysis. Finally, the output of the LDA model includes the creation of topics, depicted here as four distinct topics derived from the analysis. Each topic is characterized by a group of words that frequently co-occur in the documents. Additionally, the model outputs the frequency of topics per document, illustrated by the bar charts, showing how much each document contributes to the topics in the corpus.

Following our analysis, we identified that topic 4 had a large correlation with Bitcoin pricing, with a correlation coefficient of 0.338748. This cluster contained mostly tweets about how tweets perceive the view of the general state of the crypto which is a marker of the market rather than any technical analysis or any position in any cryptocurrency.

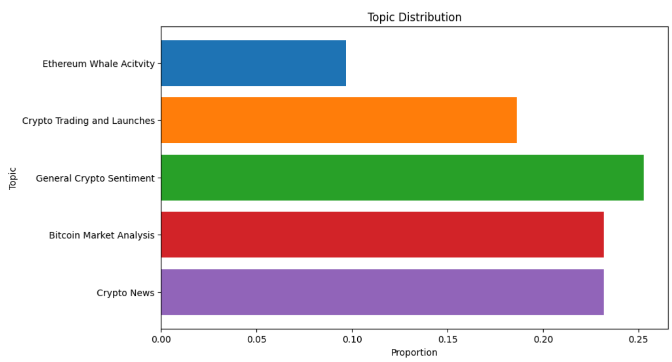


Fig 7. Topic Distribution

The conclusion of this empirical exploration will as well indicate that the strong influencers within the digital currencies community on platforms like Twitter do have the ability to shape market dynamics to instigate massive market movements. Indeed, their tweets usually cause extra trading moments where many investors follow the recommendations made by famous tweeters. Thus, some market players base their investment decisions on the intelligence and sensitivity of famous tweeters. As thus provides the basis to monitor influencer activities as they are an integral part of broader market analysis and many investment risk assessments. Besides that, the association between influencers' feelings toward Bitcoin and the changes in Bitcoin's prices requires a more profound examination into the concept of information diffusion within social networks. Having a grip on not only what kind of information influencers broadcast, but on the network effects and the fast pace with which followers are networked, is crucial. They can escalate an effect that sentiments have on market prices; so the very often the market prices drop sharply immediately after any impactful tweets.

Summing up in the end, the hypothesis of how historical sentiment data found on the Yahoo Finance Bitcoin and the perceived influence of the influencers on the general crypto sentiment may be partially correlated with the price of Bitcoin was proved in this project. Consequently, the lessening of this kind of trade can be attributed to the small impact of Tweets in the market through the stop and go method.

To enhance our analysis, future research could incorporate machine learning models to predict price fluctuations based on sentiment indicators derived from influencer tweets.
By associating the obtained topic with the historical Bitcoin Pricing from Yahoo Finance, we established a modest yet significant correlation between general crypto sentiment, mainly by influencers, and the price of Bitcoin. Therefore, this could mean that Tweets by influencers have a slightly more significant effect on the price of Bitcoin via the stop and go action of the market.

To enhance our analysis, future research could incorporate machine learning models to predict price fluctuations based on sentiment indicators derived from influencer tweets.

These models could use historical data to learn patterns of influence and sentiment spread, potentially offering more accurate predictions of market responses to new information. Additionally, investigating the role of automated trading systems in this context could provide insights into how these systems interpret and react to sentiments expressed on social media. Many trading algorithms now incorporate news and social media feeds to make split-second decisions, which can exacerbate market volatility when reacting simultaneously to influential social media posts.

4.1.3 Research Question 3 :

How do the Topic weights discussed in bitcoin forum overtime correlate with bitcoin market price ?

In this question, we aim to investigate the extent to which fluctuations in Bitcoin Forum's topic weights correlate with changes in Bitcoin market prices. To answer this question, The Bitcoin forum dataset was used since Bitcoin developers and core members have discussed the development of Bitcoin since its initial stages in this forum, this community can be considered as a representative community among various Bitcoin communities.

Initially, the Bitcoin forum dataset was subjected to rigorous preprocessing to refine the textual data and ensure its suitability for subsequent analysis. Simultaneously, historical Bitcoin market prices corresponding to the forum data dates were obtained, serving as the target variable for correlation analysis.

To do this analysis we needed to extract top topics discussed across the bitcoin forum to correlate with the frequent topic discussion with the market price of Bitcoin. To do this, topic modeling was employed, a ML technique for unsupervised learning that is used to extract coarse-grained and abstract topics that commonly occur in document corpus. This is used to uncover the hidden semantic patterns in a given corpus of text using statistical formulations to identify groups of similar words in a corpus.

For this specific analysis, Non-Negative Matrix Factorisation[NMF] for topic modeling was utilized, but before topic modeling some prerequisites like construction of Term Document Matrix(TDM) had to be satisfied for NMF. Therefore the subsequent steps involved creation of a list of unique words from the preprocessed text data to facilitate the construction of a vocabulary, essential for representing textual data numerically. Each unique word was mapped to an index using a dictionary, thereby enabling the transformation of textual data into a numerical format necessary for Non-Negative Matrix Factorization (NMF), a crucial step in our methodology. The construction of a Term Document Matrix (TDM) followed, wherein rows represented unique words, columns represented documents (forum posts), and cell values denoted the frequency of words within documents.

$$\mathrm{TDM}[i, j] = \mathrm{Frequency}(w_i, d_j), \quad \forall w_i \in V_{\mathrm{voca}}, \forall d_j \in D_{\mathrm{forum}}$$

Here, mathematically each entry is denoted ad TDMi, j, signifying the frequency of the ith word in jth document, The frequency is calculated using the function Frequency(w*i*, d*j*) where w*i* represents a unique word and d*j* signifies a specific document. These operations are performed for every unique word w*i* in the vocabulary V_{voca} and for every document in the set of forum documents.

Incorporating temporal information into the TDM involved associating each document with its corresponding date, thereby enhancing the analytical capability of the dataset. Normalization of the TDM was done to ensure that each document's vector length did not disproportionately influence the analysis, thus maintaining fairness across observations.

Leveraging NMF for topic modeling, the TDM was decomposed into two lower-dimensional matrices: W (document-topic) and H (topic-term), yielding insights into the underlying topics within the forum data, shedding light on the latent topics underlying forum discussions. Mathematically, NMF can be formulated as: TDM≈W×H.

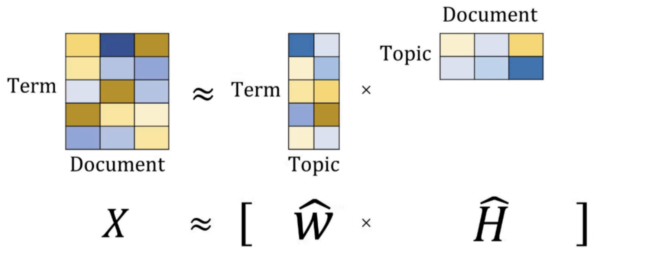


Fig 8. NMF

The Coordinate Descent Algorithm, employed for updating the factor matrices W and H in Non-negative Matrix Factorization (NMF), iteratively refines these matrices to minimize the disparity between the original data matrix and its estimated approximation.

Using NMF, top 10 topics were extracted from the bitcoin forum posts to check the correlation with the bitcoin prices. By correlating the extracted topic weights with dates and Bitcoin market prices, we aimed to find the relationship between forum discussions and bitcoin market price. Notably, our analysis revealed significant correlations between specific topics and Bitcoin prices, as evidenced by high correlation coefficients and exceptionally low p-values, signifying statistical significance.

Further exploration of topic weights over time, particularly in the year 2018, elucidated trends in forum discussions and their potential impact on Bitcoin market dynamics. Visualization techniques, such as line charts, provided intuitive representations of topic prevalence and their alignment with Bitcoin price movements, offering valuable insights for interpretation.

4.1.4 Research Question 4 :

To what extent do sentiments expressed in Bitcoin Forum affect Bitcoin market prices?

In this question, our purpose is to study how sentiments expressed over bitcoin forum affect the prices of Bitcoin in the market.

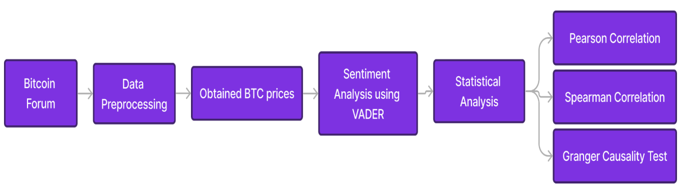


Fig 9. Data Flow Diagram

Fig 4.1.4 depicts the data flow diagram for the sentiment analysis of posts and comments on Bitcoin Forum. The Bitcoin Forum dataset consisted of 36,898 posts and 353,679 comments authored by 32,304 users, ranging from the forum's inception on November 22, 2009, to February 2, 2018.

In pre-processing, we dropped null values to improve the integrity of data and performed stopword removal, tokenization, and lemmatization.

To do this analysis, we do sentiment analysis using VADER (Valence Aware Dictionary and sEntiment Reasoner). VADER is specifically suited for the sentiment analysis of social media content, including slangs, emoticons, and emojis. Sentiment scores are calculated for each post and based on the scores, posts are categorized as “positive”, “negative”, or “neutral”. VADER uses a lexicon-based approach, where each word in the lexicon is assigned a sentiment score based on the category of sentiment (positive, neutral, negative). We have evaluated the distribution of these sentiments over the years and measured the correlation between sentiment scores and Bitcoin prices using various statistical methods using Pearson Correlation, Spearman correlation coefficient, and Granger Causality Test. These statistical method provide valuable insights into the relationship between the sentiments expressed in Bitcoin Forum dataset and the c

4.1.5 Research Question 5 :

How does the sentiment expressed in tweets about Ethereum and Binance Coin by crypto influencers affect the value of Bitcoin in the market?

The main objective of our research study here is to determine the nature of the link between the potential impact of social media influencers who tweet about Ethereum and Binance Coin on the market dynamics of Bitcoin. As the first step in our analysis, we made sure that the data set we worked with was wide enough and comprised over 70,000 tweets related to Ethereum and 30,000 tweets on Binance Coin published by influencers who are always on Twitter.

This data was obtained and thereafter we performed extensive data cleaning and filtering to remove unnecessary information and increase the consistency in the dataset. Following the excel of data cleaning of the dataset, the process of stopword removal was applied, in order to omit the common words which do not carry any substantial meaning. The second NLP technique applied was text normalization technique, Lemmatization, which helped to standardize the text and remove the word family occurrences from dimensionality of the feature space.

After topic modeling was complete, Latent Dirichlet Allocation (LDA) was used to do the topic modeling section. This is good for knowing the themes which are existing behind the dataset and why they are able to influence our collective mindsets. For both the Ethereum (ETH) and Binance Coin (BNB) datasets the top 5 topics were deemed to be the most important ones. These items, to say nothing of the general areas, serve, though, as indicators of early market trends and shifts. The next step was to introduce the analysis of this research subject in terms of their popularity in various time-periods. The average topic distribution was calculated for each the given date. Thus, all of the data was aggregated and was further grouped by a date to shape the visualization of each subject`s dominance over the timeline. The given graphs display the fluctuations of the primary topics regarding the event of Ethereum and Binance Coin. This way, the time changing impact of the cryptocurrency market is visualized.

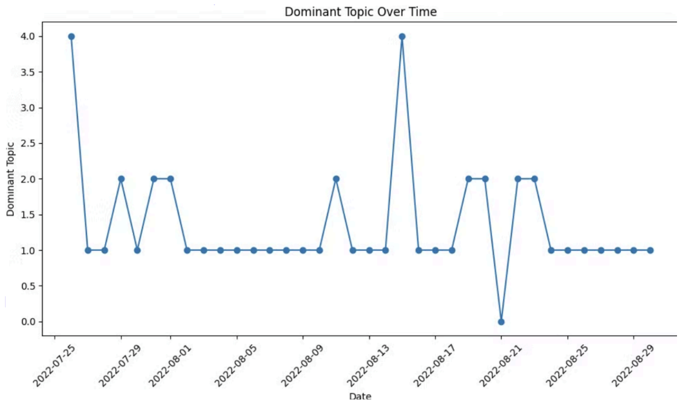


Fig 10. Ethereum Topic Dominance

Majority topic (topic 1) of the Twitter posts was shown as prevalent by the data collected within the period of time. We saw keywords 'ethereum', 'cryptonews', 'blockchain', 'bitcoin' and 'crypto' frequent in the topic 1 of the Ethereum dataset. Also, Binance Coin dataset was subjected to topic modeling like that of previously mentioned example, thus rendering the same outcomes. Another substantial issue that arises from the topic modeling of both the Ethereum and Binance Coin datasets is related to Bitcoin and Bitcoin-related topics being frequently assigned to a certain topic. Thus there is a correlation that seems to work between these other lesser size currencies and Bitcoin.

And we proceed to deepen our analysis of this interdependence which is Sentiment Analysis research part of our work. The lemmatized text, which is obtained after the process of pre-processing, becomes the model' input for the sentiment analysis module. Using the Valence Aware Dictionary and Sentiment Reasoner (VADER) that is a pre-trained lexicon-based sentiment analysis model, helped us to determine the sentiment in the @twitter's tweets with precision. This specific model architecture was used to make the sentimental polarity of each tweet in our dataset.

VADER is implemented via the NLTK library and proves to be a powerful default model suitable for sentiment detection. One advantage of the method relies on the use of the extensive dataset that consists of tweets, so it is definitely a perfect choice for our project. The model's accuracy in comprehending social media text reinforces the applicability and validity of the model in determining the sentiment patterns in our dataset that has tweets regarding cryptocurrencies.

The VADER sentiment model begins with input sentences, consisting of text data such as social media, product reviews, or news articles to analyze sentiments. It determines the punctuation marks including periods, commas, exclamation marks, and question marks which denote feelings or focus within a text.

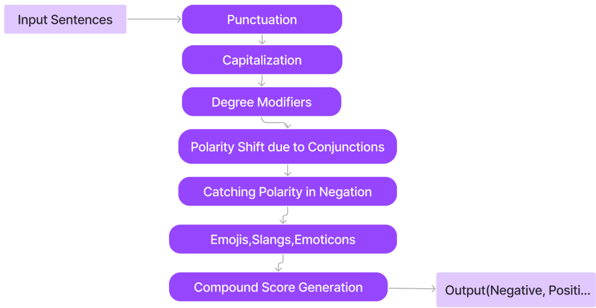


Fig 11. Model Architecture of VADER

Capitalizing usually helps to distinguish a loud or emphatic tone which, indirectly, includes the mood of speech. Degree modifiers add intensity to the existing sentiment or minimize it, whereas conjunctions are used in polarity shifts. Polarity negation through words such as 'not' or 'never' is identified that is to be properly recognized to identify sentiment. Not only emojis, slangs and emoticons but also compound scores form the basis of sentiment analysis. In this regard, the model ends up computing the sentiment scores of negative, positive, and neutral moods and users can use them to assess the sentiment of the text.

The individual sentence's sentiment scores are normalized, and they are aggregated to get the overall sentiment score for text. This number encodes the general sentiment of the text, where -1 means very negative and 1 stands for very positive. Through the sentiment score, VADER is able to classify the sentiment of the given text as one of the categories, including positive, negative, or neutral. It does this by having a compound score too, which is a normalized, weighted composite score of sentiment that is calculated by summing up the valence scores of each word in the lexicon that are weighted according to the rules..

The next thing we tried was to find out the relationship of Ethereum and Binance Coin sentiment with the price of Bitcoin. Here we depicted the correlation between these sentiments and Bitcoin prices over time. Bitcoin price data was extracted by using the Yahoo Finance library, which provided us with a trusted source of historical Bitcoin prices.

Thereafter, the emotions extracted from ETH and BNB tweets were plotted together with the Bitcoin prices using the Matplotlib library in Python. This representation through the graph enabled researchers to examine visually, if there were any conspicuous patterns or correlations between cryptocurrency sentiments and Bitcoin prices over time.

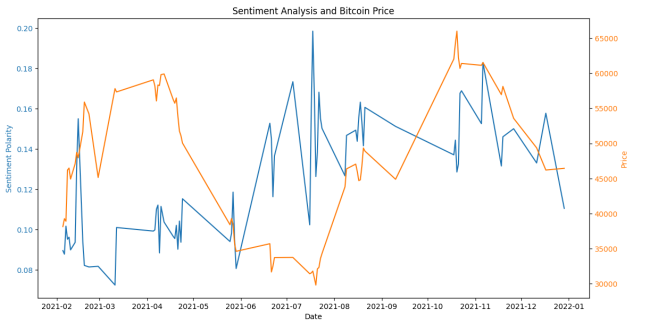
5. Results

Research Question 1

The provided visualizations display a comparison of Bitcoin prices and sentiment scores for the year 2021 and 2022. Each graph represents the fluctuations in both Bitcoin's price value and sentiment scores within distinct periods.

Year (2021)

During the year 2021 it was observed that the graph for Bitcoin closed price and Sentiment Analysis using Vedar and Text blob does not show an uniform increase in the graph. Both the plots were almost independent and showed negligible correlation with each other.



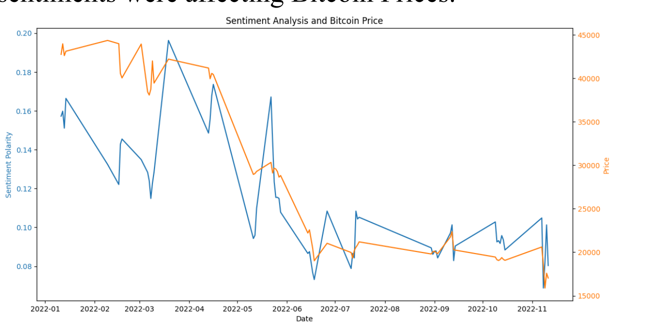
Correlation coefficient between sentiment and Bitcoin price: 0.0004906

Fig 12. Year 2021 Analysis

Both the models TextBlob and Vedar showed the same trend. From this we can conclude The overall user base's sentiment trends reveal slow shifts that generally correspond with the price of Bitcoin, although with less noticeable peaks and troughs. This implies that public opinion may not be as predictive as it is reactive to changes in prices. Longer-term trends and more general market sentiment can be better understood by analyzing user sentiment.

Year(2022)

During the year 2022 it was observed that the graph for Bitcoin closed price and Sentiment Analysis using Vedar and Text blob showed an uniform decrease in the graph. Both the plots were dependent and showed a healthy correlation with each other.We can also conclude that user sentiments were affecting Bitcoin Prices.



Correlation coefficient between sentiment and Bitcoin price: 0.11044

Fig 13. Year 2022 Analysis

We can conclude that Across a broader user base the opinion trends reflect a gradual evolving indicating the Bitcoin price fluctuations but without the peaks and troughs that are so recognizable. This even means that the public opinion cannot be regarded as a reliable basis to predict the price movements but it is rather reacting it to the changes in prices.

Medium to long-term trends and overall market sentiment as a whole can be well interpreted as per user sentiment.

Research Question 2

The below results demonstrate how bitcoin prices and the sentiment polarity of the year 2021 of the topic 4 correlate with each other in different quarters of the year. Each section of the graph depicts one quarter of the year. Topic 4 had the highest correlation with the bitcoin prices.

Topic 4 (2021)

During the first quarter, Bitcoin prices saw a significant increase from about $30,000 to just over $60,000, while the sentiment scores demonstrated high volatility but generally trended positively, suggesting an optimistic public perception as prices rose. In the second quarter, both Bitcoin prices and sentiment scores showed more fluctuation, with prices peaking and then experiencing a sharp decline by the end of June, mirrored by a decline in sentiment, indicating that the sentiment may be reactive to price movements. The third quarter showed a recovery in Bitcoin prices which did not fully recover to previous highs, and sentiment scores were less volatile but remained generally positive, pointing towards a cautiously optimistic outlook despite less price stability. Finally, in the fourth quarter, Bitcoin prices stabilized somewhat around $50,000, but sentiment scores varied, showing sharp declines at times, which could reflect mixed public reactions to the market's uncertainty. This analysis suggests that public sentiment on "Topic 4" is closely aligned with Bitcoin's price movements, potentially influencing or reflecting public confidence in the cryptocurrency's value throughout 2021.

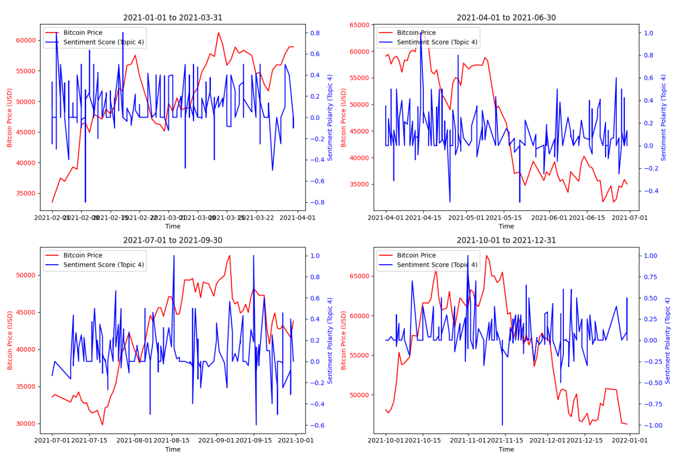


Fig 14. Topic 4 (2021) analysis

Topic 4 (2022)

In 2022, the displayed data illustrates the correlation between Bitcoin prices and the sentiment scores for "Topic 4" over four quarters. During the major quarter, there was a great deal of proving to be the case with the BTC's rate, starting with the figure of $46,000 at the top followed by $48,000, and then it was sharply down to the level of $38,000 before the BTC was remarkably up $44,000. The sentiment analysis is similarly to the price charts first with positive scores at zero. 8, falling to -0. 4 In the initial quarter, the sales volume was significantly high when the product was discounted, and then it decreased markedly at the end of the second quarter but bounced back to an all-time high at the third quarter. Through the retraced background goes the similar traits during the second quarter and its subsequent quarters where the price of Bitcoin and the sentiment score remain unchanged thus implying that as the magnitude of the swings in price heightens there exists very high market volatility coupled with shifting public opinion. By the third quarter, the PoB showed rather the suppressing price of its crypto currency, and the change was less volatile, but the fair decline was reflected in the negative sentiment scores, which were estimated to -1.

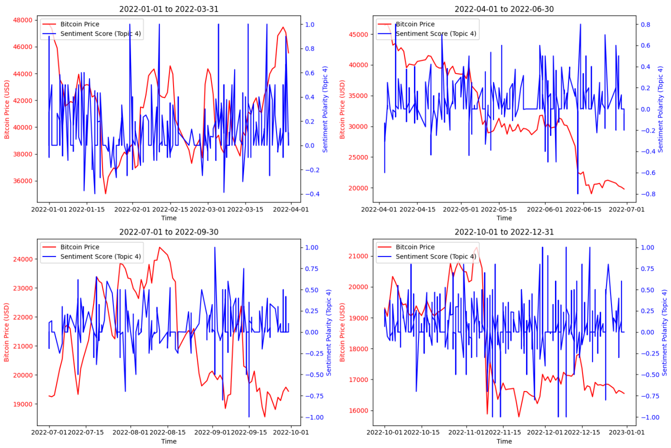


Fig 15. Topic 4 (2022) analysis

Topic 4 (2023)

The infographics shown below are depicting the bitcoin market growth in 2023 together with the polarity of "Topic 4" sentiment among the public at the same time of two halves. Over the first 6 months of this year, the value of bitcoin illustrated a trend to the path of a raising line. After a few small fluctuations they reached a high at about $24,000 by mid-February, but overall they continued to rise.
As prices increased, the sentiment polarity improved from a low starting point, suggesting that favorable price changes may have contributed to an increase in public sentiment. The mood scores, however, showed notable fluctuation despite the growing prices, indicating the erratic nature of the market.

The charts of the Bitcoin market displayed that the curve in 2023 was either upslope or downslope. Besides, the topic 4 sentiment polarity was recorded daily on the same dates that ranged from 1st to the 12th of January of two halves. In the first half of this year, following the prices of bitcoin seemed to be the scenery of expansion as well. There was a minor correction period that lasted until mid-February after which there was an upturn that got too close to $24,000; however, prices at times went higher as the prices continued to increase.

By the year's end, we hit the second half of 2023, and the charts are revealing that this sliding movement hasn't come to an end. One thing only stands stable and is gently rising - price. It was a volatile month for the cryptocurrency and at the end of May the exchange rate was $30,500, an increase from the end of April where it was around $27,000. Things simplified to end with a fall from the top, which was sudden and completely unpredictable. The price indicated a fall immediately to the $28.000 to $29.000 levels.

The sentiment polarity during this period remained highly variable, with sharp spikes both positive and negative, suggesting a reactive sentiment environment heavily influenced by price movements. Overall, the year showed that while Bitcoin prices had periods of recovery and growth, the sentiment remained uncertain, highlighting the complex dynamics between market behavior and public sentiment within the cryptocurrency space.

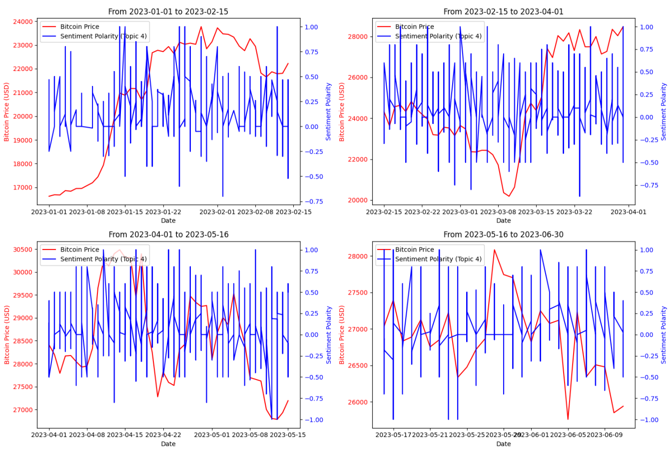


Fig 16. Topic 4 (2023) analysis

The correlation coefficients for topics 0 through 3 alongside topic 4 show varying degrees of relationship between sentiment polarity and Bitcoin price movements within different thematic discussions. Topic 0, with a coefficient of -0.148296, and Topic 3, at -0.158206, both indicate a slight negative correlation, suggesting that as sentiment within these topics becomes more negative, Bitcoin prices tend to decrease, albeit weakly. Topic 2 shows a more substantial negative correlation at -0.287846, indicating a stronger inverse relationship between sentiment polarity and Bitcoin price movements within this topic. Conversely, Topic 1 presents a small positive correlation of 0.083005, implying that more positive sentiment on this topic is slightly associated with an increase in Bitcoin prices. Topic 4 stands out with the highest positive correlation of 0.338748, indicating a more significant and direct relationship between sentiment positivity and rising Bitcoin prices. The coefficients provide a comparative summarization of how discrete thematic valences co-depend with BTC's market trajectory.

On various thematic cases that concern the space of cryptocurrency we can see how certain topics affect Bitcoin's market behavior. Matrices that tend to be connected with Ethereum whale activity and general crypto sentiment usually occur in the periods that there is a low in the Bitcoin price, possibly showing the reason which might be uncertainties in the market or a change in investors. However, some conversations that are done for the sake of seeing Bitcoin's own market behaviors end up imitating the patterns of the price movement, possibly that Bitcoin's market behaviors could affect the increase in the fiat currency valuation. On the other hand, with the development of new tokens and trading platforms, it seems that the investors' capital is about to be diverted away from the Bitcoin system, which, in a certain way, is indicative of the cut-throat nature of the crypto market, when newly launched projects are given a temporary privilege over the established profits-makers, such as the Bitcoin system. Such discoveries emphasize in turn the tangled link between chattering about crypto targeted at Bitcoin and the currency's trading rate that show the divergent influences of the investor concentration and sentiment on its costing.

Research Question 3

It made some intriguing discoveries about the range of topics discussed, the relationship between these topic weights and bitcoin market values, and the variety of issues discussed in the bitcoin community by using Non-negative Matrix Factorization (NMF) for topic modeling.

Exploration of the topic co-occurrence network revealed distinct user groups within the community, delineated by their discussions' thematic focus. These groups encompassed discussions related to market sentiment and trading strategies, developer-centric topics concerning mining activities and network operations, and conversations surrounding Bitcoin's global currency status and adoption trends.

Firstly, discussions pertaining to market sentiment and trading strategies were characterized by terms such as "market," "back," "sell," "low," "hold," and "buy." Secondly, the developer community engaged in discussions primarily revolving around mining activities and network operations, featuring terms like "fee," "long," "day," "network," "transaction," and "block." Lastly, discussions concerning Bitcoin as a global currency and its adoption trends were underscored by terms such as "Bitcoin," "value," "currency," "start," "country," and "world."Through these clusters, it became evident that the Bitcoin community engages in multifaceted discussions encompassing various aspects of Bitcoin's ecosystem and its broader implications. Further our analysis with these topics and market price showed that Topic 6 was exhibiting the most significant positive correlation coefficient (0.5813) with the market price of Bitcoin. This topic included words like "price", "bitcoin", "coin", "buy", "exchange", "money", "market", "sell", "crypto" ", "currency", invest "," "high" "big" "government".

This finding suggests that as Topic 6 gains prevalence in Bitcoin forum discussions, a corresponding increase in Bitcoin's market price is observed. The statistical significance of this correlation was affirmed by an exceptionally low p-value of 1.67e-88, indicating the rarity of such a strong correlation occurring by random chance alone.

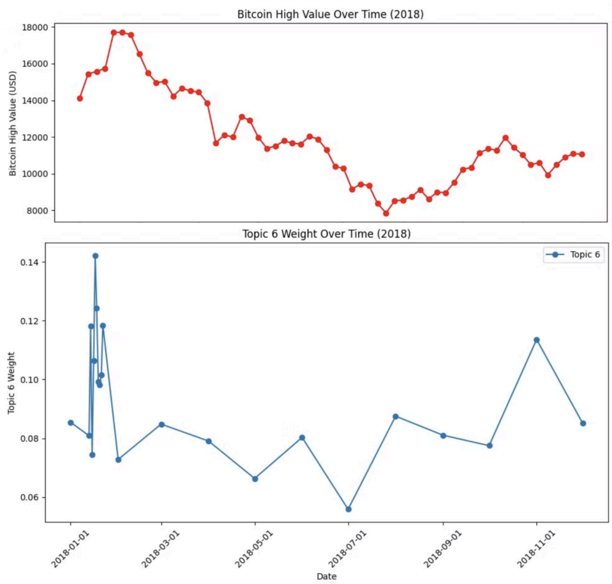


Fig 17. Topic weight 6 vs bitcoin Price in 2018

The figure showcases the correlation between bitcoin price and topic 6 weight in 2018, it can be visualized that from january to february there was a hike in the topic discussion which also increased the market value of bitcoin. Similarly from July to August it can be seen that a low dip in topic discussion showed a dip in market price too.
In the below image it can be seen that as the topic weight increased from 2017 may to 2018 January the market price also started surging together.

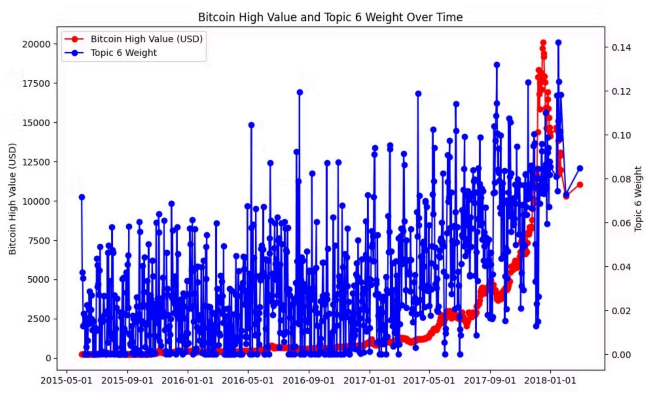


Fig 18. Topic weight 6 vs bitcoin Price from 2015 to 2018

Conversely, Topic 3 demonstrated a negative correlation (-0.3954) with Bitcoin market prices. This negative correlation implies that as Topic 3 becomes more prevalent in forum discussions, the market price of Bitcoin tends to decrease. The associated p-value of 1.51e-37 further supported the statistical significance of this negative correlation.

This analysis indicates a statistically significant positive correlation between the prevalence of Topic 6 in Bitcoin forum discussions and an increase in Bitcoin's market price, suggesting that Topic 6 discussions might predict or drive market behavior. Conversely, Topic 3 shows a significant negative correlation with Bitcoin prices, indicating that increases in discussions around this topic could be predictive of a decrease in market value. These findings highlight the potential impact of specific conversational topics on Bitcoin's market dynamics.

Research Question 4

On performing sentiment analysis using VADER, we found that the most prevalent content in the forum was positive, followed by negative sentiment posts and comments, and neutral posts.

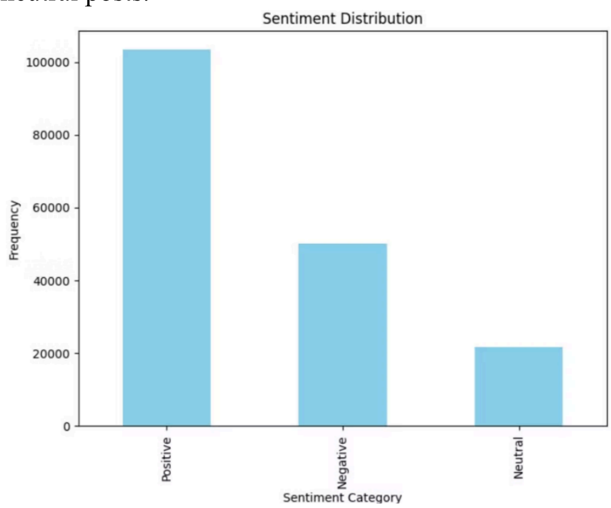


Fig 19. Sentiment Distribution of Bitcoin Forum

We visualized the correlation between sentiment and bitcoin price (btc_price) using a correlation heatmap. This color coded matrix is a visual representation of closely related two variables. Darker colors indicate higher correlation values. Lighter colors indicate lower or no correlation. The diagonal from top left to bottom right has a value of 1, indicating perfect positive correlations between Sentiment-Sentiment and btc_price-btc_price pairs.

The off-diagonal elements have a value of 0.042, indicating a very weak positive correlation between Sentiment and btc_price. This means that as sentiment scores increase (or decrease), there is a slight tendency for bitcoin price to move in the same direction.

We measure the Spearman coefficient, often denoted by the symbol ϱρ (rho), and Pearson correlation coefficient (often denoted as "r") between two variables: Spearman coefficient and Pearson coefficient.

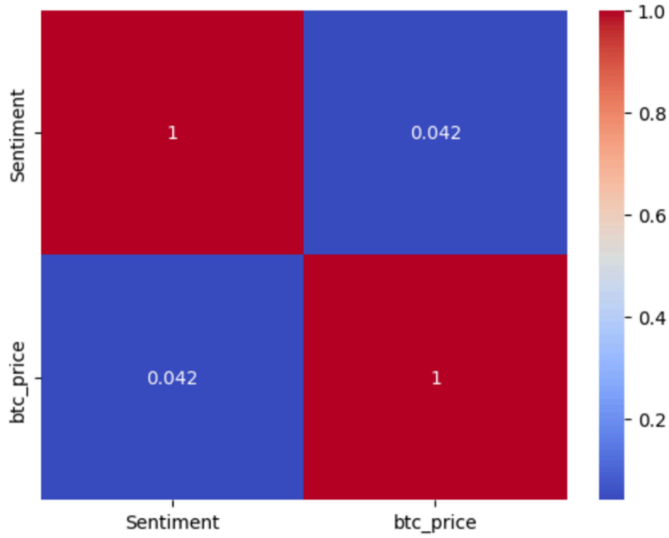


Fig 20. Correlation Heatmap

Spearman Coefficient: 0.05
Pearson Coefficient: 0.04

These values indicate weak positive correlation between sentiment and Bitcoin prices.

We also performed Granger causality test, a statistical test to determine whether one variable could be potentially useful in predicting another. It is commonly used in econometrics and time series analysis to assess the causal relationship between two variables.

We measure the F statistic and p-value for 5 lags.

Lag 1: F-statistic = 140.58875917442518, p-value = 2.0364960849911447e-32

Lag 2: F-statistic = 50.518621443612325, p-value = 1.1651064555691996e-22

Lag 3: F-statistic = 28.724516860294003, p-value = 1.4680092670491898e-18

Lag 4: F-statistic = 20.150726545524442, p-value = 1.309474108584644e-16
Lag 5: F-statistic = 16.27263417230707, p-value = 4.3892798446336113e-16

The Granger causality test results suggest that past values of 'Sentiment' contain information that helps predict current and future values of 'btc_price', and vice versa.

Research Question 5

The sentiment graph plotted against Bitcoin prices over time provides insights into the potential relationship between market dynamics of Bitcoin and sentiments associated with Ethereum and Binance Coin.

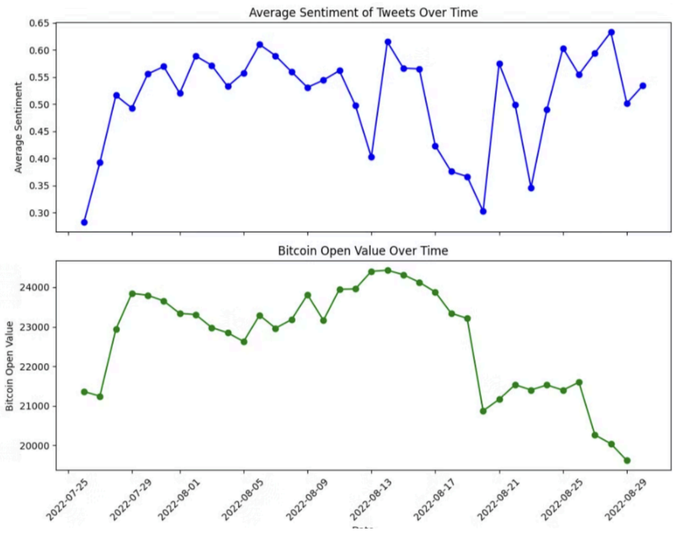


Fig 21. Ethereum Sentiment v/s Bitcoin Prices

From the graph, it's apparent that whenever Ethereum sentiments are positive, Bitcoin prices tend to rise, indicating a positive correlation between Ethereum sentiment and Bitcoin price. To validate this correlation, we computed both the Pearson correlation coefficient and the Spearman correlation coefficient between the open value of Bitcoin and the average sentiment of Ethereum.

Spearman Coefficient: 0.05

Pearson Coefficient: 0.14

The Spearman coefficient, calculated to be 0.05, suggests a weak positive correlation, while the Pearson coefficient, calculated as 0.14, indicates a slightly stronger positive correlation. Although both coefficients confirm a positive relationship between Ethereum sentiment and Bitcoin price, they also imply that this correlation is relatively weak. Despite the positive correlation, it's essential to recognize that other factors may influence Bitcoin prices, and sentiment expressed in Ethereum tweets is just one of many variables affecting market dynamics.

Similar graph was plotted for the average Binance Coin sentiment against the Bitcoin Prices to find if Binance Coin sentiments have any effect on the market behavior of bitcoin or not.

The relationship between Binance Coin (BNB) sentiments and the Bitcoin prices as shown in the analysis is not strong. While analyzing sentiments from Binance Coin associated tweets and plotting them against Bitcoin prices, I cannot find strong evidence of a direct connection between Binance Coin sentiment and behavior of the Bitcoin market. To verify this relationship we computed both the Pearson correlation coefficient and the Spearman correlation coefficient between the open value of Bitcoin and the median sentiment of Binance Coin.

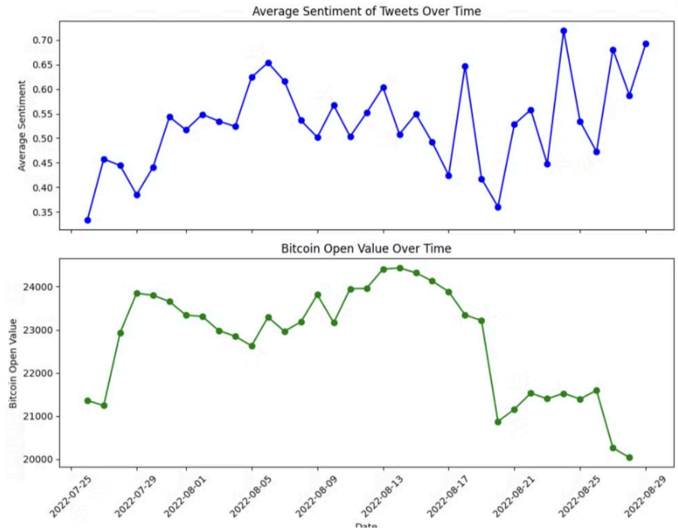


Fig 22. Ethereum Sentiment v/s Bitcoin Prices

Spearman Coefficient: -0.01

Pearson Coefficient: 0.06

The coefficients of sentiment show the weak relationship between the sentiment towards Binance Coin and the price of Bitcoin in the given period - Spearman coefficient around zero indicates no monotonic relationship whilst the Pearson coefficient 0 suggests similar 0.06 depict a weak positive relationship or association. Nevertheless, all of these coefficients are still within a range where the positive relationship between Binance Coin sentiment and Bitcoin prices is not strongly declared.

These findings corroborate the earlier observation that Binance Coin sentiments do not appear to have a significant impact on Bitcoin market behavior. Despite fluctuations in Binance Coin sentiment, the correlation coefficients suggest that other factors may primarily drive Bitcoin price movements.

It is observed in the analysis, Ethereum sentiment has a higher correlation with Bitcoin prices than the correlation level of Binance coin sentiment. This suggests that the popularity of Ethereum can be that it is already well-established as a major player in the crypto market and it follows Bitcoin, which is the leader by its market capitalization. But at present, the BNB is under the position of developing into a highly stable cryptocurrency.

Applying that Bitcoin market capitalization exceeds Ethereum's by nearly fivefold and Binance Coin's by tenfold, we can see that Bitcoin's price is less dependent on sentiment towards other cryptocurrencies than these two. The reason why Bitcoin prices can affect the prices of other cryptocurrencies, including Ethereum and Binance Coin, is not due to these movements in prices.

6. Conclusion

The goal of our project was to study the underlying trends of sentiments expressed on various platforms by different communities of users. We studied the impact of sentiment on the bitcoin price market. Through a thorough evaluation of the data and research, we conclude that although public sentiment plays a role in market price of Bitcoin, the effect is minimal. Future research could include improving our analysis by incorporating more factors such as geographical locations, socio-economic conditions, and such.

References

[1] Burnie Andrew and Yilmaz Emine 2019Social media and bitcoin metrics: which words matterR. Soc. Open Sci.6191068

[2] Albanese, F., Pinto, S., Semeshenko, V., & Balenzuela, P. (2020). Analyzing mass media influence using natural language processing and time series analysis. Journal of Physics. Complexity, 1(2), 025005. https://doi.org/10.1088/2632-072x/ab8784

[3] Kang, K., Choo, J., & Kim, Y. B. (2019). Whose opinion matters? Analyzing relationships between Bitcoin prices and user groups in online community. Social Science Computer Review, 38(6), 686–702. https://doi.org/10.1177/0894439319840716

[4] Di Tollo, G., Andria, J., & Filograsso, G. (2023). The Predictive Power of Social Media Sentiment: Evidence from Cryptocurrencies and Stock Markets Using NLP and Stochastic ANNs. Mathematics, 11(16), 3441. https://doi.org/10.3390/math11163441

[5] Phillips, R. C., & Gorse, D. (2017). Predicting cryptocurrency price bubbles using social media data and epidemic modelling. IEEE Xplore. https://doi.org/10.1109/ssci.2017.8280809

[6] Valencia, F., Gómez-Espinosa, A., & Valdés-Aguirre, B. (2019). Price movement prediction of cryptocurrencies using sentiment analysis and machine learning. Entropy, 21(6), 589. https://doi.org/10.3390/e21060589

[7] Analysis on relationship between bitcoin price trend and sentiment of bitcoin related tweets by ML and NLP. (2021, December 1). VDE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/9736780

[8] Predicting Fluctuations in Cryptocurrencies' Price using users' Comments and Real-time Prices. (2018, August 1). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/8748792

[9] Seroyizhko, P., Zhexenova, Z., Shafiq, M. Z., Merizzi, F., Galassi, A., & Ruggeri, F. (2022). A Sentiment and Emotion Annotated Dataset for Bitcoin Price Forecasting Based on Reddit Posts. IEEE Xplore. https://doi.org/10.18653/v1/2022.finnlp-1.27

[10] Mai, F., Shan, Z., Bai, Q., Wang, X., & Chiang, R. H. L. (2018). How does social media impact bitcoin value? A test of the silent majority hypothesis. Journal of Management Information Systems, 35(1), 19–52. https://doi.org/10.1080/07421222.2018.1440774