VILNIUS GEDIMINAS TECHNICAL UNIVERSITY

FACULTY OF BUSINESS AND MANAGEMENT

DEPARTMENT OF FINANCE ENGINEERING

APPLICATION OF TEXTUAL ANALYSIS FOR FINANCIAL INFORMATION

Behavioral Finance

Homework

Author: Romain Didier Taugourdeau, Flfcu-20

Lecturer: Dr. Nijolė Maknickienė

VILNIUS 2024

TABLE OF CONTENTS

INTRODUCTION

Cryptocurrencies Bull Market. In my project for the Behavioral Finance course at Vilnius Gediminas Technical University, I explore how social media sentiment can predict cryptocurrency market trends, focusing on Bitcoin. This study combines technical analysis of Bitcoin's price movements with an analysis of social media sentiments to identify correlations between public sentiment and market behavior. By analyzing tweets, I aim to understand the predictive power of social media on cryptocurrency markets, contributing to the field of behavioral finance by highlighting the influence of digital sentiment on market dynamics.

Problem - Is There Currently An Evidence of a Bull Run, and Can Social Media Sentiments Predict Such Market Phases ?

Research object - Evaluating the Current State of Cryptocurrency Markets

Aim - Show that Social Media Sentiments can Predict such Market Phases to a certain extent

Tasks:

1. Decoding Bitcoin's Trend: A Technical Analysis
2. Sentiment Spectrum: A Deep Dive into Twitter Emotions
3. Digital Echoes: Linking Cryptocurrency Movements to Social Media Waves

Research Methods.

- **Data Collection**: Gathered social media and Bitcoin market data for the selected period.
- **Technical Analysis**: Used indicators like moving averages to identify market trends.
- **Sentiment Analysis**: Applied sentiment data analysis techniques via R, Python, and MatLab to evaluate Bitcoin-related sentiments on social media.
- **Data Visualization**: Created word clouds, sentiment graphs, machine learning tools and discussion networks to illustrate findings.
- **Bonus Statistical Analysis**: Employed Principal Component Analysis (PCA) to explore relationships between social media sentiments and market indicators.

1. CRYPTOCURRENCIES REVIEW

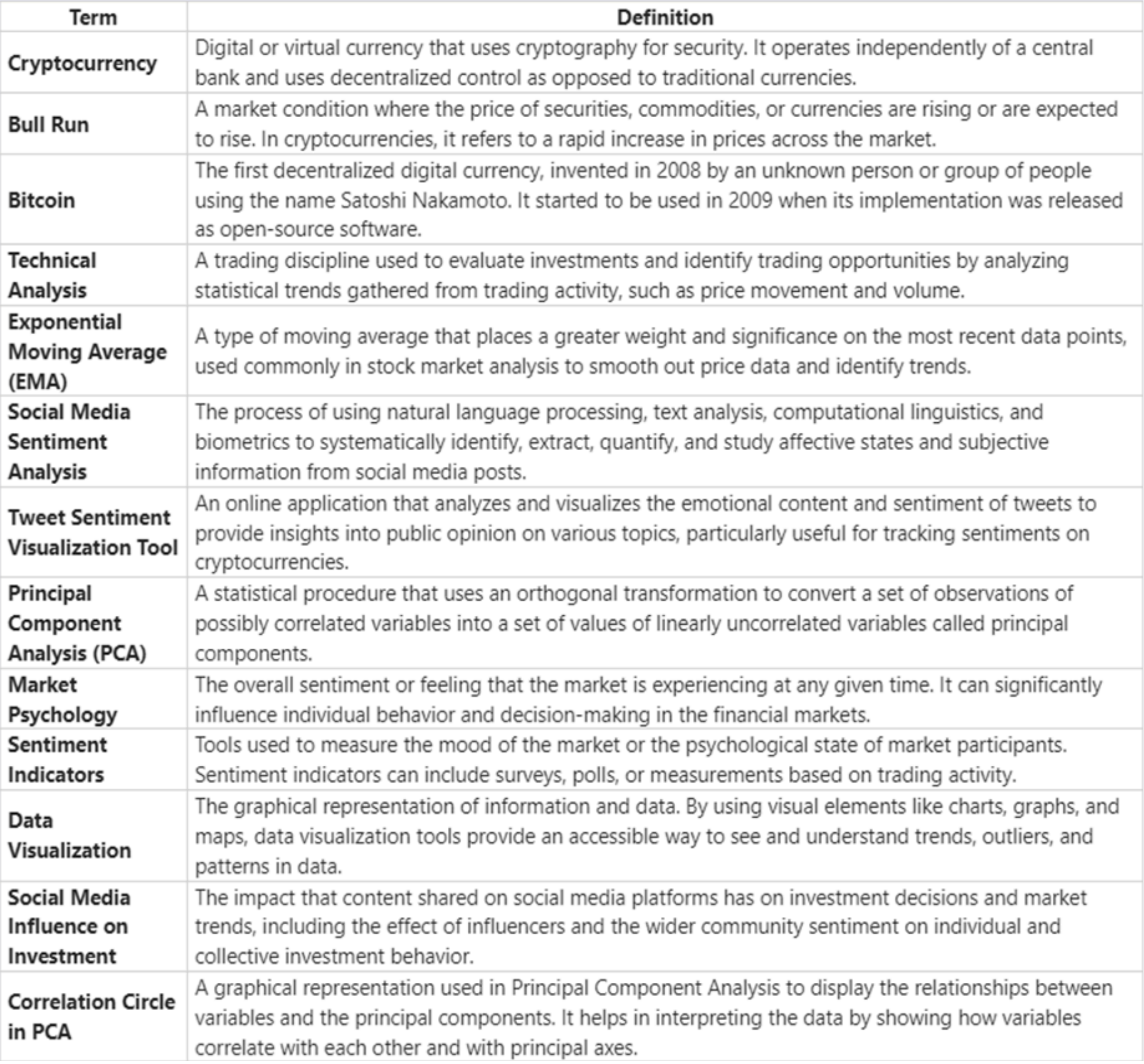
1.1. Popular Opinion

At the heart of behavioral finance analysis, especially in the realm of cryptocurrencies, the contrasting views of influential figures such as Elon Musk and Warren Buffett highlight the spectrum of sentiments and expectations characterizing this notoriously volatile sector. On one hand, Elon Musk exhibits pronounced optimism about the future of cryptocurrencies, stating : "Bitcoin is on the verge of being widely accepted by conventional financiers" (Musk, 2021), suggesting a vision where cryptos gain legitimacy and seamlessly integrate into the global financial system.

On the opposite end of the spectrum, Warren Buffett takes a far more critical stance, asserting : "I can say with almost certainty that cryptocurrencies will come to a bad ending" (Buffett, 2020), reflecting his deep-seated belief in the non-viability of cryptocurrencies as a long-term investment.

These divergent viewpoints underscore the inherently speculative and volatile nature of cryptocurrencies, a reality encapsulated by George Soros when he notes : "If investing is a process, speculation is the art of foreseeing market psychology" (Soros, 2019). Thus, exploring sentiments surrounding cryptocurrencies reveals a fascinating analysis of market psychology, showing how investors' hopes, conjectures, and attitudes significantly influence market movements in an environment as fraught with uncertainties and prone to speculation as that of cryptocurrencies.

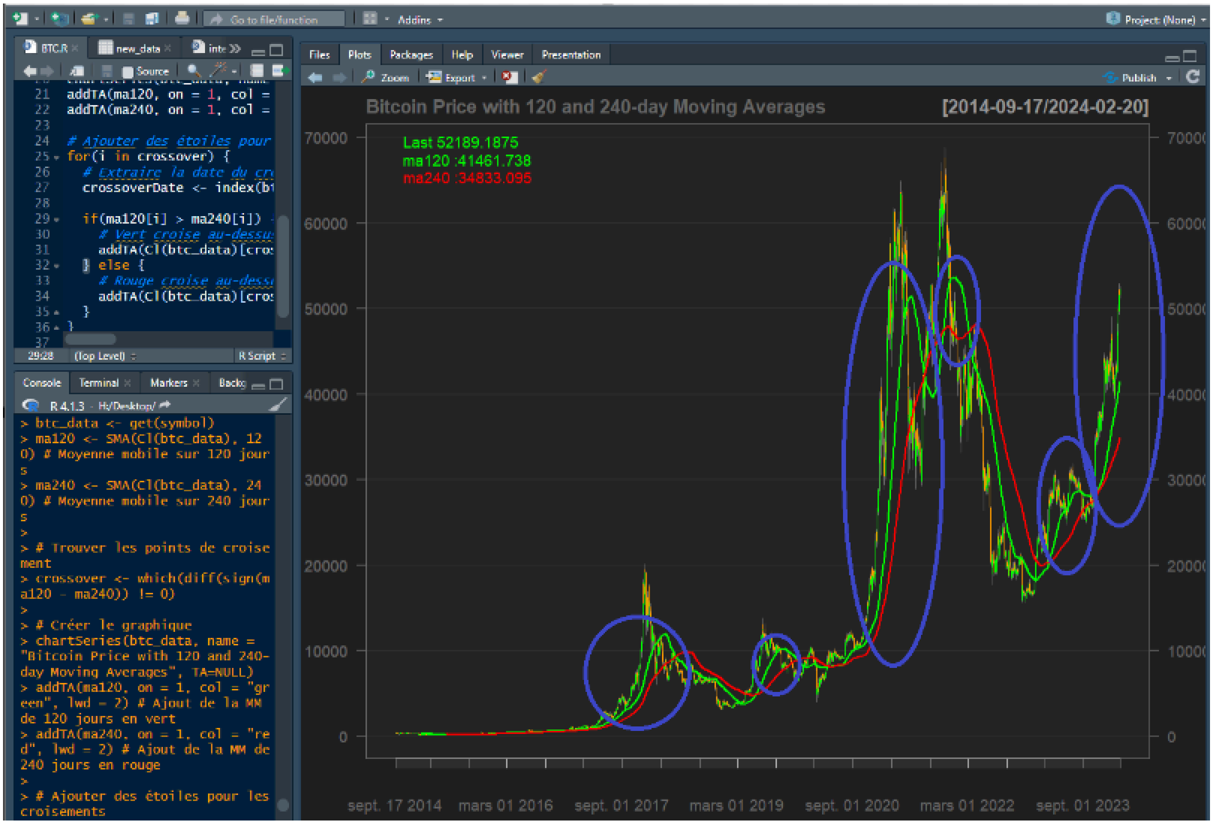
1.2. Definition of Thesis Terms



Term	Definition
Cryptocurrency	Digital or virtual currency that uses cryptography for security. It operates independently of a central bank and uses decentralized control as opposed to traditional currencies.
Bull Run	A market condition where the price of securities, commodities, or currencies are rising or are expected to rise. In cryptocurrencies, it refers to a rapid increase in prices across the market.
Bitcoin	The first decentralized digital currency, invented in 2008 by an unknown person or group of people using the name Satoshi Nakamoto. It started to be used in 2009 when its implementation was released as open-source software.
Technical Analysis	A trading discipline used to evaluate investments and identify trading opportunities by analyzing statistical trends gathered from trading activity, such as price movement and volume.
Exponential Moving Average (EMA)	A type of moving average that places a greater weight and significance on the most recent data points, used commonly in stock market analysis to smooth out price data and identify trends.
Social Media Sentiment Analysis	The process of using natural language processing, text analysis, computational linguistics, and biometrics to systematically identify, extract, quantify, and study affective states and subjective information from social media posts.
Tweet Sentiment Visualization Tool	An online application that analyzes and visualizes the emotional content and sentiment of tweets to provide insights into public opinion on various topics, particularly useful for tracking sentiments on cryptocurrencies.
Principal Component Analysis (PCA)	A statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components.
Market Psychology	The overall sentiment or feeling that the market is experiencing at any given time. It can significantly influence individual behavior and decision-making in the financial markets.
Sentiment Indicators	Tools used to measure the mood of the market or the psychological state of market participants. Sentiment indicators can include surveys, polls, or measurements based on trading activity.
Data Visualization	The graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data.
Social Media Influence on Investment	The impact that content shared on social media platforms has on investment decisions and market trends, including the effect of influencers and the wider community sentiment on individual and collective investment behavior.
Correlation Circle in PCA	A graphical representation used in Principal Component Analysis to display the relationships between variables and the principal components. It helps in interpreting the data by showing how variables correlate with each other and with principal axes.

1.3. Bitcoin Current Trend

1.3.1. Bitcoin Technical Analysis Chart : Identification of Bull Runs - btc.r



```
addTA(ma120, on = 1, col =
addTA(ma240, on = 1, col =

# Ajouter des étoiles pour
for(i in crossover) {
  # Extraire la date du cr
  crossoverDate <- index(bt

  if(ma120[i] > ma240[i])
    # Vert croise au-dessu
    addTA(Cl(btc_data)[cros
  } else {
    # Rouge croise au-dess
    addTA(Cl(btc_data)[cros
  }
}
```

```
> btc_data <- get(symbol)
> ma120 <- SMA(Cl(btc_data), 12
0) # Moyenne mobile sur 120 jour
s
> ma240 <- SMA(Cl(btc_data), 24
0) # Moyenne mobile sur 240 jour
s
>
> # Trouver les points de croise
ment
> crossover <- which(diff(sign(m
a120 - ma240)) != 0)
>
> # Créer le graphique
> chartSeries(btc_data, name =
"Bitcoin Price with 120 and 240-
day Moving Averages", TA=NULL)
> addTA(ma120, on = 1, col = "gr
een", lwd = 2) # Ajout de la MM
de 120 jours en vert
> addTA(ma240, on = 1, col = "re
d", lwd = 2) # Ajout de la MM de
240 jours en rouge
>
> # Ajouter des étoiles pour les
croisements
```

The chart created by the file 'btc.r' illustrates a technical analysis of Bitcoin, using the 120-day and 240-day moving averages to identify the onset of bull markets. When the green line (the 120-day exponential moving average) crosses above the red line (the 240-day average), it traditionally signals the start of a bull market. It's a period anticipated to witness a rise in the price of Bitcoin and other cryptocurrencies. These occurrences are highlighted with blue circles on the chart. According to the present analysis, we seem to be on the cusp of a new bull market. The subsequent phase of my thesis will explore whether public feelings and discussions about Bitcoin correspond with the upward trend indicated by the technical analysis.

Bull markets in the cryptocurrency space are periods of rising prices, influenced by a variety of factors like increasing adoption by institutions and the general public, technological innovations, favorable regulation, integration into payment systems, positive investor sentiment, macroeconomic conditions, fear of missing out (FOMO), and the Bitcoin halving every four years. The Bitcoin halving is an event that cuts the reward for mining each block in half, occurring every four years, reducing the supply of new bitcoins and potentially increasing the price if demand remains constant or increases. These elements can boost demand and confidence in cryptocurrencies, leading to a price increase.

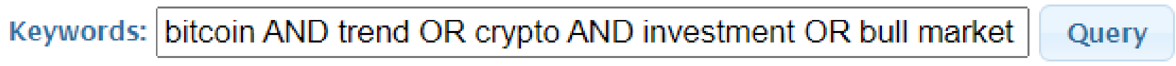
Currently, we are in a bull run, also notably due to BlackRock, the world's largest asset manager, entering the Bitcoin ETF market on the 10 January 2024, marking a significant step toward institutional acceptance of cryptocurrencies, offering increased credibility and accessibility for traditional investors to invest in Bitcoin.

2. EXAMINATION AND DESCRIPTION OF SENTIMENTS

2.1. Choice of the three phrases

For creating a sentiment database, we have to select phrases on the Social Media Sentiment Tool based on the topic we want to investigate. They do not necessarily have to describe the same subject, but they should be related in a way that allows them to collectively give insight into a broader topic of interest.

2.1.1. Screenshot of the phrases selected related to of my topic - Social Media Sentiment Visualization



Keywords: bitcoin AND trend OR crypto AND investment OR bull market

My thesis focuses on the impact of sentiments expressed on social media regarding the cryptocurrency market, unfolding across three principal axes :

- **Bitcoin Trend** : This section is dedicated to analyzing Bitcoin price fluctuations, with a particular focus on how online sentiments can predict these changes because it's the most famous cryptocurrency.
- **Crypto Investment** : Here, I explore how perceptions and discussions on social media influence investment strategies in cryptocurrencies by popular opinion, assessing their impact on investor decisions.
- **Bull Run/Bull market**: This part aims to identify whether sentiments expressed on social media can anticipate or coincide with rapid growth periods in the cryptocurrency market, marking significant bullish phases.

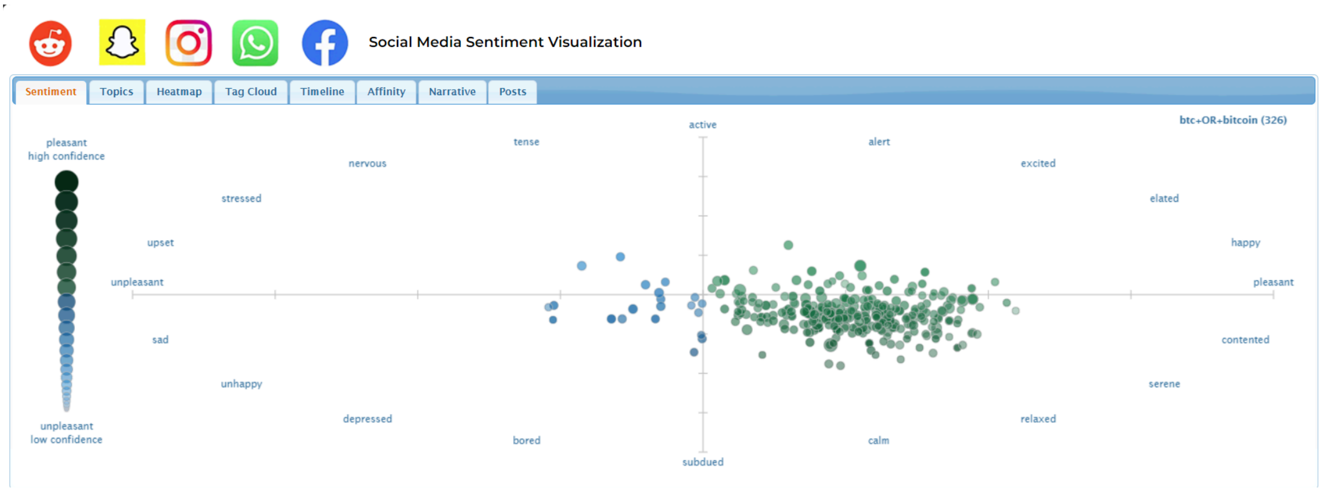
2.2. All the entries that match the selected phrases

In my sentiment analysis project on Bitcoin through social media, the initial use of "btc OR bitcoin" on the Tweet Sentiment Visualization Tool quickly flooded the data system within three days, making multi-week analysis impossible. In response, I refined my search with "AND trend" witch drastically reduced the volume of tweets by targeting discussions on Bitcoin trends. I further filtered the results with "btc," which cleaned the data from irrelevant content even more. This adjusted method facilitated efficient data flow management and enabled a qualitative analysis over five weeks, offering nuanced insights into public opinion that were initially unattainable.

Since the tweet sentiment visualization tool did not offer an option to directly download the data in a tabular form, I developed an R script 'data.r' to structure this plain text into a usable tabular format. This process first involved splitting the text into individual lines and determining the columns, then using specific R functions like strsplit to segment the text and read.delim to organize these segments into a dataframe. I converted the date strings into actual date objects to sort and group them effectively by week. Using the openxlsx package, I programmed the script to create a separate Excel sheet for each week, inserting the corresponding tweets into it. This approach successfully transformed a mass of plain text into a structured and well-organized dataset, distributed by week in an Excel file, thus facilitating temporal and sentimental analysis of tweets. This file was generated using the 'data2.r' script and is named 'output_weekly.xlsx' even if it was not an obligation for sorting informations..

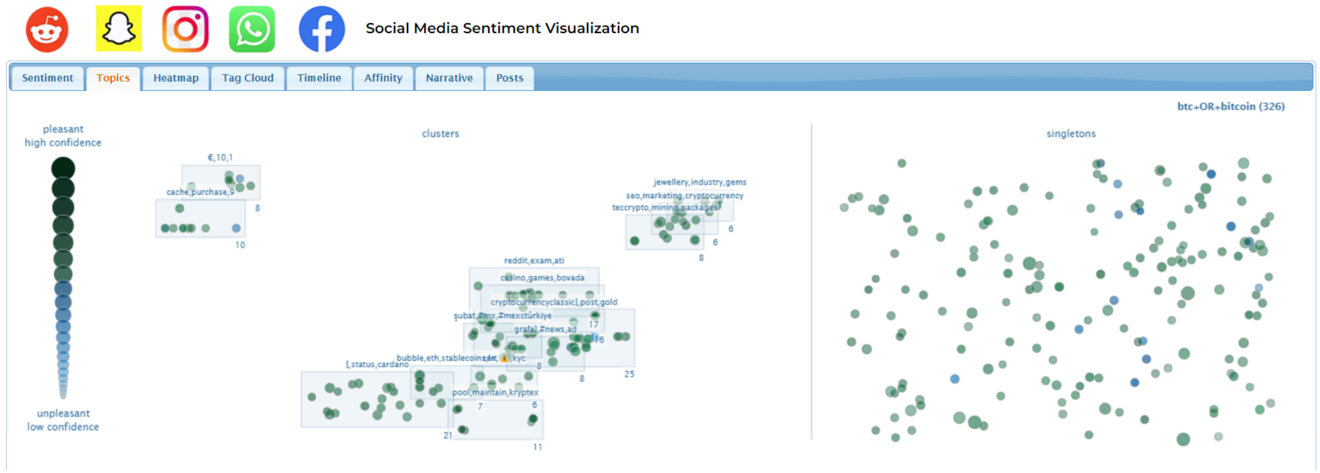
2.3. Description of the data received

2.3.1. Sentiment and Activity Scatter Plot - Social Media Sentiment Visualization



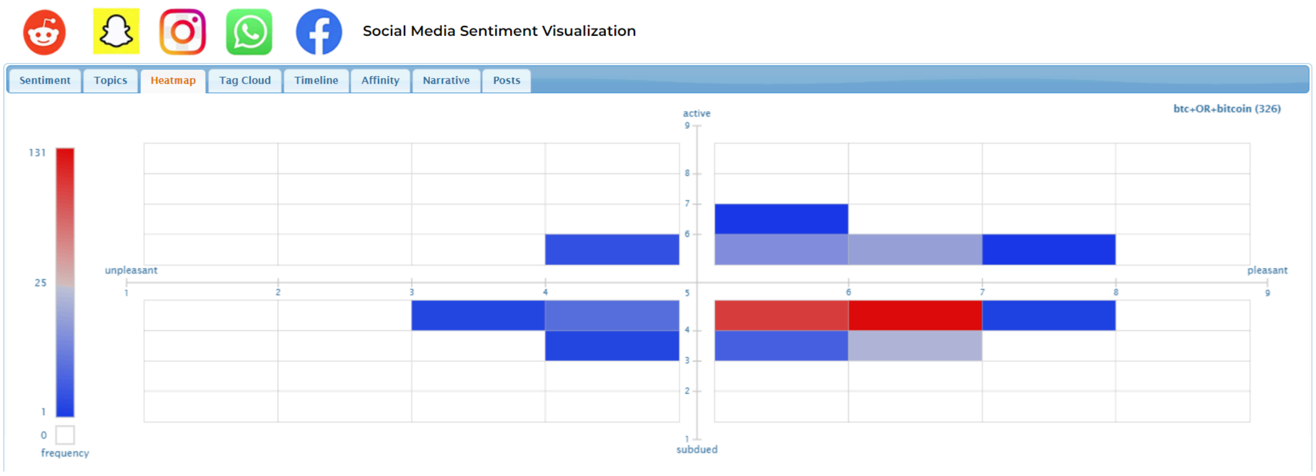
The application provides a sentiment scatter plot visualizing the emotional content of tweets. Each tweet is positioned on a chart that maps sentiment on two axes : one ranging from unpleasant to pleasant and another from relaxed to active. This visualization allows users to quickly grasp the overall mood of social media discussions related to Bitcoin and to identify whether the sentiment is predominantly positive or negative, as well as passive or active. We can see in the graph that the feelings are mostly green which means pleasant with a high confidence. That seems logical because we're probably currently in a bull market period.

2.3.2. Word Cloud Topic Visualization - Social Media Sentiment Visualization



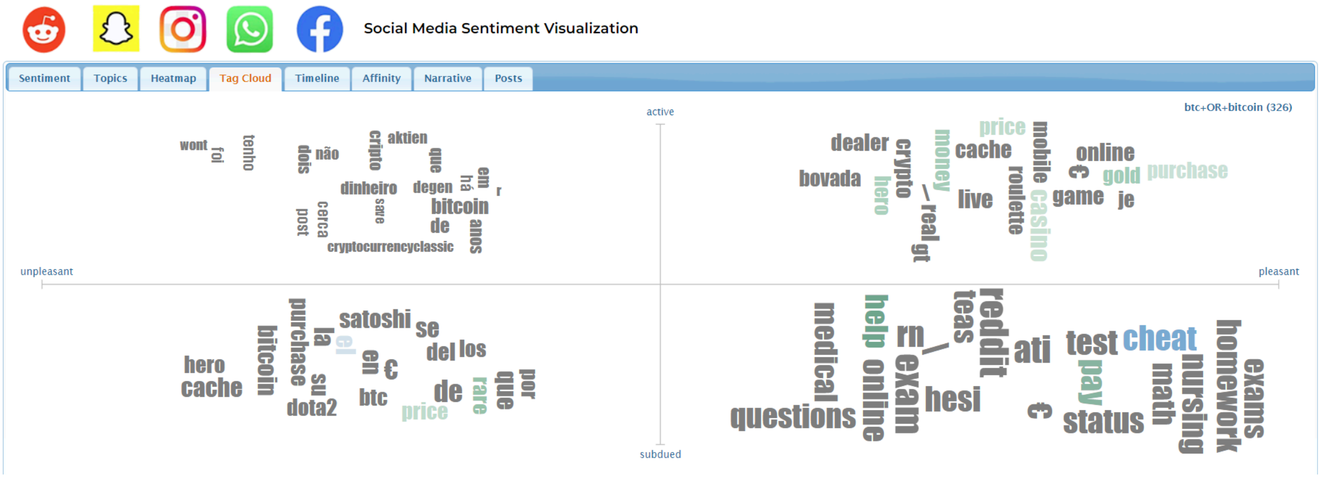
A prominent feature of the application is the word cloud, which aggregates the most frequently used words in the Bitcoin-related tweets. The size of each word in the cloud indicates its frequency of occurrence. This provides a visual representation of key terms and topics that are currently trending in discussions about Bitcoin on social media platforms and clusters are created. The sentiments of posts are visualized on the Sentiment tab, where each post is represented as a circle. The position of these circles is determined based on the estimated sentiment of the words of the post's text. Posts with unpleasant sentiments are shown as blue circles on the left side, while those with pleasant sentiments are depicted as green circles on the right. The vertical positioning of the circles indicates the activity level of the post, with sedate posts represented by darker circles at the bottom and more active posts by brighter circles at the top.

2.3.3. Sentiment Bar Chart Heatmap - Social Media Sentiment Visualization



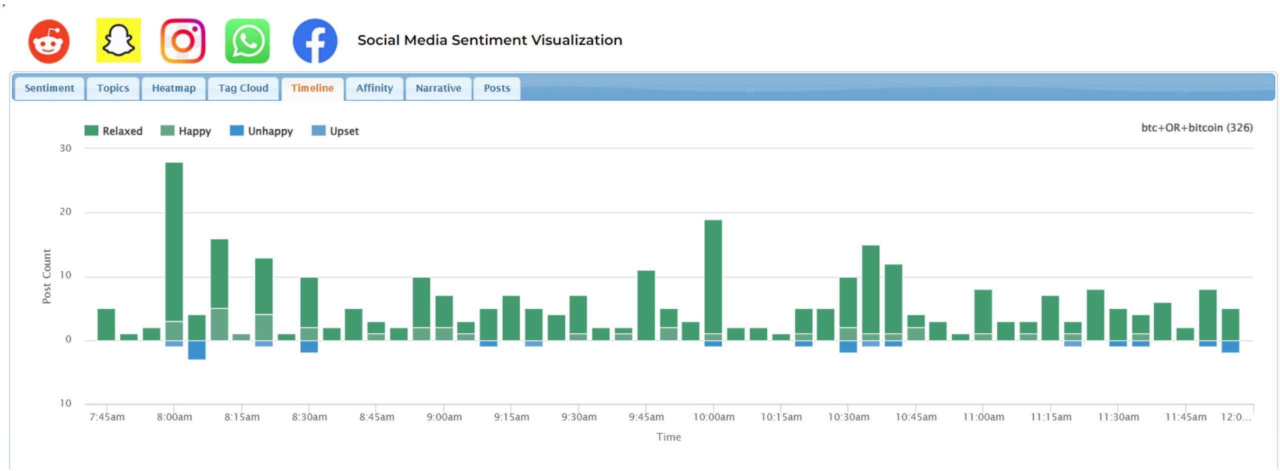
The application includes a bar chart that tracks the frequency of sentiments. This view of sentiment can offer insights into how public emotions and opinions fluctuate in response to market events or news developments related to Bitcoin.

2.3.4. Network Graph Tag Cloud - Social Media Sentiment Visualization



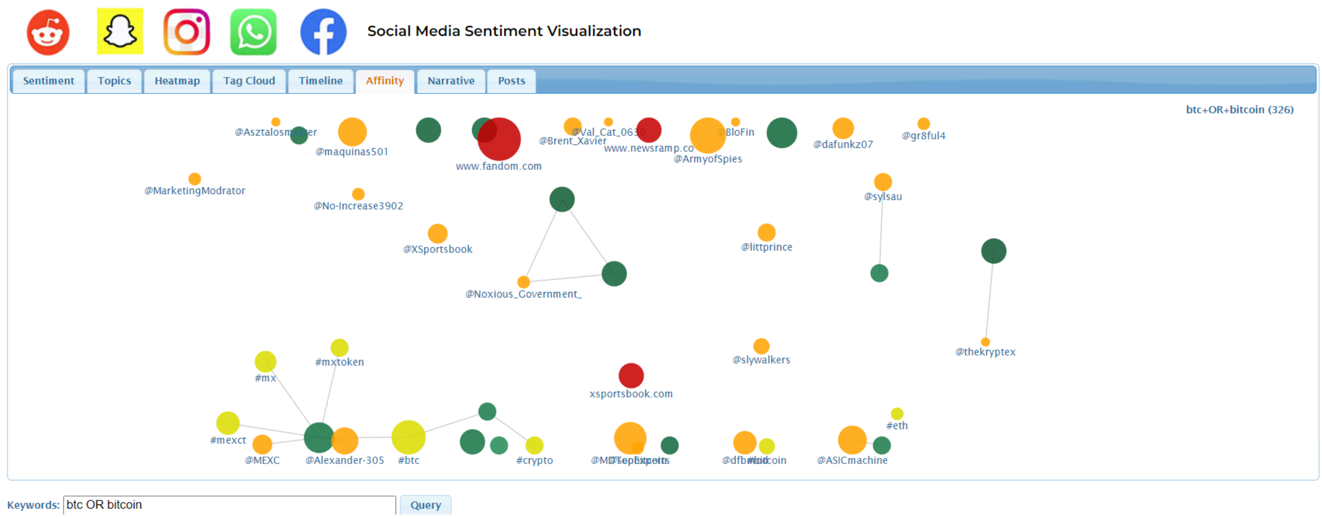
There's also a network graph that illustrates the connections between different Twitter users or hashtags. This helps in understanding how information spreads across the network and which users or hashtags are most influential in the Bitcoin conversation space.

2.3.5. Sentiment Timeline - Social Media Sentiment Visualization



A sentiment trend line is displayed, indicating sentiment scores across a timeline. This line graph may represent individual tweet sentiments or aggregate sentiment over time, providing a clear visual of how overall sentiment is trending over a set period.

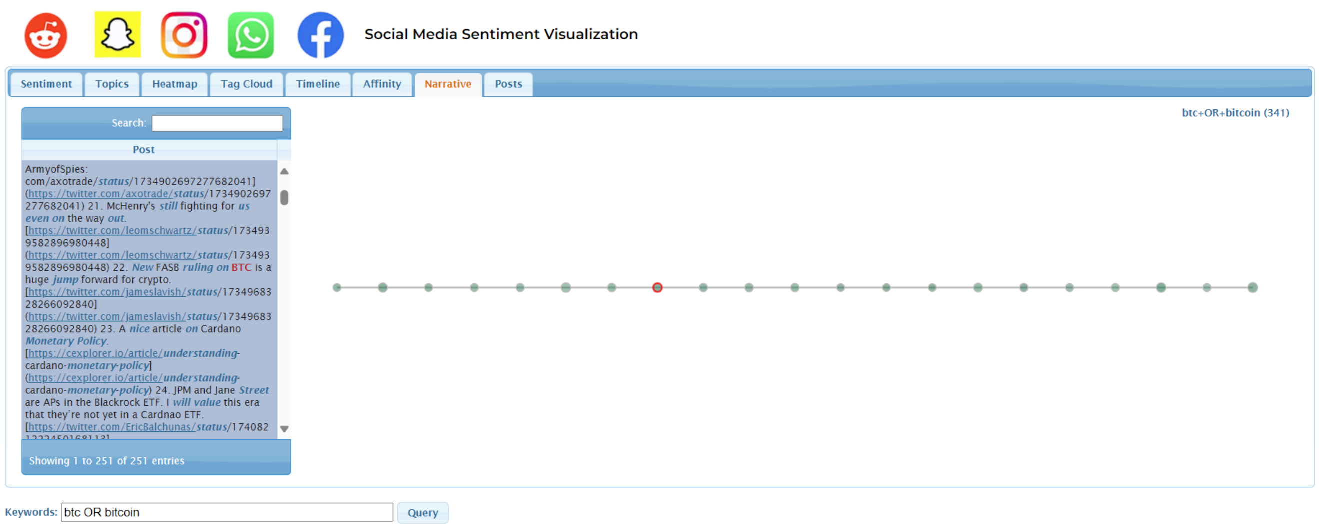
2.3.6. Network Graph Affinity - Social Media Sentiment Visualization



Keywords: btc OR bitcoin

There's also a network graph that illustrates the connections between different Twitter users or hashtags. This helps in understanding how information spreads across the network and which users or hashtags are most influential in the Bitcoin conversation space. Record clusters and single records are organized in the Topics tab. Here, posts that discuss similar themes are grouped into topic clusters, with keywords displayed above each cluster to identify its main theme. Single records, or posts that don't align with any specific topic cluster, are displayed as singletons on the right side. This layout allows users to see at a glance how posts are grouped by common topics or stand alone based on their content.

2.3.7 Tabular Data and Tweet Excerpts Narrative - Social Media Sentiment Visualization

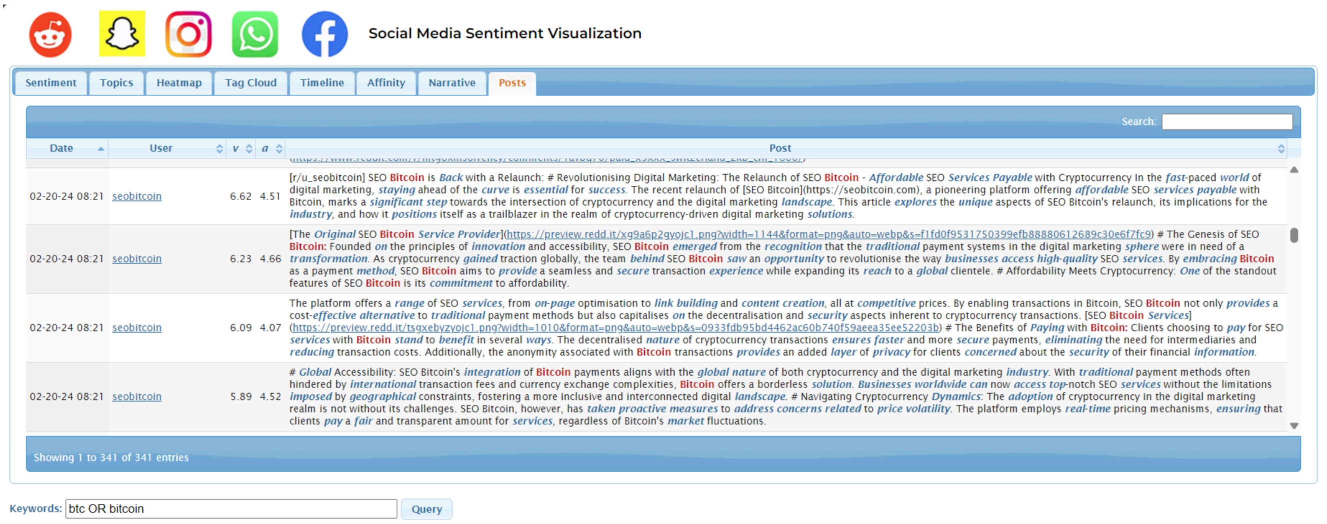


ArmyofSpies: com/axotrade/status/1734902697277682041] (https://twitter.com/axotrade/status/1734902697277682041) 21. McHenry's still fighting for us even on the way out. [https://twitter.com/leomschwartz/status/1734939582896980448] (https://twitter.com/leomschwartz/status/1734939582896980448) 22. New FASB ruling on BTC is a huge jump forward for crypto. [https://twitter.com/jameslavish/status/1734968328266092840] (https://twitter.com/jameslavish/status/1734968328266092840) 23. A nice article on Cardano Monetary Policy. [https://cexplorer.io/article/understanding-cardano-monetary-policy] (https://cexplorer.io/article/understanding-cardano-monetary-policy) 24. JPM and Jane Street are APs in the Blackrock ETF. I will value this era that they're not yet in a Cardnao ETF. [https://twitter.com/EricBalchunas/status/174082

Keywords: btc OR bitcoin

The application presents tabular data showing excerpts from tweets along with their associated sentiment scores and the dates of posting. This table allows for a more detailed textual analysis of individual tweets and provides a structured way to review the sentiment data over time.

2.3.8. Every type of Posts - Social Media Sentiment Visualization



Date	User	v	a	Post
02-20-24 08:21	seobitcoin	6.62	4.51	[r/u_seobitcoin] SEO Bitcoin is Back with a Relaunch: # Revolutionising Digital Marketing: The Relaunch of SEO Bitcoin - Affordable SEO Services Payable with Cryptocurrency In the fast-paced world of digital marketing, staying ahead of the curve is essential for success. The recent relaunch of [SEO Bitcoin](https://seobitcoin.com), a pioneering platform offering affordable SEO services payable with Bitcoin, marks a significant step towards the intersection of cryptocurrency and the digital marketing landscape. This article explores the unique aspects of SEO Bitcoin's relaunch, its implications for the industry, and how it positions itself as a trailblazer in the realm of cryptocurrency-driven digital marketing solutions.
02-20-24 08:21	seobitcoin	6.23	4.66	[The Original SEO Bitcoin Service Provider](https://preview.redd.it/xg9a6p2gyojc1.png?width=1144&format=png&auto=webp&s=f1fd0f9531750399efb88880612689c30e6f7fc9) # The Genesis of SEO Bitcoin: Founded on the principles of innovation and accessibility, SEO Bitcoin emerged from the recognition that the traditional payment systems in the digital marketing sphere were in need of a transformation. As cryptocurrency gained traction globally, the team behind SEO Bitcoin saw an opportunity to revolutionise the way businesses access high-quality SEO services. By embracing Bitcoin as a payment method, SEO Bitcoin aims to provide a seamless and secure transaction experience while expanding its reach to a global clientele. # Affordability Meets Cryptocurrency: One of the standout features of SEO Bitcoin is its commitment to affordability.
02-20-24 08:21	seobitcoin	6.09	4.07	The platform offers a range of SEO services, from on-page optimisation to link building and content creation, all at competitive prices. By enabling transactions in Bitcoin, SEO Bitcoin not only provides a cost-effective alternative to traditional payment methods but also capitalises on the decentralisation and security aspects inherent to cryptocurrency transactions. [SEO Bitcoin Services] (https://preview.redd.it/tsgxebyzyojc1.png?width=1010&format=png&auto=webp&s=0933fdb95bd4462ac60b740f59aeea35ee52203b) # The Benefits of Paying with Bitcoin: Clients choosing to pay for SEO services with Bitcoin stand to benefit in several ways. The decentralised nature of cryptocurrency transactions ensures faster and more secure payments, eliminating the need for intermediaries and reducing transaction costs. Additionally, the anonymity associated with Bitcoin transactions provides an added layer of privacy for clients concerned about the security of their financial information.
02-20-24 08:21	seobitcoin	5.89	4.52	# Global Accessibility: SEO Bitcoin's integration of Bitcoin payments aligns with the global nature of both cryptocurrency and the digital marketing industry. With traditional payment methods often hindered by international transaction fees and currency exchange complexities, Bitcoin offers a borderless solution. Businesses worldwide can now access top-notch SEO services without the limitations imposed by geographical constraints, fostering a more inclusive and interconnected digital landscape. # Navigating Cryptocurrency Dynamics: The adoption of cryptocurrency in the digital marketing realm is not without its challenges. SEO Bitcoin, however, has taken proactive measures to address concerns related to price volatility. The platform employs real-time pricing mechanisms, ensuring that clients pay a fair and transparent amount for services, regardless of Bitcoin's market fluctuations.

Keywords: btc OR bitcoin

The dataset from the "Social Media Sentiment Visualization" tool is organized into a structured table, designed for sentiment analysis on social media platforms. The table comprises key columns:

- **Date** : Timestamps of each post, detailed to the minute, facilitating temporal analysis.
- **User** : Identifiers for the individuals posting, allowing for user-specific trend analysis.
- **Sentiment (v)**: Numeric indicator probably showing if a post's sentiment is viral.
- **Intensity (a)** : Numeric scores gauging the emotional intensity of posts, enriching sentiment analysis.
- **Post**: The full text of each message, including links and hashtags.

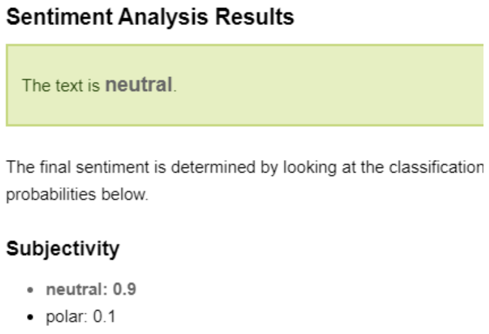
Each row represents a unique message, ready for analysis to uncover trends and sentiment patterns in social media discourse.

3. ANALYSIS OF COLLECTED TEXTUAL INFORMATION

3.1. Using Python Online Program and Sentigem API

3.1.1. Week 1 - 22/01/2024 to 28/01/2024

With Python



Sentiment Analysis Results

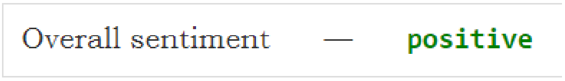
The text is **neutral**.

The final sentiment is determined by looking at the classification probabilities below.

Subjectivity

- **neutral: 0.9**
- polar: 0.1

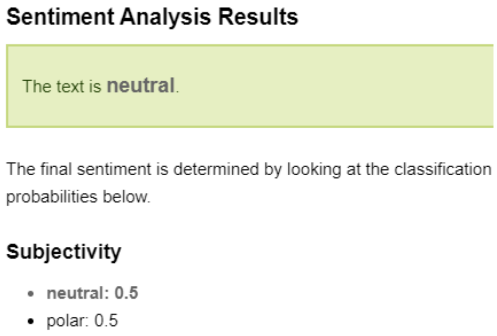
With Sentigem



Overall sentiment — **positive**

3.1.2. Week 2 - 29/01/2024 to 04/02/2024

With Python



Sentiment Analysis Results

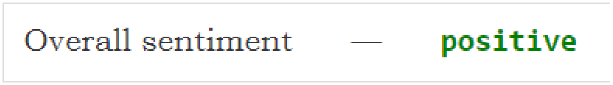
The text is **neutral**.

The final sentiment is determined by looking at the classification probabilities below.

Subjectivity

- **neutral: 0.5**
- polar: 0.5

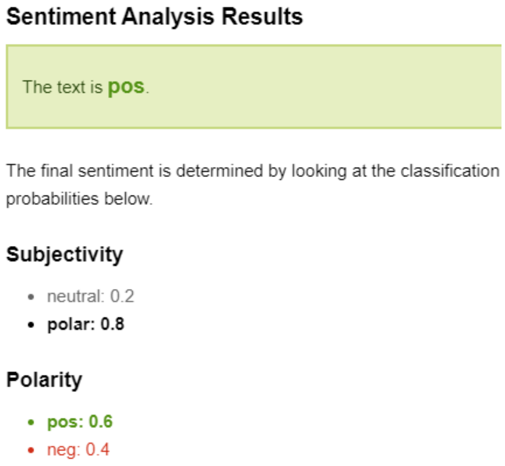
With Sentigem



Overall sentiment — **positive**

3.1.3. Week 3 - 05/02/2024 to 11/02/2024

With Python



Sentiment Analysis Results

The text is **pos**.

The final sentiment is determined by looking at the classification probabilities below.

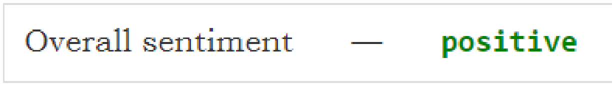
Subjectivity

- neutral: 0.2
- **polar: 0.8**

Polarity

- **pos: 0.6**
- neg: 0.4

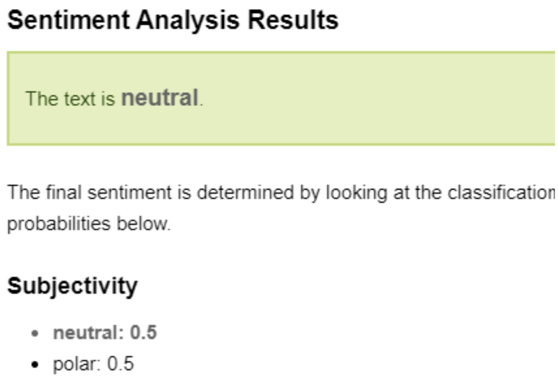
With Sentigem



Overall sentiment — **positive**

3.1.4. Week 4 - 12/02/2024 to 18/02/2024

With Python



Sentiment Analysis Results

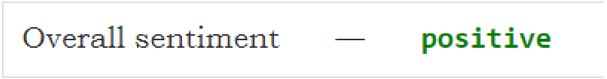
The text is **neutral**.

The final sentiment is determined by looking at the classification probabilities below.

Subjectivity

- neutral: 0.5
- polar: 0.5

With Sentigem



Overall sentiment — positive

The sentiment analysis of Bitcoin-related tweets over the observed period demonstrates a prevalence of record clusters and single records are organized in the Topics tab. Here, posts that discuss similar themes are grouped into topic clusters, with keywords displayed above each cluster to identify its main theme. Single records, or posts that don't align with any specific topic cluster, are displayed as singletons on the right side. This layout allows users to see at a glance how posts are grouped by common topics or stand alone based on their content. eelings. By utilizing two analytical methodologies API, with Python and Sentigem, we note a distinctly favorable trend in the expressed reactions. Sentigem, which specifically assesses the sentiment conveyed by full sentences, reflects a consistently positive perception throughout the month. Concurrently, the Python-based approach, which concentrates on the analysis of specific keywords or text snippets, indicates a surge of positivity during the third week, with more neutral outcomes for the other weeks. This comprehensive assessment of tweet sentiments provides an insight into the general mood prevailing in the cryptocurrency market and could be instrumental in deciphering market trends during this evidently very positive period.

3.2. Using MatLab algorithms

To assess the trend in the cryptocurrency market, particularly to identify a potential bull market, the analysis of textual data from social media can be an invaluable tool. Two MATLAB scripts will be used to process and analyze text content drawn from discussions on Bitcoin and other cryptocurrencies, from an Excel file named 'tweets.xlsx' which collects tweets over a month. These scripts deploy natural language processing techniques to extract information that could indicate the current state of the market.

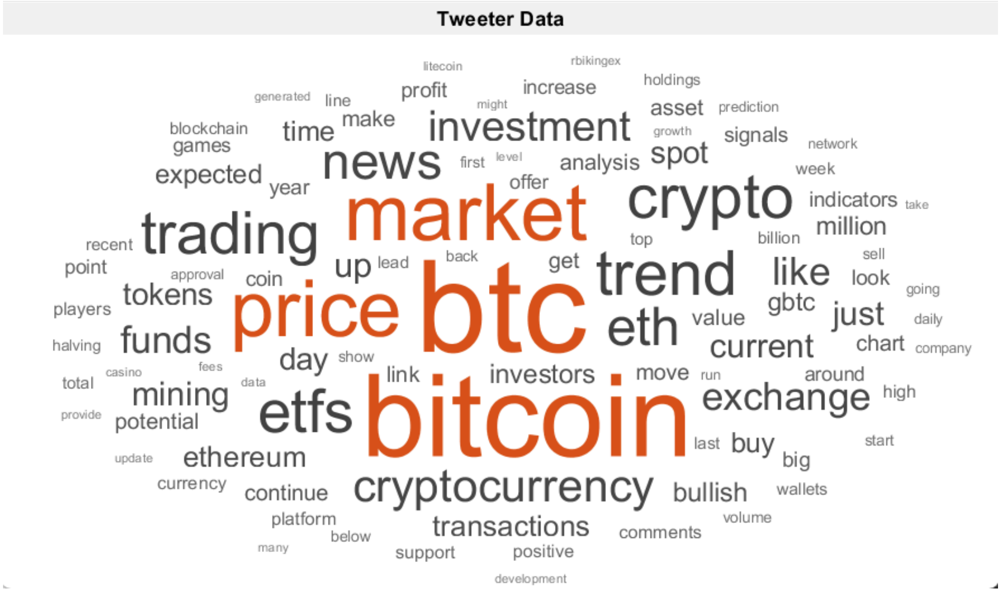
The first script takes an initial approach aiming to generate a word cloud from the posts. This method is straightforward: it cleanses the data by removing URLs and punctuation and displays the most frequent terms to give a general overview of the most discussed topics. This provides an immediate snapshot of word frequency, which can signal the dominant themes in public discourse.

The second script provides a far more detailed analysis. It not only preprocesses the text more thoroughly but also applies more sophisticated techniques such as n-gram analysis and Latent Dirichlet Allocation (LDA). These methods capture the subtleties of language by recognizing not just individual words but also phrases and sentence structures, as well as identifying recurring themes or patterns in the data. The addition of word normalization, such as lemmatization, and the removal of insignificant words further sharpens the analysis, potentially revealing more precise insights into market sentiment.

While the first code may suggest general market activity and an increased presence of certain keywords, the second code delves deeper, potentially exposing not just whether the market is active, but the nature of that activity: optimistic, cautious, or otherwise. Together, these scripts provide a powerful foundation for those looking to understand market dynamics through the lens of social media.

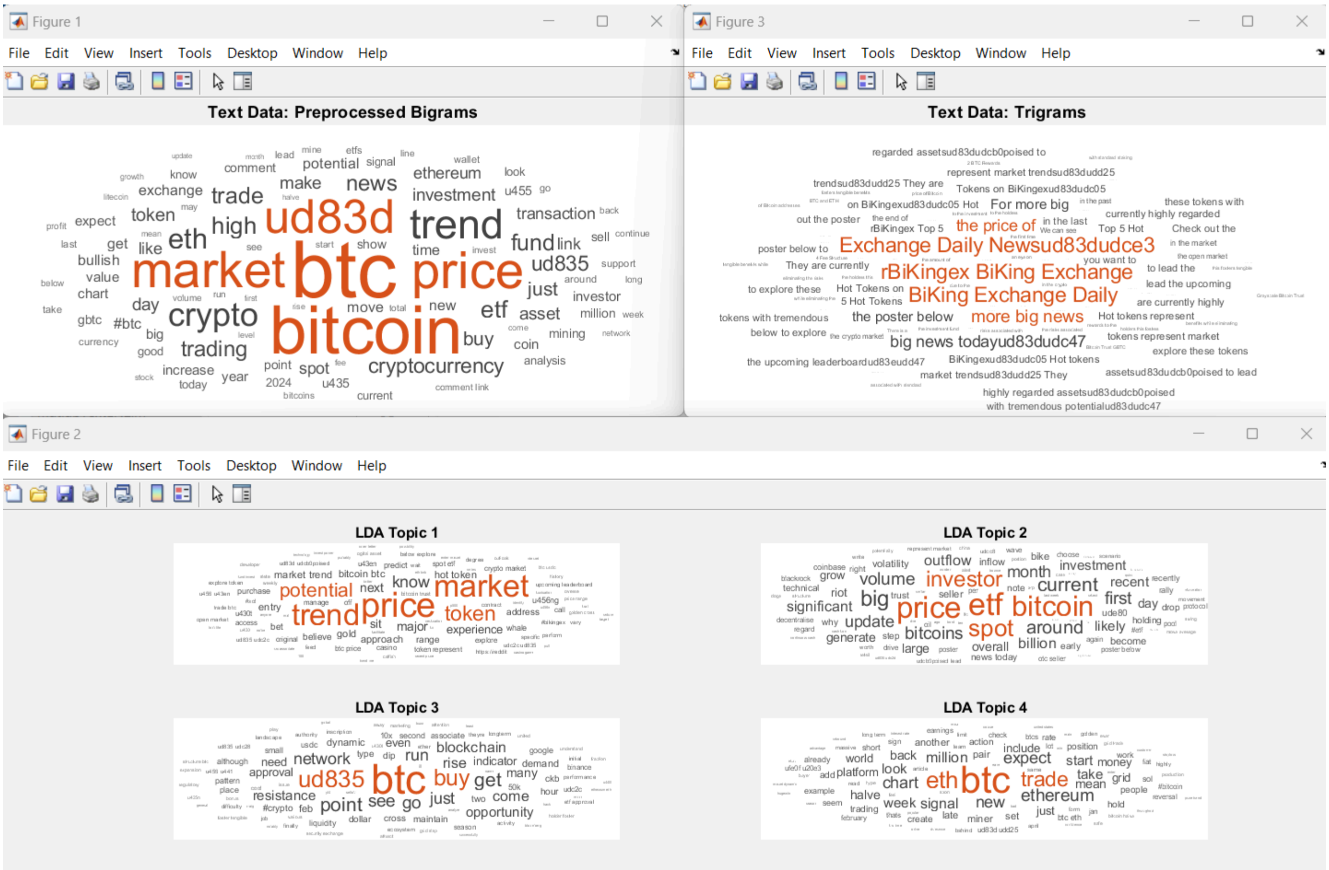
These three different scripts all generate different kinds of word clouds about the cryptocurrencies's topic.

3.2.1. Word Cloud 1 - MatlabTextEN.m



The examination of the word cloud reveals a dominant use of terms directly related to the realm of cryptocurrencies, such as "market," "crypto," "bitcoin," "btc," "price," "trading," "etfs," "investment," and "trend," which attests to the relevance and focus of the keywords in relation to the subject of my thesis. Furthermore, the notable presence of positively connoted terms such as "bullish," "buy," "positive," "big," "approval," "growth," "top," "high," "continue," and "potential" strengthens the assumption of an optimistic sentiment and could be interpreted as an indicator of a bullish trend in the cryptocurrency market.

3.2.2. Word Cloud 2 - MatlabTextEN2.m



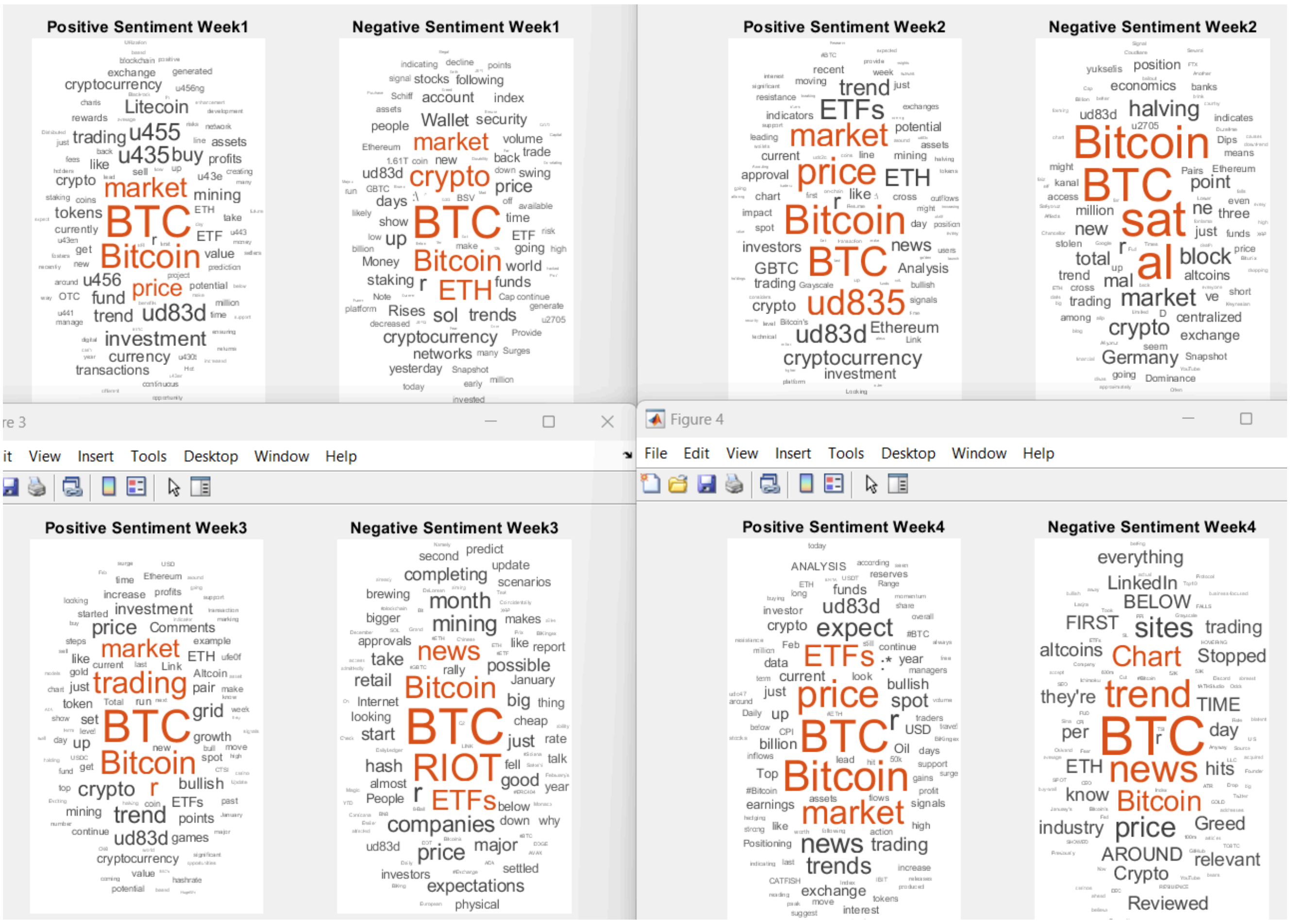
In the bigram word cloud, dominant terms such as "market," "btc," "bitcoin," "price," and "trading," along with the presence of positive words like "high," "bullish," and "trend," may indicate a focus on price evolution and could reflect a growing interest or positive market dynamics.

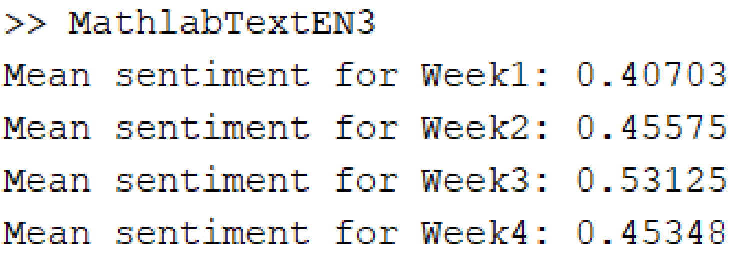
The trigram word cloud reveals notable terms, including hashtags and specific entity names such as "#BikingExchange," "Exchange Daily," and "big news today." These terms suggest specific announcements or news that could be related to positive market events.

The LDA topics present a variety of thematic discussions, with some centered on prices and market movements and others on investor sentiment and technical aspects. Keywords for Topic 1 include "potential," "market," "price," and "token," while Topic 2 focuses on "outflow," "investor," "month," and "investment." Topic 3 highlights terms like "network," "btc," and "resistance point," and Topic 4 focuses on "trade," "eth," "btc," and "ethereum."

Based solely on these words, there are indications of robust market activity and discussions about prices and investments, often associated with a bull market. Terms such as "high," "bullish," “good” and "growth" imply a positive sentiment, but a detailed sentiment analysis of individual tweets would be necessary to determine with certainty whether the market is in a bullish phase or not.

3.2.3. Word Cloud 3 - MatlabTextEN3.m





```
>> MathlabTextEN3
Mean sentiment for Week1: 0.40703
Mean sentiment for Week2: 0.45575
Mean sentiment for Week3: 0.53125
Mean sentiment for Week4: 0.45348
```

Week 1 - 22/01/2024 to 28/01/2024:

- Average Sentiment: 0.40703, indicating a slightly positive sentiment.
- Word Clouds: In the positive cloud, terms like "market" and "investment" suggest optimistic discussions about investment and the market. In the negative cloud, terms such as "volatility" and "security" may reflect concerns regarding market security and volatility.

Week 2 - 29/01/2024 to 04/02/2024:

- Average Sentiment: 0.45575, a higher score than Week 1, indicating an improvement in sentiment.
- Word Clouds: Words like "bullish" and "growth" in the positive cloud may indicate expectations of growth and a bullish trend, while "halving" and "volatility" in the negative cloud could signal uncertainties related to specific events in the crypto world.

Week 3 - 05/02/2024 to 11/02/2024:

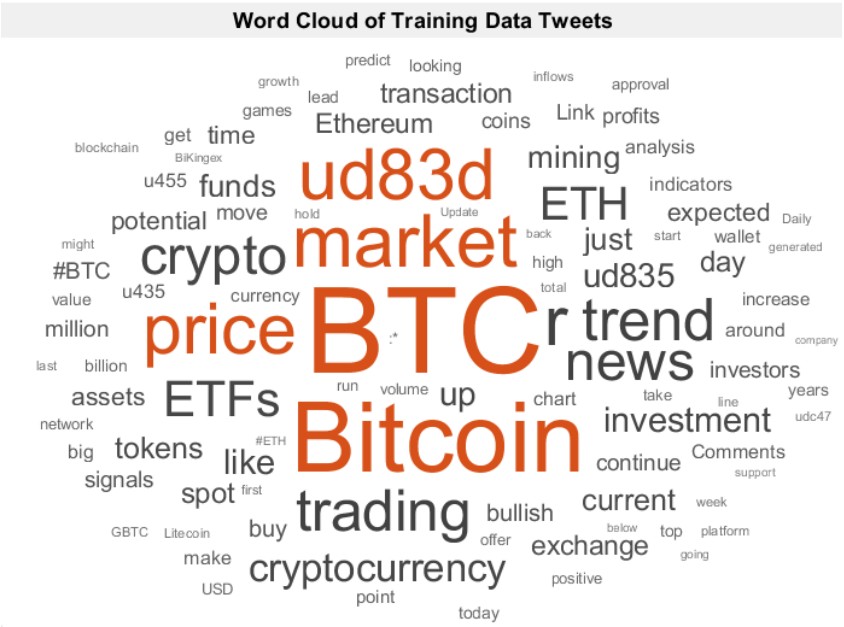
- Average Sentiment: 0.53125, the highest score among the four weeks, indicating a significantly positive sentiment.
- Word Clouds: Words like "bullish" and "high" in the positive cloud are clearly associated with positive outlooks. However, even with an overall positive sentiment, words like "riot" in the negative cloud could be related to specific events or protest movements in the crypto world.

Week 4 - 12/02/2024 to 18/02/2024

- Average Sentiment: 0.45348, a slight decline compared to Week 3, but still an overall positive sentiment.
- Word Clouds: The positive cloud contains terms like "bullish" and "top," while the negative cloud includes "stopped" and "greed," which could reflect a reaction to market slowdown or controversial practices in the industry.

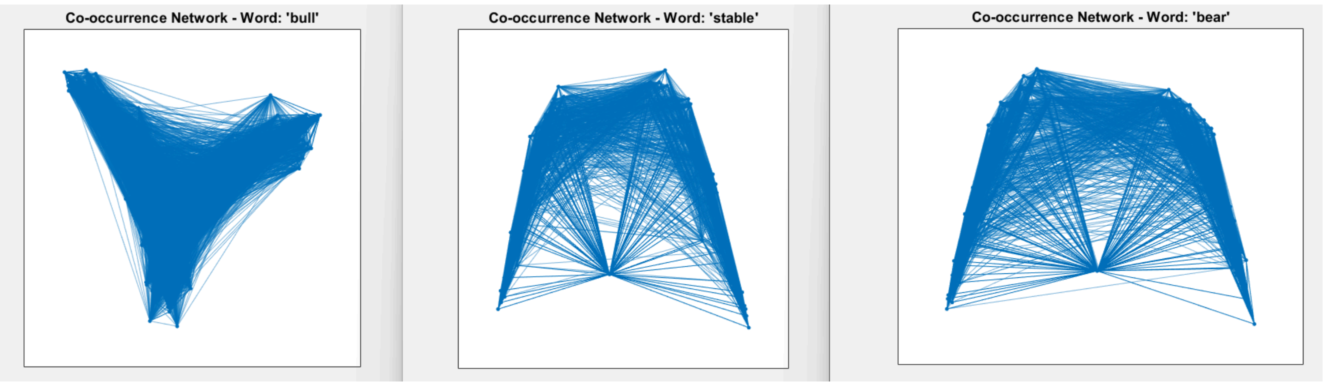
The sentiment analysis of cryptocurrency tweets over the month reveals an upward trend in sentiment, indicating increasing positivity among users discussing crypto-related topics. This trend aligns with the rising prices of cryptocurrencies during the same period

3.2.4. Word Cloud 4 - MatlabTextSTM8EN.m



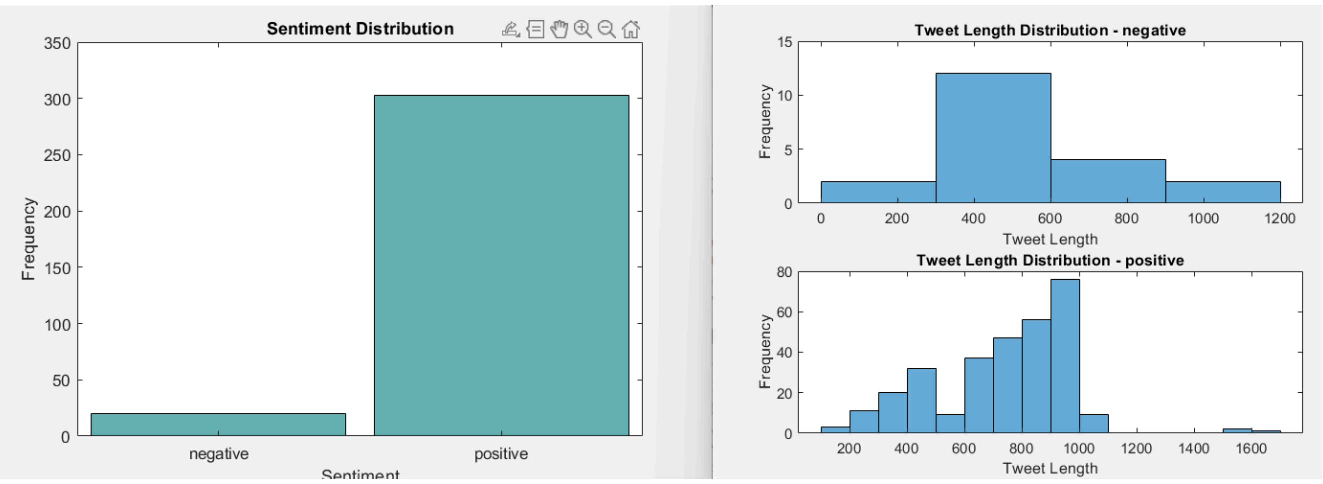
The word cloud showcases key terms from a dataset intended for training a neural network to perform sentiment analysis or topic modeling within the cryptocurrency domain. The dominant words such as "market," "BTC," "Bitcoin," "crypto," "trading," "ETFs," and "investment" signal a strong focus on cryptocurrency market activities, trading dynamics, and investment vehicles. The term "ETFs" might particularly refer to discussions about BlackRock's foray into cryptocurrency through ETF offerings, which is a notable development in the financial sector. Such topics represented in the training data are crucial as they likely influence the neural network's ability to detect and interpret market sentiment and trends in cryptocurrency, potentially making it adept at recognizing nuanced conversations, including investor responses to ETFs like those offered by BlackRock.

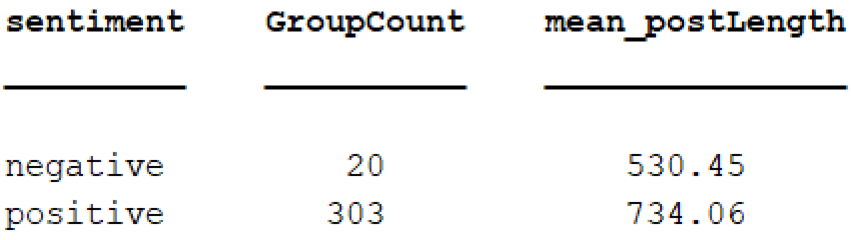
3.2.5. Co-occurrence Network - MatlabTextEN4.m



The network visualizations for "bull," "stable," and "bear" reflect discussions around cryptocurrency. Notably, the network linked to "bull" shows a high density of connections, suggesting that conversations are heavily focused on the concept of a bull market. This indicates active discussions within the community about periods of growth or optimism regarding cryptocurrency prices. In contrast, the networks for "stable" and "bear" display different patterns, representing other market aspects like stability or a bearish trend. However, the pronounced density and centrality of the "bull" network highlight a strong focus on an upward trend in the market sentiment within the discussions captured.

3.2.6. Sentiment and Tweet Length Distribution - MatlabTextEN3.m



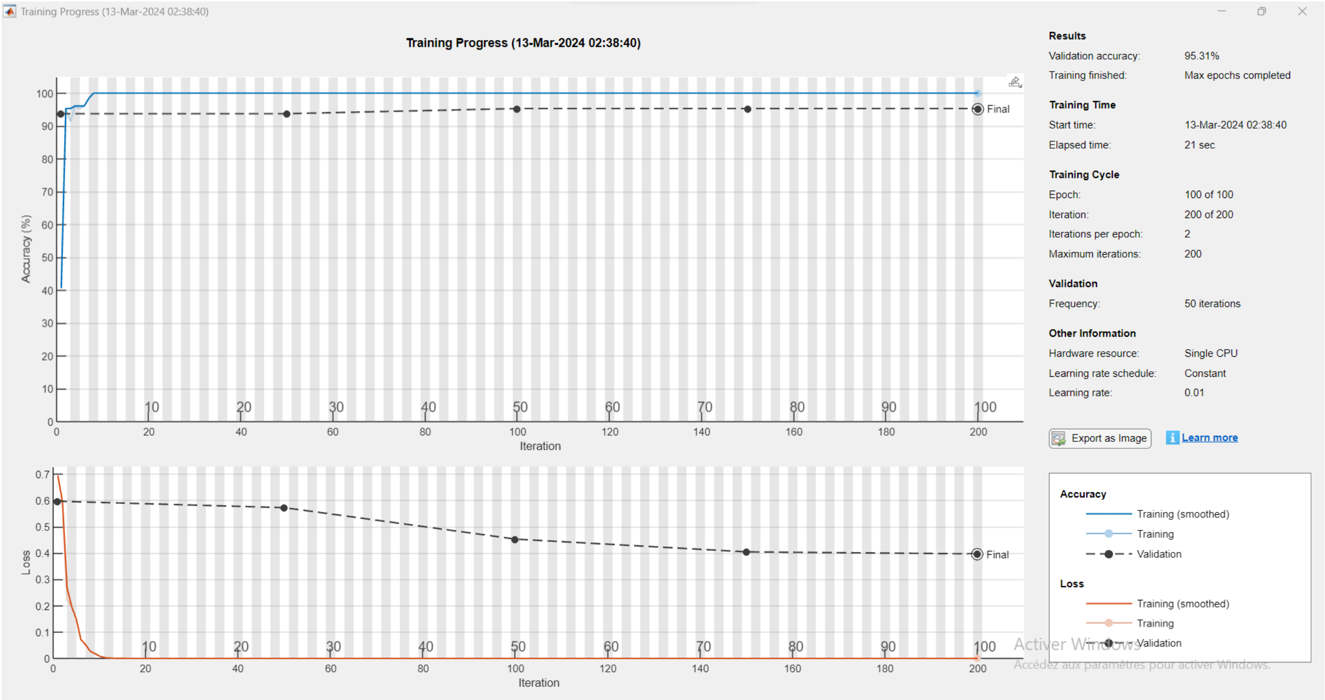


```
sentiment    GroupCount    mean_postLength
_________    __________    _______________

negative         20            530.45
positive        303            734.06
```

The sentiment analysis of tweets during a bull market for cryptocurrencies shows a much higher count of positive tweets as opposed to negative ones, with 303 positive tweets compared to 20 negative tweets. The average length of positive tweets is longer than that of negative ones, with positive tweets averaging 734.06 characters compared to 530.45 characters for negative tweets. This suggests more detailed or elaborate discussions in favor of cryptocurrencies, reflecting an overall optimistic sentiment during this market phase.

3.2.7. Training of the Neural Network Model - MatlabSTM8EN.m



Results

Validation accuracy: 95.31%

Training finished: Max epochs completed

Training Time

Start time: 13-Mar-2024 02:38:40

Elapsed time: 21 sec

Training Cycle

Epoch: 100 of 100

Iteration: 200 of 200

Iterations per epoch: 2

Maximum iterations: 200

Validation

Frequency: 50 iterations

Other Information

Hardware resource: Single CPU

Learning rate schedule: Constant

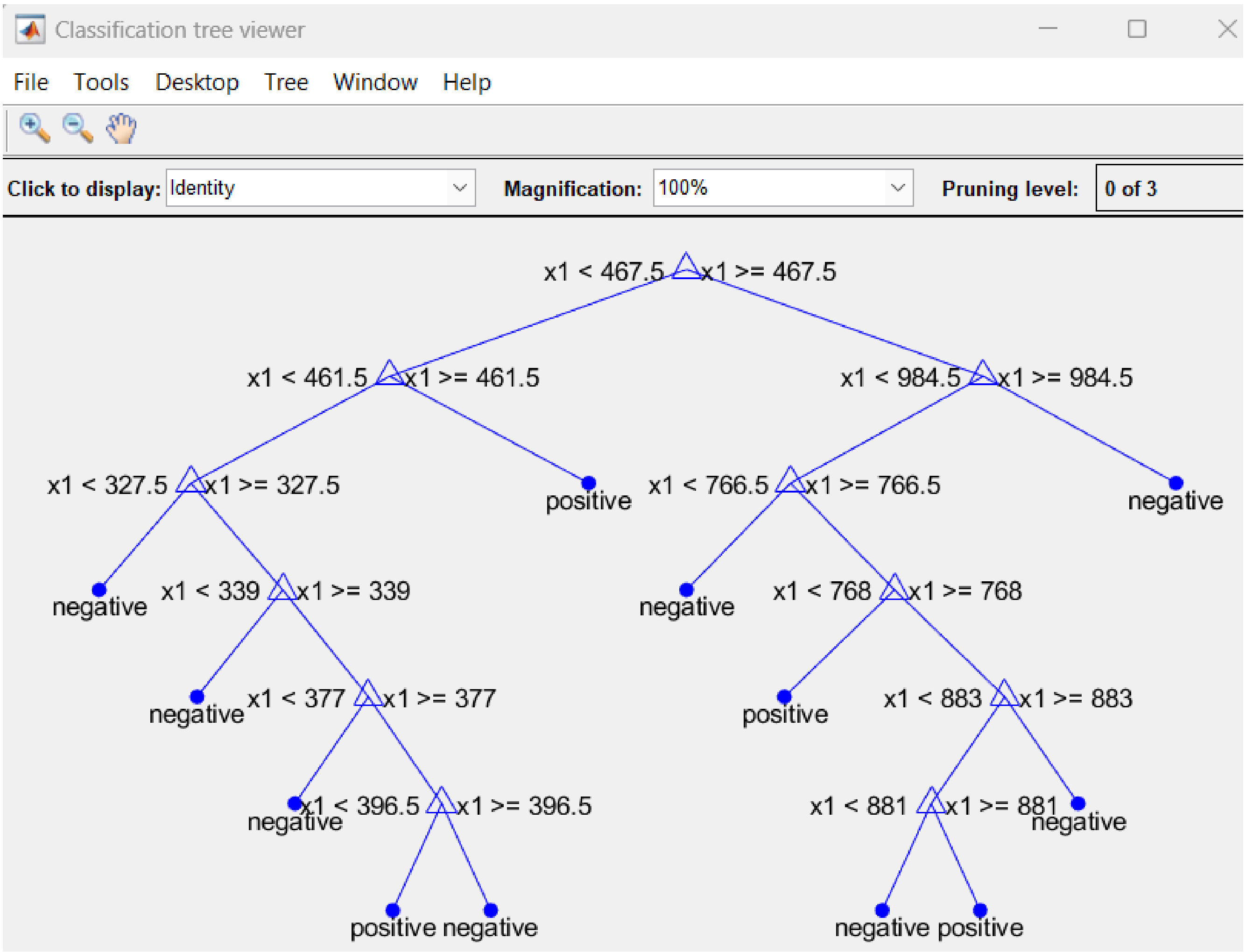
Learning rate: 0.01

```
                                        Phrase                                                 PredictedSentiment
    ____________________________________________________________________________________      __________________

    "Cryptocurrencies are a total scam, leading countless investors to ruin with their volatile markets."    "negative"
    "Cryptocurrencies represent the future of finance, offering unprecedented growth and opportunities."     "positive"
    "I can say with almost certainty that cryptocurrencies will come to a bad ending (Buffett, 2020)"       "negative"
    "Bitcoin is on the verge of being widely accepted by conventional financiers (Musk, 2021)"              "positive"
```

The image displays the training about the length of the postresults of my sentiment analysis model on cryptocurrency-related tweets, achieving a validation accuracy of 95.31%. This high performance suggests the model's proficiency in accurately categorizing tweet sentiments as either positive or negative. It has successfully classified phrases referencing influential figures like Buffett and Musk, indicating an ability to grasp the nuanced context and the impact of prominent individuals on market sentiment. However, tweet classification isn't infallible upon code execution, and the learning curve begins to plateau after 100 epochs, which is why I strategically chose this number to mitigate overfitting, ensuring the model retains a balance between learning and generalization, as indicated by the optimal validation percentage achieved.

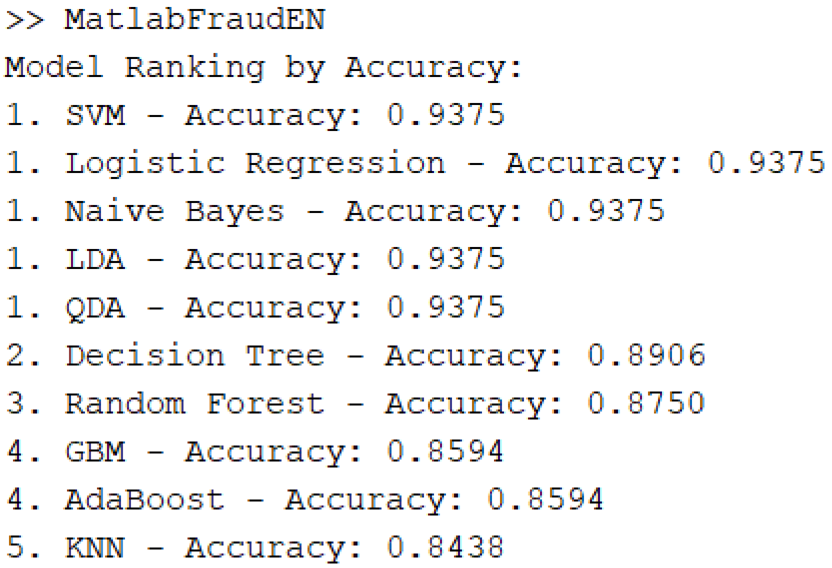
3.2.8. Decision Tree Model - MatlabFraudEN.m



Click to display: Identity **Magnification:** 100% **Pruning level:** 0 of 3

The decision tree depicted is a model derived from sentiment analysis, where the primary feature used for classification is the length of the text (a numeric variable) of tweets. In this visualization, decisions are made at nodes where the dataset is split based on this length. For example, tweets shorter than a certain threshold are classified as negative, while longer ones might be deemed positive. This indicates that, according to this model, the sentiment conveyed in a tweet could be linked to its length, with shorter messages perhaps being more likely to be negative and longer ones positive. However, without additional information, such as model accuracy or comparative analysis with other models, it's challenging to determine the effectiveness of text length alone as a predictor of sentiment. The presence of warnings and errors related to model convergence in the code suggests potential issues with the dataset or the model's assumptions, which would need to be addressed to ensure reliable predictions.

3.2.9. Best models ranking - MatlabFraudEN.m



```
>> MatlabFraudEN
Model Ranking by Accuracy:
1. SVM - Accuracy: 0.9375
1. Logistic Regression - Accuracy: 0.9375
1. Naive Bayes - Accuracy: 0.9375
1. LDA - Accuracy: 0.9375
1. QDA - Accuracy: 0.9375
2. Decision Tree - Accuracy: 0.8906
3. Random Forest - Accuracy: 0.8750
4. GBM - Accuracy: 0.8594
4. AdaBoost - Accuracy: 0.8594
5. KNN - Accuracy: 0.8438
```

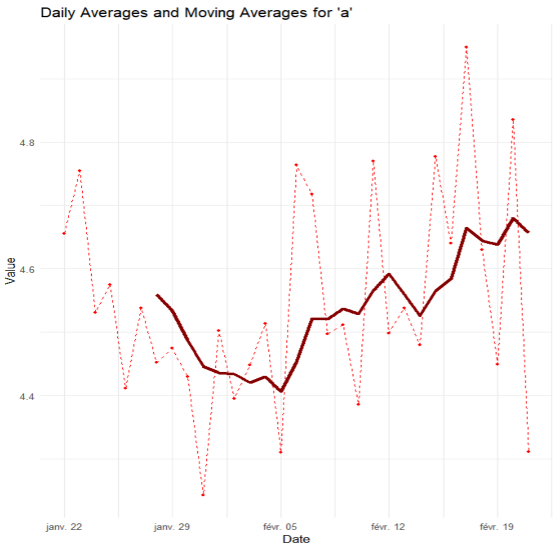
In the analysis of machine learning models aimed at classifying texts based on their length, the neural network exhibits the highest performance with an accuracy of 95.31%. This precision places it above the other evaluated models. In comparison, the SVM (Support Vector Machine), Logistic Regression, Naive Bayes, LDA (Linear Discriminant Analysis), and QDA (Quadratic Discriminant Analysis) models all display an accuracy of 93.75%, which is also commendable but slightly lower than that of the neural network. Decision tree methods show proper performance, and ensemble methods such as Random Forest and AdaBoost also present solid results. The KNN model is the least performant on the list, with an accuracy of 84.38%. To determine the most suitable for the task of text length-based classification, one must consider accuracy as well as other factors such as model comprehensibility, computational cost, and the model's ability to avoid overfitting the data.

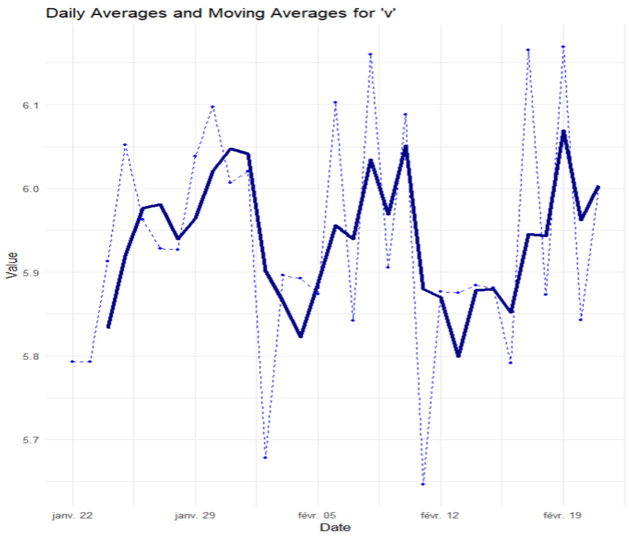
4. LINKING CRYPTOCURRENCY TO SOCIAL MEDIA - BONUS

4.1. Personal Analysis using ‘a’ and ‘v’ Indicators

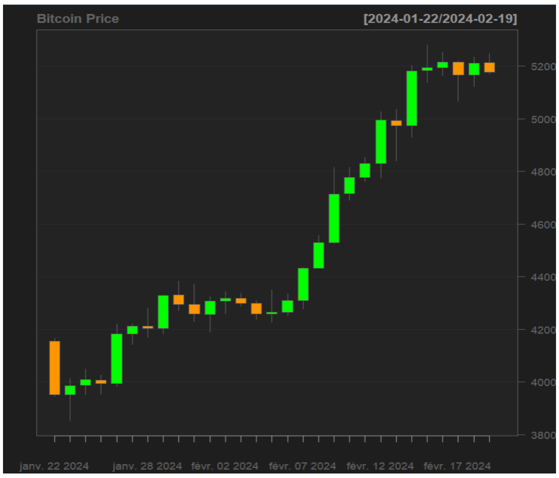
Each day, a multitude of tweets are posted, and to analyze this vast amount of information. On the file ‘analyse2.r’ I have computed a daily average for the 'a' and 'v' indicators from each tweet, to derive a single data point per day. This method has enabled me to create a new data series for 'a' and 'v', which I can then compare with the daily transaction volume and closing price of Bitcoin, respectively. It is important to recall that the 'a' and 'v' indicators are generated by the software based on an analysis of the sentiment keywords and their intensity.

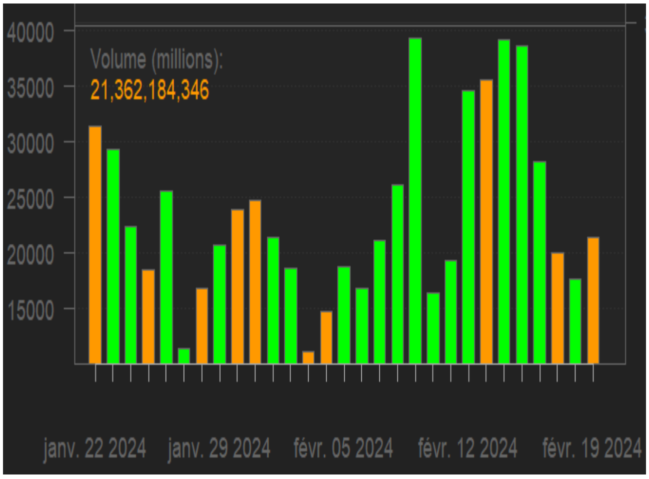
4.1.1. Sentiment Indicators 'a' and 'v' with Their 7-Day and 2-Day Moving Averages - analyse2.r





4.1.2. Bitcoin Price and Trading Volume - btc2.r





An initial examination suggests there are similarities between the evolution of Bitcoin's price and the 7-day moving average of 'a', which represents the intensity of reactions in tweets. A similar trend is also observed between the 2-day moving average of 'v', which reflects the volume of mentions or engagement around Bitcoin on Twitter, and the actual trading volume of Bitcoin. These preliminary observations indicate that sentiment indicators derived from social media expressions could mirror or even influence the dynamics seen in the Bitcoin market.

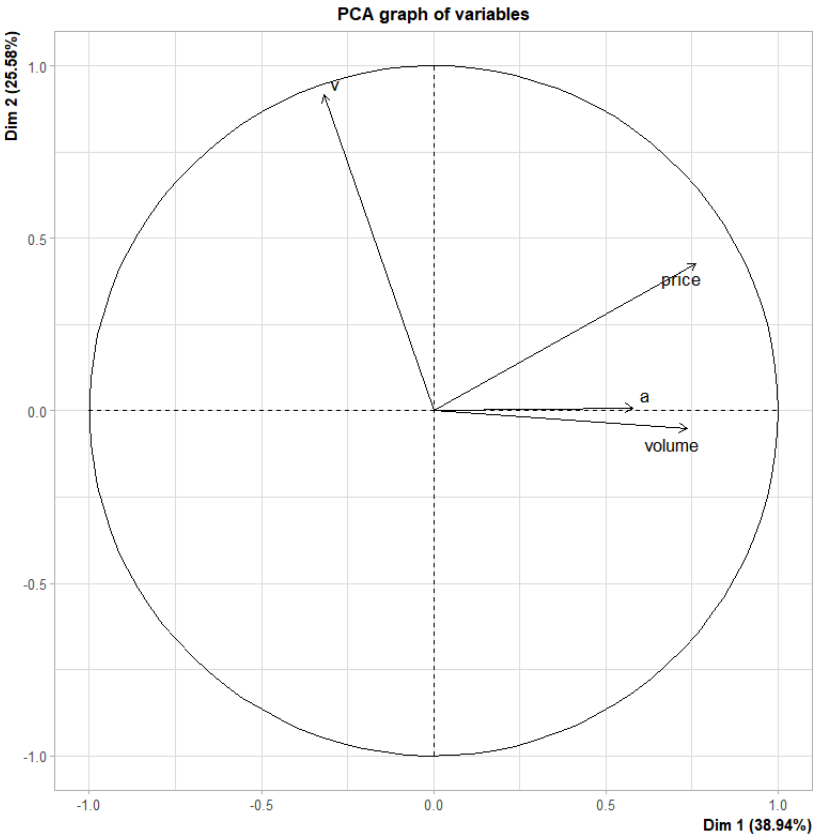
4.2. Correlation Circle Analysis

To verify the hypothesis of correlations, we will conduct a Principal Component Analysis (PCA) with a correlation circle :

A Principal Component Analysis (PCA) is a powerful dimensionality reduction method that simplifies the complexity of datasets. By identifying the principal axes that maximize data variance, PCA reorganizes information into principal components. These new dimensions, or axes, unveil the underlying dynamics of the data by capturing its essence with minimal information loss. The orientation of vectors in this new space of acute, right, or obtuse angles indicates positive correlation, no direct correlation, or negative correlation between variables, respectively.

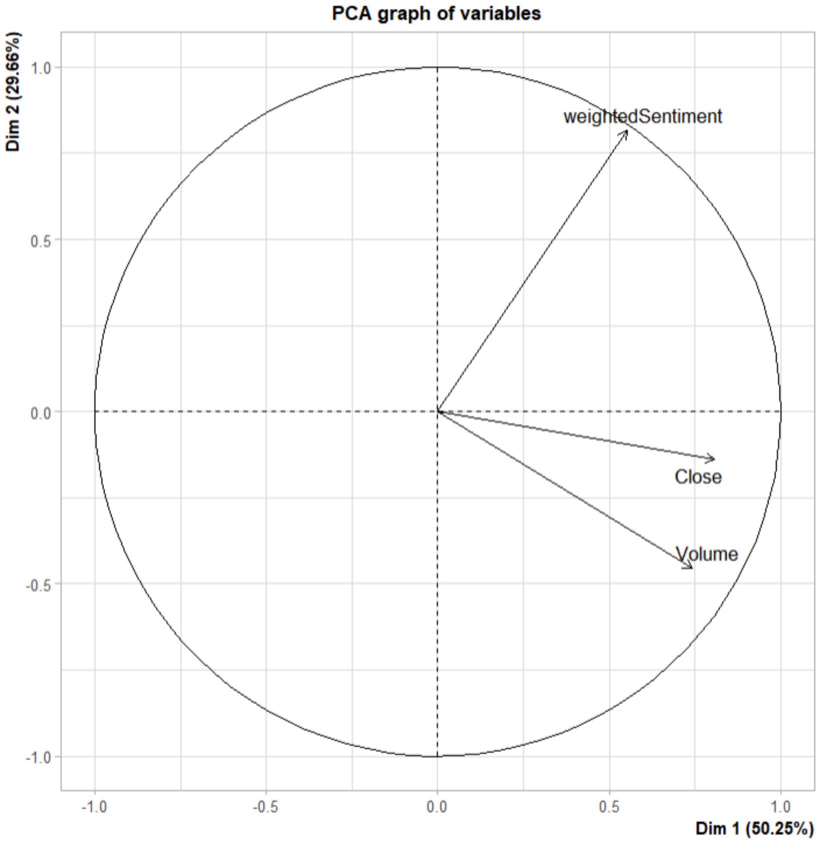
This is the angles between vectors witch reflects their correlation: acute angles (<90°) suggest positive correlation, right angles (=90°) indicate no correlation, and obtuse angles (>90°) hint at negative correlation. The cosine of these angles quantifies the correlation, with 1 (0°) for perfect positive, -1 (180°) for perfect negative, and 0 (90°) for no correlation. Thus, by calculating the cosine of the angle between two vectors, one can directly assess their correlation coefficient, offering a precise measure of how the represented variables relate to each other.

4.2.1. Correlation Analysis of Bitcoin and tweets indicators - ACPbtc.r



The PCA chart analysis of vector orientations offers valuable insights into the relationships among various variables related to the Bitcoin market from January 22 to February 19. The 'volume' and 'price' exhibit similar orientations, demonstrating a positive correlation: an increase in trading volume tends to coincide with a rise in Bitcoin prices, which is logical during a bull run as people buy more Bitcoin. The 'a' indicator, projecting in a direction close to that of 'volume', suggests that intense reactions on Twitter are closely linked to high trading volumes, indicating that social media activity may reflect or influence market movements. 'a' also maintains a positive relationship with 'price', albeit less pronounced. On the other hand, 'v' is distinguished by its divergent orientation from the other vectors, signaling a dimension of sentiment on Twitter that does not directly follow the trends of trading volume or prices.

4.2.2. Correlation Analysis of Bitcoin and weighted tweet indicators - ACPbtc2.r



The image displays a PCA (Principal Component Analysis) graph illustrating the interplay among three variables related to Bitcoin: 'Close', 'Volume', and 'weightedSentiment'. 'weightedSentiment' is computed by multiplying the difference between a tweet's sentiment score and the neutral value of 5 by the tweet's virality, thus yielding a measure that encapsulates the emotional intensity and reach of the sentiment expressed. On the chart, the horizontal axis (Dim 1) accounts for 50.25% of the total data variance, and the vertical axis (Dim 2) for 28.26%. 'Close' and 'Volume' appear close together, suggesting a degree of correlation and similar influence on the principal data structure. However, the acute angle between 'weightedSentiment' and 'Close' indicates a weaker but existing correlation and it is well represented because the length is near one, implying that the weighted sentiment variably influences the variance pattern of Bitcoin prices.

SYNTHESIS

Conclusions

My thesis examined the potential of social media sentiment analysis to forecast market trends in cryptocurrencies, with a focus on Bitcoin. By integrating technical analysis with an exploration of online sentiment, we uncovered a notable correlation between public sentiment and market movements. The study highlighted the speculative nature of cryptocurrencies, influenced by the divergent views of figures like Elon Musk and Warren Buffett, underscoring the importance of behavioral finance in understanding market dynamics.

Our findings suggest that social media sentiment can indeed serve as an early indicator for market phases, particularly bull runs. However, this study also acknowledges the complexities of cryptocurrency markets, influenced by a wide array of external factors. Future research could further dissect the impact of these factors on sentiment and market behavior, broadening our understanding of digital finance.

In summary, my thesis reinforces the value of combining sentiment analysis with traditional financial analysis to gain insights into cryptocurrency market trends, showing the importance of behavioral finance principles to the digital economy.

Potential improvements

Automating the entire analysis process in a single programming language could significantly enhance the system, enabling comprehensive retrieval of tweets to analyze correlations between social sentiment and cryptocurrency market movements, as well as other financial products. By employing scraping techniques to collect data from sentiment analysis sites and deploying the system on a free AWS server for daily updates to our new data set, this method offers an efficient and cost-effective way to develop a new investment indicator with a daily data point. Not only does this strategy streamline analysis, making the process more consistent and less prone to errors, but it also opens the door to creating valuable predictive tools for investors, leveraging the potential of real-time sentiment analysis to anticipate market trends.

Moreover, incorporating portfolio optimization (as discussed at the beginning of the semester) for cryptocurrencies would have added an extra layer of analysis, by measuring performance and risk using the Sharpe ratio, for instance. This metric assesses an investment's effectiveness by comparing its excess return to the risk incurred, providing a framework to evaluate risk-adjusted profitability. However, our primary focus was on exploring the correlation between tweet sentiments and market fluctuations.

The Mean-Variance and Black-Litterman models would also have been applicable in this context. The former aims to maximize return for a given level of risk, based on the normal distribution of returns, while the latter adjusts portfolio allocations based on investor outlooks. Due to their inherently high volatility, cryptocurrencies would have required a specific adaptation of these models to effectively capture their risk and return potential.

REFERENCES

Quotes

- Musk, E. (2021). "Bitcoin is on the verge of being widely accepted by conventional financiers", Twitter.
- Buffett, W. (2020). "I can say with almost certainty that cryptocurrencies will come to a bad end", CNBC Interview.
- Soros, G. (2019). "If investing is a process, speculation is the art of foreseeing market psychology", The Alchemy of Finance.

Analysis Sentiments Tools

- Tweet Sentiment Visualization: https://www.csc2.ncsu.edu/faculty/healey/tweet_viz/tweet_app/
- Python Text Processing Sentiment Analysis Demo: http://text-processing.com/demo/sentiment/
- Sentigem API: http://sentigem.com/

Articles on Bull Runs

- "What is a Bull or Bear Market?" Coinbase. https://www.coinbase.com/learn/crypto-basics/what-is-a-bull-or-bear-market
- "Crypto ‘bull trap’ could be lying in wait for investors in 2024." Fortune. https://fortune.com/2024/01/26/crypto-bull-trap-2024-btc-eth-etf-halving-outlook-finance-rachel-lin/
- "Understanding the Market." Binance. https://www.binance.com/en/feed/post/546925

Videos on Bitcoin

- "Bitcoin Bull & Bear Markets Explained." By Coin Bureau. https://www.youtube.com/watch?v=r20Oxccvs6I
- "The Truth About Bitcoin's Bull Market." By Andrei Jikh. https://www.youtube.com/watch?v=wrNn5Q-JA48

Code Listings

- Bitcoin Technical Analysis: btc.r, btc2.r
- Data Transformation Script: data.r, data2.r
- Word Cloud Textual: MatlabTextEN.m, MatlabTextEN2.m, MatlabTextEN3.m, MatlabTextEN4.m
- Machine Learning: MatlabSTM8EN.m, MatlabTreeEN.m
- Personal Analysis with ‘a’ and ‘v’ Indicators: analyse2.r
- Correlation Circle Analysis: ACPbtc.r and ACPbtc.r
- Cryptocurrency Investment Platform Simulation (OOP) : Crypto.py
- Cryptocurrencies Predictions : Crypto Predictions (shinyapps.io)

Selection of level:

- Threshold: text analysis by using tweet sentiment visualization tool.
- Typical: analysis by additionally obtaining word clouds.
- Great: analysis by additionally obtaining word nets, clusters or artificial intelligence forecasting.

Term of ending the course project: 28 March 2024.

Lecturer of course project: Assoc. prof. dr. Nijolė Maknickienė
(name, surname)

I got the task:

………………RT ………………
(Signature)

……Romain Taugourdeau……
(name, surname)

…………21-03-2024……………
(Date)

https://www.csc2.ncsu.edu/faculty/healey/tweet_viz/tweet_app/