OPEN ACCESS
Indo-JC
Indonesia Journal of Computing
ISSN 2460-9056
socj.telkomuniversity.ac.id/indojc

Ind. Journal on Computing
Vol. 9, Issue. 2, August 2024. pp. 185-198
doi:10.34818/indojc.2024.9.2.980

Public Perception of Buying and Selling Bitcoin Using Lexicon Sentiment Analysis

Muhammad Rahman Ali[#1], Rifki Wijaya[*2], Prasti Eko Yunanto[#3]

Faculty of Informatics, Informatics, Telkom University, Bandung, Indonesia
[1] rahmanali@student.telkomuniversity.ac.id
[2] rifkiwijaya@telkomuniversity.ac.id
[3] gppras@telkomuniversity.ac.id

Abstract

This study investigates public perceptions of Bitcoin (BTC) trading using sentiment lexicon analysis. The rapid growth of cryptocurrency trading has attracted significant public interest and investment, making it crucial to understand the sentiments and opinions surrounding BTC transactions. This research employs various sentiment analysis methods, including AFINN, Bing, and National Research Council (NRC), to analyze tweets and social media posts to determine public sentiment. Additionally, data imputation methods such as linear interpolation, polynomial interpolation, and moving average are used to address missing data. The final results are analyzed using a correlation heatmap to identify trends and patterns in public opinion and their impact on BTC trading behavior. Preliminary results indicate a correlation between positive sentiment and increased trading activity, while negative sentiment correlates with market declines. This research contributes to a better understanding of the role of public sentiment in the volatile cryptocurrency market.

Keywords : Bitcoin, BTC Trading, Sentiment Analysis, Sentiment Lexicon, Public Perception, Cryptocurrency Market, Bitcoin Prediction

Abstrak

Studi ini menyelidiki persepsi publik terhadap perdagangan Bitcoin (BTC) menggunakan analisis sentimen leksikon. Pertumbuhan pesat perdagangan cryptocurrency telah menarik minat dan investasi publik yang signifikan, sehingga penting untuk memahami sentimen dan opini yang mengelilingi transaksi BTC. Penelitian ini menggunakan berbagai metode analisis sentimen, termasuk AFINN, Bing, dan National Research Council (NRC), untuk menganalisis tweet dan postingan media sosial guna menentukan sentimen publik. Selain itu, metode imputasi data seperti interpolasi linear, interpolasi polinomial, dan moving average digunakan untuk menangani data yang hilang. Hasil akhir dianalisis menggunakan heatmap korelasi untuk mengidentifikasi tren dan pola dalam opini publik serta dampaknya terhadap perilaku perdagangan BTC. Hasil awal menunjukkan adanya korelasi antara sentimen positif dan peningkatan aktivitas perdagangan, sedangkan sentimen negatif berkorelasi dengan penurunan pasar. Penelitian ini memberikan kontribusi pada pemahaman yang lebih baik tentang peran sentimen publik dalam pasar cryptocurrency yang volatil.

Kata kunci : Bitcoin, Perdagangan BTC, Analisis Sentimen, Sentimen Lexicon, Persepsi Publik, Pasar Cryptocurrency, Prediksi Bitocin

Received on April, 2024. Accepted on August, 2024

I. Introduction

Cryptocurrency, often described as digital or virtual currency, utilizes cryptography for security, making it difficult to counterfeit. The advent of cryptocurrency has revolutionized the financial world, offering a decentralized alternative to traditional financial systems [1]. Bitcoin, created in 2008 by an anonymous person or group known as Satoshi Nakamoto, is the first cryptocurrency that garnered widespread attention. Operating on a peer-to-peer network without a central authority, Bitcoin relies on blockchain technology to ensure transparency and security[2].

Bitcoin’s technological foundation and significant market capitalization have made it a focal point for investors and researchers alike. Its decentralized nature and limited supply have led to its characterization as "digital gold," providing a hedge against traditional market volatility. The strength of Bitcoin lies not only in its underlying technology but also in the widespread public interest and perception, which significantly influences its market value[3].

The rise of social media platforms like Twitter has provided a vast and dynamic dataset for researchers. Twitter, with its real-time updates and widespread usage, offers insights into public sentiment and trends. Researchers have increasingly utilized Twitter data to gauge public perception on various topics, including financial markets[4]. Public perception plays a crucial role in shaping behaviors and decisions. In the context of financial markets, the sentiments expressed by the public can have profound effects [5]. Positive or negative perceptions can drive market movements, influencing prices and investment strategies. Understanding public sentiment towards Bitcoin is essential, as it can lead to better predictions of price fluctuations and market trends [6].

Sentiment analysis, a branch of natural language processing (NLP), involves the computational study of opinions, sentiments, and emotions expressed in text. It has become a powerful tool for analyzing public perception, particularly in understanding how opinions shared on social media can impact financial markets [7]. During the corona virus disease of 2019 (*COVID-19)* pandemic, sentiment analysis has been instrumental in assessing public sentiment towards various assets, including Bitcoin . The COVID-19 pandemic has accelerated the adoption of digital technologies and the use of cryptocurrencies as investment and transaction tools. During this period, public sentiment towards Bitcoin experienced sharp fluctuations, influenced by global news, economic events, and technological developments[8][9]. Sentiment analysis during the pandemic has provided strong evidence that public perception can impact Bitcoin prices, reflecting investors' uncertainty and expectations in a rapidly changing global situation[10].

The use of sentiment analysis has shown that public opinion can influence stock prices, as evidenced by studies conducted in India where sentiment analysis of social media data led to significant changes in stock prices [11]. Similarly, incorporating sentiment analysis into the study of Bitcoin prices has shown that public sentiment can significantly impact its value[12].

In this paper, we employ lexicon-based sentiment analysis to study public perception of Bitcoin. Lexicon-based approaches involve using predefined lists of words associated with positive or negative sentiments to analyze text data [13].

By applying this method, we aim to understand how public sentiment towards Bitcoin, as expressed on social media, influences its buy and sell prices. This study is particularly relevant in the current climate, where digital currencies are becoming increasingly integral to the global financial ecosystem.

Over the past decade, Bitcoin has evolved from an experimental concept to a major financial asset recognized worldwide. Its extensive influence on global financial markets has made it a highly intriguing subject for academic research. Analyzing public sentiment towards Bitcoin, particularly through social media platforms like Twitter, provides critical insights into how public perception can affect price volatility and market trends [3].

This research aims to deepen our understanding of how sentiment analysis, specifically lexicon-based approaches, can be used to predict Bitcoin price changes. By leveraging data from social media and advanced sentiment analysis methods, we hope to provide valuable insights for investors, researchers, and policymakers about the dynamics influencing the Bitcoin market.

II. ARCHITECTURE AND LITERATURE REVIEW

This study uses Twitter post data available on the website Kaggle. The research utilizes Twitter post data and historical BTC cryptocurrency data for the period from February 5, 2021, to April 24, 2021.

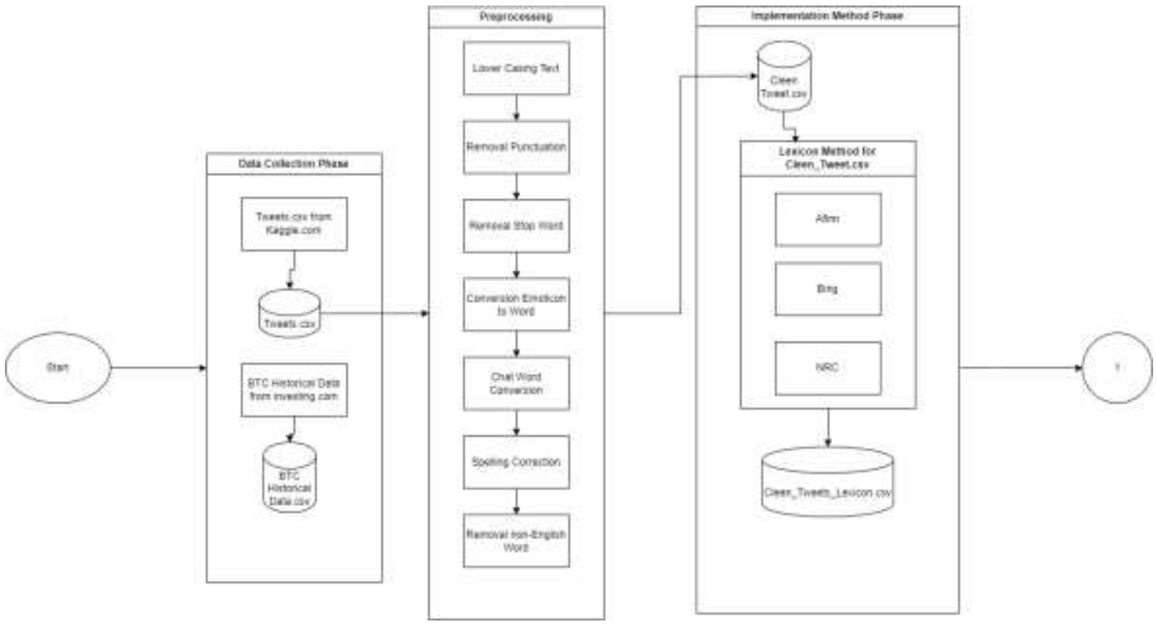

Fig. 1 Flowchart System

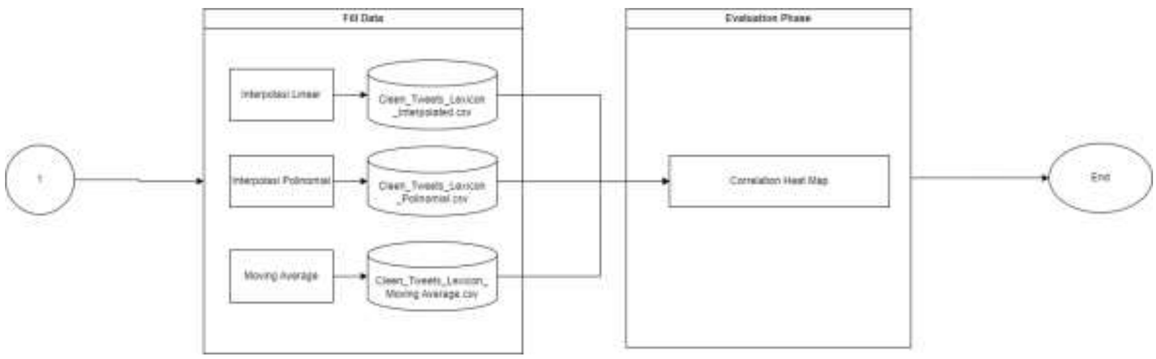

Fig. 2 Flowchart System

A. Afinn Method

The Afinn lexicon is a robust and flexible tool for sentiment analysis, providing effective sentiment assessment for texts on microblogging platforms. Developed by Finn Årup Nielsen, this lexicon rates words based on a valence scale from -5 to +5 and has shown higher correlation with human ratings compared to other lexicons like anew and General Inquirer. Its adaptability to multiple languages and spelling variations makes Afinn an excellent choice for sentiment analysis in diverse contexts [14].

Fig. 3 Afinn Method Scientific Arrchitecture

B. Bing Method

Bing is a sentiment analysis lexicon described by Bing Liu in his book "Sentiment Analysis and Opinion Mining". This lexicon includes a list of positive and negative words that are used to determine the sentiment of a text. Unlike Afinn, the Bing lexicon does not assign specific sentiment scores to each word but rather categorizes them simply as positive or negative. Bing Liu's approach is appreciated for its straightforward implementation and efficiency in providing a general sentiment assessment. This lexicon is particularly effective for analyzing texts with clear emotional polarity, making it a valuable tool for sentiment analysis tasks[15].

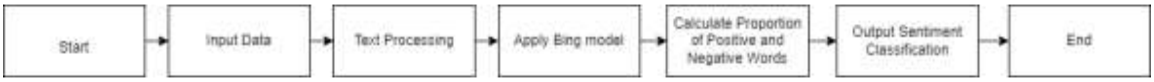

Fig. 4 Bing Method Scientific Arrchitecture

C. National Research Council (NRC) Method

NRC Emotion Lexicon (EmoLex) stands as a valuable resource for various NLP applications, facilitating the understanding of how specific words can evoke different emotions. It enables sophisticated emotion detection algorithms and contributes to cross-cultural and cross-linguistic studies on emotion associations. The lexicon's extensive coverage and high-quality data make it a reliable tool for both research and practical applications in analyzing emotional content in text [16].

Fig. 5 NRC Method Scientific Arrchitecture

III. Research Method

A. Data Collection

This research utilized 100,000 rows of data from Twitter posts and replies out of a total of 4,688,657 available rows. Given that 4,688,657 rows of data are too large to process, this study only used 100,000 rows of the total dataset. The data used spans from February 5, 2021, to April 24, 2021, and can be accessed on the Kaggle site. Likewise, BTC historical data uses data from the period of February 5, 2021, to April 24, 2021.

B. Preprocessing

Data preprocessing is a crucial step in data analysis aimed at enhancing the quality and usability of data before applying it to analytical models or machine learning algorithms. In the context of big data, preprocessing encompasses a range of techniques that help manage the high volume, variety, and velocity of data. Effectively applying these techniques can significantly improve model performance and result in more accurate and reliable analytical outcomes[17].

In summary, data preprocessing is an essential step to ensure that data is in an optimal condition for further analysis.

In this research, preprocessing involves several steps: converting to lowercase, removing punctuation, removing stop words, converting emoticons to words, chat words conversion, correcting spelling errors and removing non English words.

1) Lower Casing Text

In the first stage of data preprocessing, lowercasing is carried out. In text analysis, this process is an important step because text data can be diverse, including uppercase letters, lowercase letters, and special characters. By converting text to lowercase, the process and analysis become easier.

2) Removal Punctuation

Removing punctuation is an important step. In text, especially in social media posts, punctuation is often used differently in various contexts. By carrying out the punctuation removal stage, the text can be standardized, making it easier to analyze.

3) Remove Stop Words

The removal of stop words was done because these words are frequently used but do not provide significant meaning or information for analysis. By removing stop words, noise can be reduced, and the analysis can focus more on words that have meaning. This stop word list uses natural language toolkit (NLTK).

4) Conversion Emoticon to Words

Converting emojis into words is an important step because emojis often convey actions, emotions, or objects that are crucial for understanding the context of a message. By converting emojis into words, analytics can capture additional meanings conveyed by the emojis and improve the performance of text analysis models.

5) Chat Words Conversion

Converting chat words into standard language is crucial because chat words, slang, and various abbreviations are often difficult to understand outside of specific contexts. By converting these terms into standard language, we can ensure text consistency and improve the accuracy of the analysis.

6) Spelling Correction

Correct spelling supports data capture systems and reflects data consistency, all of which are critical for obtaining accurate analysis.

7) Removing Non English Words

Because the focus of this study is on Twitter post data in English, removing non-English words is an important step to maintain data accuracy and quality. Non-English words can lead to inaccurate results or interpretations of sentences.

C. Implementation Method Phase

Once data preprocessing is complete, the data is ready for sentiment analysis. In this study, sentiment analysis was performed using the Lexicon sentiment, Lexicon sentiment analysis is a method that utilizes a dictionary of words with pre-assigned sentiment scores to determine the emotions or attitudes expressed in a text. This approach doesn't require training data, making it an unsupervised technique. The text is broken down into tokens, and the sentiment scores of each token (positive, negative, or neutral) are aggregated to give an overall sentiment score for the text [13], applying three methods—Afinn, Bing, and NRC—to identify the most effective approach.

D. Fill Data

At this stage, missing data is filled using three methods: linear interpolation, polynomial interpolation, and moving average. Before the data filling process, the results of the sentiment analysis conducted earlier are grouped by day to facilitate the analysis.

1) Linear Interpolation

Linear interpolation is used to address short-term missing data by leveraging the values available around the missing data to estimate the accurate value. This technique works by connecting two known data points around the missing point with a straight line and using the value on this line as an estimate for the missing data[18]. To use this method, use the formula below.

$$y = y_1 + \frac{(x - x_1)(y_2 - y_1)}{(x_2 - x_1)} \tag{1}$$

Description :

(x_1, y_1) and (x_2, y_2)	= Two known data points
x	= The point between x_1 and x_2 where we want to calculate the interpolation value.
y	= The interpolated value at x.

2) Polynomial Interpolation

Polynomial interpolation is a numerical method used to estimate unknown values that fall within the range of known data points. The method involves constructing a polynomial that passes exactly through a given set of points. To use this method, use the formula below [18].

The general form of the polynomial of degree nnn is:

$$P(x) = \sum_{i=0}^{n} y_i \prod \frac{0 < j < n}{j \neq i} \frac{x - x_j}{x_i - x_j} \tag{2}$$

Description :

$P(x)$	= Polynomial Interpolation.
n	= Number of data points.
(x_i, y_i)	= Known data points

However, power polynomials have limitations, such as the potential for creating fictive extremes, especially when the number of interpolation nodes is small or the distance between them is large. As a result, trigonometric polynomials based on the summation of Fourier series are often preferred for applications requiring better approximation [19].

3) Moving Average

The moving average method technique for smoothing historical data and predicting future values. This method assigns higher weights to recent data, reducing the influence of older data to produce a moving average that is more responsive to recent changes. It is particularly useful in monitoring processes with autocorrelated data, enhancing the detection of small shifts in the process with greater accuracy[20]. To use this method, use the formula below.

$$M\ A_k = \frac{1}{N} \sum_{i=k}^{k+N-1} x\_i \tag{3}$$

Description :

$M\,A_k$ = Moving average at point k.
N = Window length (number of data points in the subset).
x_i = Data value at point i.

E. Evaluation Phase

Correlation is a statistical method used to assess the relationship between two variables. The correlation coefficient (r) quantifies the strength and direction of this relationship, ranging from -1 to 1. A positive correlation implies that as one variable increases, the other also increases, while a negative correlation means that as one variable increases, the other decreases [21][22].

Interpretation of the Correlation Coefficient

r = 1: Perfect positive correlation, indicating a direct relationship.
r = -1: Perfect negative correlation, indicating an inverse relationship.
r = 0: No correlation, indicating no linear relationship.

Values near 1 or -1 suggest a strong relationship, while values close to 0 suggest a weak relationship [21].

This stage is the final step in the analysis process for this study. The results of combining sentiment analysis data with historical crypto price data will be evaluated to determine the correlation between Twitter posts and Bitcoin prices on the same day. This study uses the correlation formula given by Equation 4.

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \tag{4}$$

Description

r = is the pearson correlation coefficient
x_i and y_i = are the values of the two variables
$\bar{x}$ and $\bar{y}$ = are the means of the two variables

Correlation is essential for understanding relationships between variables, aiding in prediction and pattern identification. However, it is crucial to remember that correlation does not imply causation, and confounding variables might affect the results.

IV. RESULTS AND DISCUSSION

A. Preprocessing Implementation

In this study, several pre-processing stages were carried out: converting text to lowercase, removing punctuation, eliminating stop words, converting emoticons to words, standardizing chat words, correcting spelling errors, and removing non-English words. This section will explain each pre-processing stage in detail to ensure the accuracy of the resulting data.

1) Lowercase, Removing Punctuation, Eliminating Stop Words

The first stage is lowercasing, which involves changing all text to lowercase without exception. Next is the process of removing punctuation, which includes characters such as ~!@#$%^&*()-=[]<> as well as spaces, numbers, and uniform resource locator (URL)s. Following that, the stop word removal stage is carried out, eliminating words like “blah,” “ble,” and “blu” that are considered unimportant.

TABLE I
RESULT TABLE FOR LOWERCASE, REMOVING PUNCTUATION AND ELIMINATING STOP WORDS

Before	After
Guys evening, I have read this article about BTC and would like to share with you all – https://t.co/QxCZgmuy3B...https://t.co/o6wn7ppkVY	guys evening read article btc would like share httpstcoqxczgmuy3b…httpstcoo6wn7ppkvy

$BTC A big chance in a billion! Price: \4872644.0 (2021/02/11 08:51) #Bitcoin #FX #BTC #crypto	btc big chance billion price 48726440 20210211 0851 bitcoin fx btc crypto
This network is secured by 9 508 nodes as of today. Soon, the biggest bears will recognise: #BTC in too big to fail… https://t.co/1XovDA8rKw	network secured 9 508 nodes today soon biggest bears recognise btc big fail… httpstco1xovda8rkw

2) Convert Emoticon to Words

This stage aims to convert emoticons into words. For example, the illustration of this conversion process can be seen in Table II.

TABLE II
ILLUSTRATION TABLE FOR CONVERT EMOTICON TO WORDS

Before	After
I am so D:	I am so sadness
Good morning :*	Good morning kiss
What happen!! oO	What happen!! surprised

3) Chat Words Conversion

This stage involves converting slang words or abbreviations into their full forms. The goal is to avoid ambiguity in the text and ensure data accuracy. The process of converting abbreviations or slang words to their original forms can be seen in Table III.

TABLE III
ILLUSTRATION TABLE FOR CHAT WORDS CONVERSION

Before	After
I will brb	I will Be Right Back
HI! WB	Hi! Welcome Back
THAT’S SO FUNNY LOL	THAT’S SO FUNNY Laughing Out Loud

4) Correcting Spelling Errors

At this stage, corrections are made to writing and spelling errors in the data to avoid ambiguity in the context of the text. The process of correcting spelling errors is shown in Table IV.

TABLE IV
ILLUSTRATION TABLE FOR CORRECTING SPELLING ERRORS

Before	After
I will brb	I will Be Right Back
Come here ASAP	Come here As Soon As Possible
That’s so funy LOL	That so fun Laughing Out Loud

5) Removing Non-English Words

To ensure data accuracy, non-English words will be removed since this study focuses on Twitter posts in English. Thus, non-English words are considered noise that can interfere with and reduce data accuracy. This process is shown in Table V.

TABLE V
ILLUSTRATION TABLE FOR REMOVING NON-ENGLISH WORDS

Before	After
My name rahman, I love kucing	My name , i love

Cepat datang, why so long	, why so long
Please deh, don't do that again	Please, don't do that again

B. Implementation Method Phase

In the implementation phase of the lexicon, weighting the preprocessed data will utilize AFINN, Bing, and NRC for scoring.

TABLE VI
RESULT TABLE FOR IMPLEMENTATION LEXICON METHOD

Date	Text	Afinn_value	Bing_value	NRC_value
2021-02-10 23:53:30	trade crypto binance enjoy cashback 10 trading…	2	2	7.0
2021-02-10 23:53:17	ltfire amp mangt bitcoin crypto btc httpstcocv…	0	0	0.0
2021-02-10 23:52:42	prices update eur 1 hour btc 370821 € 051 eth…	0	0	0.0
2021-02-10 23:52:25	btc bitcoin ethereum eth crypto cryptotrading…	0	0	0.0
2021-02-10 23:52:08	tesla ‘ bitcoin investment revolutionary crypt…	0	1	4.0

C. Fill data

After performing sentiment analysis using the Lexicon method with Afinn, Bing, and NRC, the data were consolidated into the same date range to facilitate further analysis. Following this consolidation, it was observed that some data points were still missing, as shown in Table VII, where data for February 11 and 12, 2021, were absent. Therefore, this stage involves filling in the missing data.

In this process, three methods were tested: linear interpolation, polynomial interpolation, and moving average, in order to determine the most effective approach.

TABLE VII
RESULT TABLE BEFORE PROCESS FILL DATA

date	Afinn_value	Bing_value	NRC_value
2021-02-10	0.723766	0.260352	2.485536
2021-02-11	NaN	NaN	NaN
2021-02-12	NaN	NaN	NaN
2021-02-13	0.621882	0.242013	2.537418
2021-02-14	0.997256	0.386381	2.799950

TABLE VIII
RESULT TABLE AFTER LINEAR INTERPOLATION PROCESS

date	Afinn_value	Bing_value	NRC_value
2021-02-10	0.723766	0.260352	2.485536
2021-02-11	0.689805	0.254239	2.502830
2021-02-12	0.655843	0.248126	2.520124
2021-02-13	0.621882	0.242013	2.537418
2021-02-14	0.997256	0.386381	2.799950

TABLE IX
RESULT TABLE AFTER POLYNOMIAL INTERPOLATION PROCESS

date	Afinn_value	Bing_value	NRC_value
2021-02-10	0.723766	0.260352	2.485536

2021-02-11	0.623417	0.223454	2.493981
2021-02-12	0.482024	0.183212	2.436158
2021-02-13	0.621882	0.242013	2.537418
2021-02-14	0.997256	0.386381	2.799950

TABLE X
RESULT TABLE AFTER MOVING AVERAGE PROCESS

date	**Afinn_value**	**Bing_value**	**NRC_value**
2021-02-10	0.723766	0.260352	2.485536
2021-02-11	0.722113	0.263394	2.429665
2021-02-12	0.723766	0.260352	2.485536
2021-02-13	0.621882	0.242013	2.537418
2021-02-14	0.997256	0.386381	2.799950

D. Evaluation Phase

After merging the data, we observed the results of the correlation between sentiment analysis data and historical Bitcoin data. Sentiment analysis was conducted using three lexicon-based methods: AFINN, Bing, and NRC. Each method will be tested to determine the most effective one.

1) Corelation Heatmap with Intepolation Linear

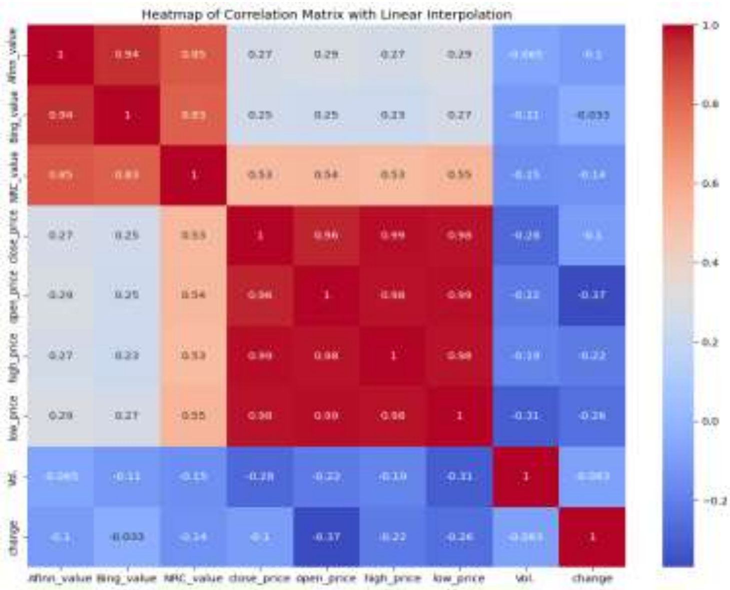

Fig .6 Corelation Heatmap with Linear Interpolation

NRC_value : Moderate positive correlation with prices (close_price: 0.53, open_price: 0.54, high_price: 0.55, low_price: 0.55).

Afinn_value: Weak positive correlation with prices (close_price: 0.27, open_price: 0.29). Bing_score: Weak positive correlation with prices (close_price: 0.25, open_price: 0.25).

2) Corelation Heatmap with Polinomial Interpolation

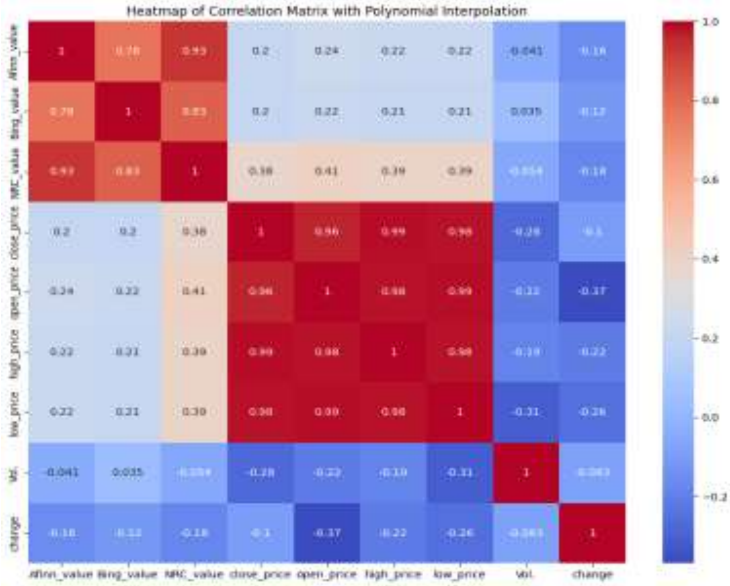

Fig. 7 Corelation Heatmap with Polinomial Interpolation

NRC_value: Moderate positive correlation with prices (close_price: 0.38, open_price: 0.41, high_price: 0.39, low_price: 0.39).
Afinn_value: Weak positive correlation with prices (close_price: 0.20, open_price: 0.24).
Bing_score: Weak positive correlation with prices (close_price: 0.20, open_price: 0.22).

3) Corelation Heatmap with Moving Average

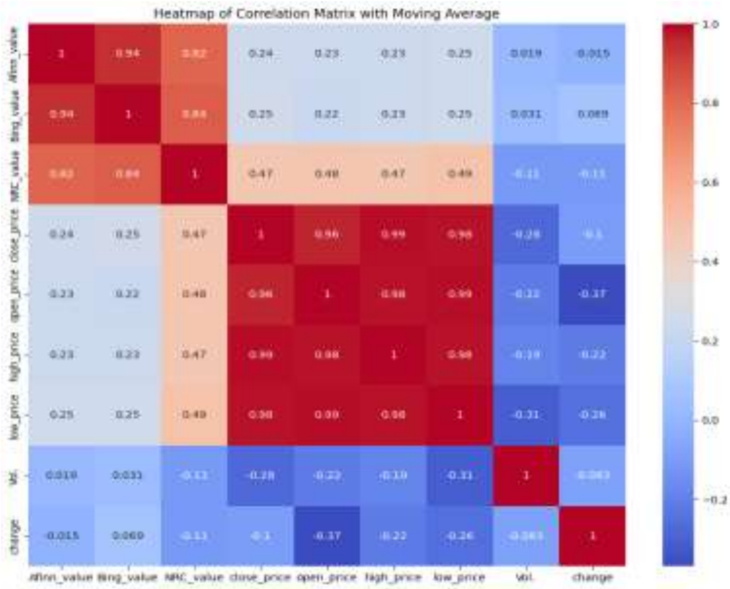

Fig. 8 Corelation Heatmap with Polinomial Interpolation

NRC_value: Moderate positive correlation with prices (close_price: 0.47, open_price: 0.48, high_price: 0.47, low_price: 0.49).
Afinn_value: Weak positive correlation with prices (close_price: 0.24, open_price: 0.23).
Bing_score: Weak positive correlation with prices (close_price: 0.25, open_price: 0.22).

NRC_value consistently shows a moderate positive correlation with price variables across all interpolation methods, with the highest correlation observed in linear interpolation. This indicates that the sentiment captured by NRC_value has a noticeable influence on Bitcoin price fluctuations.

Afinn_value and Bing_value exhibit weak positive correlations with price variables, indicating a minimal influence on Bitcoin prices. The correlations are slightly higher in linear interpolation compared to polynomial and moving average methods.

NRC_value with Linear Interpolation, NRC_value Shows a moderate positive correlation with Bitcoin price variables:

- Correlation with closing price (close_price): 0.53
- Correlation with opening price (open_price): 0.54
- Correlation with highest price (high_price): 0.55
- Correlation with lowest price (low_price): 0.55

This indicates that the sentiment measured by NRC_value has a relatively strong relationship with Bitcoin price fluctuations, suggesting that the emotional value captured by NRC_value can be used as an indicator for predicting Bitcoin prices.

Therefore, NRC_value with linear interpolation provides a clear view of the moderate relationship between emotional sentiment and Bitcoin prices, showing that changes in NRC_value tend to be followed by similar directional changes in Bitcoin prices.

V. CONCLUSION

This research indicates that there is a moderate positive correlation between public sentiment, as measured by NRC_value, and BTC trading behavior. Specifically, the correlation values are as follows: 0.53 with closing price, 0.54 with opening price, 0.55 with the highest price, and 0.55 with the lowest price. These findings suggest that the emotional value captured by NRC_value has a relatively strong relationship with Bitcoin price fluctuations, indicating that it can be used as an indicator for predicting Bitcoin prices.

However, while NRC_value demonstrates a clear and moderate relationship between emotional sentiment and Bitcoin prices, it is important to note that this correlation is not strong enough to act as a sole predictor of market behavior. This underscores the importance of considering other factors in sentiment analysis to improve prediction accuracy. Therefore, it is crucial to integrate a multi-faceted approach that includes additional variables and external factors, such as macroeconomic indicators, regulatory changes, and technological advancements in blockchain technology, to better understand and predict BTC trading patterns.

Further research is needed to identify and incorporate additional variables that could strengthen this correlation. For instance, macroeconomic indicators, regulatory changes, and technological advancements in blockchain technology could also impact BTC trading behavior and should be considered in future studies. Additionally, the dynamic and volatile nature of cryptocurrency markets necessitates continuous monitoring and updating of predictive models to ensure their relevance and accuracy over time.

Moreover, the use of advanced machine learning techniques and more comprehensive sentiment analysis tools may provide deeper insights into the nuances of public sentiment and its impact on BTC trading. By expanding the scope of research to include these elements, we can develop more robust models that capture the complexity of the cryptocurrency market.

In conclusion, while this study provides valuable insights into the relationship between public sentiment and BTC trading behavior, it also underscores the need for a more holistic approach in future research. By doing so, we can enhance our understanding and prediction capabilities, ultimately leading to more informed trading strategies and better decision-making in the cryptocurrency market.

REFERENCES

[1] A. Olbrecht and G. Pieters, “Crypto-Currencies and Crypto-Assets: An Introduction,” Apr. 01, 2023, *Palgrave Macmillan*. doi: 10.1057/s41302-023-00246-1.

[2] S. Nakamoto, “Bitcoin: A Peer-to-Peer Electronic Cash System.” [Online]. Available: www.bitcoin.org

[3] K. John, M. O’Hara, and F. Saleh, “Bitcoin and Beyond,” *Annual Review of Financial Economics*, vol. 14, no. 1, pp. 95–115, Nov. 2022, doi: 10.1146/annurev-financial-111620-011240.

[4] K. Chen, Z. Duan, and S. Yang, “Twitter as research data,” *Politics and the Life Sciences*, vol. 41, no. 1, pp. 114–130, Aug. 2022, doi: 10.1017/pls.2021.19.

[5] M. van Klingeren, D. Trilling, and J. Möller, “Public opinion on Twitter? How vote choice and arguments on Twitter comply with patterns in survey data, evidence from the 2016 Ukraine referendum in the Netherlands,” *Acta Politica*, vol. 56, no. 3, pp. 436–455, Jul. 2021, doi: 10.1057/s41269-020-00160-w.

[6] O. Sattarov, H. S. Jeon, R. Oh, and J. D. Lee, “Forecasting bitcoin price fluctuation by twitter sentiment analysis,” in *2020 International Conference on Information Science and Communications Technologies, ICISCT 2020*, Institute of Electrical and Electronics Engineers Inc., Nov. 2020. doi: 10.1109/ICISCT50599.2020.9351527.

[7] V. Koltun and I. P. Yamshchikov, “Pump It: Twitter Sentiment Analysis for Cryptocurrency Price Prediction,” *Risks*, vol. 11, no. 9, Sep. 2023, doi: 10.3390/risks11090159.

[8] C. Chen, L. Liu, and N. Zhao, “Fear Sentiment, Uncertainty, and Bitcoin Price Dynamics: The Case of COVID-19,” 2020. [Online]. Available: https://coinmarketcap.com/

[9] D. Güler, “The Impact of Investor Sentiment on Bitcoin Returns and Conditional Volatilities during the Era of Covid-19,” *Journal of Behavioral Finance*, vol. 24, no. 3, pp. 276–289, 2023, doi: 10.1080/15427560.2021.1975285.

[10] I. I. Japar, D. Pathmanathan, U. Malaya, and S. Azman, “Bitcoin price uctuation by Twitter sentiment analysis pre-and post-COVID-19 pandemic,” 2022, doi: 10.21203/rs.3.rs-2288064/v1.

[11] A. Baruah and B. Changkakati, “VADER SENTIMENT ANALYSIS ON TWITTER: PREDICTING PRICE TRENDS AND DAILY RETURNS IN INDIA’S STOCK MARKET,” *Bangladesh Journal of Multidisciplinary Scientific Research*, pp. 45–54, Jul. 2024, doi: 10.46281/bjmsr.v9i2.2226.

[12] M. Frohmann, M. Karner, S. Khudoyan, R. Wagner, and M. Schedl, “Predicting the Price of Bitcoin Using Sentiment-Enriched Time Series Forecasting,” *Big Data and Cognitive Computing*, vol. 7, no. 3, p. 137, Jul. 2023, doi: 10.3390/bdcc7030137.

[13] A. Rufaida, A. Permanasari, and N. Setiawan, “Lexicon-Based Sentiment Analysis Using Inset Dictionary: A Systematic Literature Review,” European Alliance for Innovation n.o., Jul. 2023. doi: 10.4108/eai.5-10-2022.2327474.

[14] F. Årup and F. Nielsen, “General rights A new ANEW: Evaluation of a word list for sentiment analysis in microblogs A new ANEW: Evaluation of a word list for sentiment analysis in microblogs.” [Online]. Available: http://research.hypios.com/msm2011/http://www.imm.dtu.dk/~fn/

[15] B. Liu, “Sentiment Analysis and Opinion Mining,” *Synthesis Lectures on Human Language Technologies*, vol. 5, no. 1, pp. 1–167, May 2012, doi: 10.2200/S00416ED1V01Y201204HLT016.

[16] S. M. Mohammad and P. D. Turney, “CROWDSOURCING A WORD-EMOTION ASSOCIATION LEXICON,” 2012. [Online]. Available: http://crowdsourcing.typepad.com/cs/2006/06

[17] J. Luengo, D. García-Gil, S. Ramírez-Gallego, S. García, and F. Herrera, “Big Data Preprocessing Enabling Smart Data.”

[18] G. Huang, “Missing data filling method based on linear interpolation and lightgbm,” in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Feb. 2021. doi: 10.1088/1742-6596/1754/1/012187.

[19] A. Makarchuk, I. Kal’chuk, Y. Kharkevych, and A. Yakovleva, “The Usage of Interpolation Polynomials in the Studying of Data Transmission in Networks,” in *2020 IEEE 2nd International Conference on System Analysis & Intelligent Computing (SAIC)*, IEEE, Oct. 2020, pp. 1–4. doi: 10.1109/SAIC51296.2020.9239180.

[20] K. Mabude, J. Malela-majika, P. Castagliola, S. Shongwe, and M.-M. J-c, “Generally weighted moving average monitoring schemes: Overview and perspectives Generally Weighted Moving Average monitoring schemes-Overview and Perspectives,” *Qual Reliab Eng Int*, vol. 2021, no. 2, 2021, doi: 10.1002/qre.2765ï.

[21] SpringerLink, “Correlational Research,” in *The ECPH Encyclopedia of Psychology*, Springer, 2024.

[22] O. Academic, “Conducting correlation analysis: important limitations and pitfalls,” *Oxford Academic*, 2024.