**PREDICTING BITCOIN PRICE TRENDS USING**

**TWITTER SENTIMENT ANALYSIS**

by

Tung Pham Duy, Nguyen Thanh Liem

Faculty of Finance and Banking, University of Economics and Law, Ho Chi Minh City, Vietnam.

Vietnam National University, Ho Chi Minh City, Vietnam.

*Submitted to Innovation & Sustainability - Impact on Economics & Business*

*International Conference*

Ho Chi Minh City, 2024

**TABLE OF CONTENT**

[1. INTRODUCTION 8](#_Toc162844936)

[2. LITERATURE REVIEW 9](#_Toc162844937)

[2.1. Background theories 9](#_Toc162844938)

[2.1.1. Market Sentiment theory 9](#_Toc162844939)

[2.1.2. Information Cascade theory 10](#_Toc162844940)

[2.1.3. Social Learning theory 11](#_Toc162844941)

[2.2. Literature review 12](#_Toc162844942)

[2.2.1. Sentiment analysis 12](#_Toc162844943)

[2.2.2. Previous researches 13](#_Toc162844944)

[3. METHODOLOGY 17](#_Toc162844945)

[3.1. Data 17](#_Toc162844946)

[3.2. Data cleaning and preprocessing 17](#_Toc162844947)

[3.3. Steps of conducting the study 20](#_Toc162844948)

[4. EXPLORATORY DATA ANALYSIS 20](#_Toc162844949)

[5. IMPLEMENT LSTM MODEL 26](#_Toc162844950)

[5.1. Construct LSTM model architecture and determine parameters 27](#_Toc162844951)

[5.2. Result of the LSTM model for sentiment classify 29](#_Toc162844952)

[5.3. Result of the LSTM model for crypto trend classify 30](#_Toc162844953)

[5.4. Discussion on the results of the LSTM model application 31](#_Toc162844954)

[6. INTRODUCING ADDITIONAL VARIABLES AND IMPLEMENTING MACHINE LEARNING MODELS 31](#_Toc162844955)

[6.1. Introducing additional variables 31](#_Toc162844956)

[6.1.1. Separated sentiment variables from VADER 32](#_Toc162844957)

[6.1.2. Variables from TextBlob 32](#_Toc162844958)

[6.1.3. Technical indicator variables 33](#_Toc162844959)

[6.2. Implementing machine learning models 35](#_Toc162844960)

[6.2.1. The aggregated datasets 35](#_Toc162844961)

[6.2.2. Result of machine learning models 37](#_Toc162844962)

[7. CONCLUSION 41](#_Toc162844963)

[8. LIMITATION AND FURTHER RESEARCH 42](#_Toc162844964)

[REFERENCE 43](#_Toc162844965)

**ABSTRACT**

This study aimed to investigate whether Twitter sentiment influences Bitcoin price fluctuations. Our approaches ranged from using common correlation verification methods (Pearson, Kendall, and Spearman cross-correlation) to more manual and in-depth methods, such as constructing specialized deep learning models for text processing to classify tweet content at the character level (building and implementing LSTM model). Additionally, we employed specialized NLP models with high accuracy to determine user sentiment. The results achieved high accuracy rates of 73.1% and 76.1% for the AdaBoost and XGBoost models, respectively, utilizing combined data from tweet user sentiment, proposed custom coefficients, and technical indicators.

*Keyword: sentiment analysis, bitcoin, twitter, price prediction*

**ABBREVIATIONS**

|  |  |
| --- | --- |
| **Abbreviation** | **Full Name** |
| AdaBoost | Adaptive Boosting |
| ADL | Accumulation/Distribution Indicator |
| Aroon | Aroon Indicator |
| CV | Cross-Validation |
| DT | Decision Tree |
| F1-score | F1 Score |
| KNN | K-Nearest Neighbors |
| LSTM | Long Short-Term Memory |
| MLP | Multilayer Perceptron |
| NLP | Natural Language Processing |
| NB | Naive Bayes |
| PA | Passive Aggressive |
| RC | Ridge Classifier |
| RF | Random Forest |
| SGD | Stochastic Gradient Descent |
| SVC | Support Vector Classifier |
| VADER | Valence Aware Dictionary and sEntiment Reasoner |
| XGBoost | Extreme Gradient Boosting |

**TABLES AND FIGURES**

|  |  |
| --- | --- |
| **Table** | **Description** |
| Table 1 | The aggregated datasets including data from score index, sentiment coefficients from models, and technical indicators |
| Table 2 | Summarizes the comparison of scores among the models and datasets |

|  |  |
| --- | --- |
| **Figure** | **Description** |
| Figure 1 | Stop words and their respective volume within the tweet content correspond to the display size. |
| Figure 2 | Top 20 most frequently used words within the tweet content. |
| Figure 3 | The compound score and the price of bitcoin. |
| Figure 4 | The distribution ratio of user sentiment (left) and the distribution ratio of bitcoin price movement (right). |
| Figure 5 | The volume of each tweet sentiment and the price of bitcoin. |
| Figure 6 | The sentiment score and the price of bitcoin. |
| Figure 7 | Compare the correlation between bitcoin price movement and sentiment score over lag. (from left to right: Pearson, Kendall, and Spearman) |
| Figure 8 | The architecture of the LSTM model. |
| Figure 9 | Result of the LSTM model for sentiment classify. |
| Figure 10 | Result of the LSTM model for crypto trend classify. |
| Figure 11 | The calculated values of the Aroon Up and Down. |
| Figure 12 | The calculated values of the Accumulation/Distribution Line. |
| Figure 13 | Results of the accuracy on the sentiment dataset |
| Figure 14 | Results of the F1-score on the sentiment dataset. |
| Figure 15 | Results of the accuracy on the sentiment dataset with technical indicators. |
| Figure 16 | Results of the F1-score on the sentiment dataset with technical indicators. |

# **1. INTRODUCTION**

The introduction of Bitcoin by Satoshi Nakamoto (Nakamoto, 2008) has brought a significant transformation in the financial system, as it introduced a decentralized electronic money system. This exemplifies advancements in the realm of information technology, particularly focusing on peer-to-peer networks and encryption techniques. Bitcoin operates independently from any government or bank, as a consequence of its decentralized nature and electronic system. The primary objective of Bitcoin is to facilitate transactions for goods and services. Bitcoin has undergone significant changes and has successfully attracted a large user base, becoming extremely popular due to frequent media coverage and widespread adoption. Due to its broad appeal, the price of Bitcoin is constantly changing in real-time, much like the stock exchange. It is intriguing to explore the possibility of developing a model that could forecast Bitcoin's price using data from internet social media. If it can be predicted with a reasonable degree of accuracy, it can be valuable for investors, companies, banks, organizations, and others who rely on Bitcoin for transactions.

The Internet has undergone significant changes in the last decade, making it easier than ever to share information and experiences through various communication networks like Facebook, Instagram, Twitter, blogs, and more. Every day, a massive volume of Bitcoin-related post tweets data is being generated by millions of Twitter users. This vast amount of data can be valuable for studying Bitcoin trends through advanced technologies like Natural Language Processing, Deep Learning and Time Series Analysis... leverage the constantly updated real-time data from social media platforms. Given the novelty of Bitcoin, there is still much progress to be made in applying Natural Language Processing along with Deep learning techniques to analyze social media data with improved accuracy and speed. While there has been extensive research on using machine learning techniques to predict time series, there is a noticeable gap in the literature when it comes to applying these techniques specifically to Bitcoin. This study aims to address the existing gap by creating a model that could accurately predict the future price trend of Bitcoin using Twitter sentiment analysis. To accomplish this objective, we gather Bitcoin-related Tweets and extract their price histories prior to conducting sentiment analysis on the collected data. In the next section we will focus on establishing background theories explaining why sentiment analysis affects the value of Bitcoin and why Twitter (currently renamed as X) has been chosen as the platform for determining the psychology of investors in the crypto market.

# **2. LITERATURE REVIEW**

## **2.1. Background theories**

### **2.1.1. Market Sentiment theory**

The Market Sentiment theory explores the domain of investor psychology, with a particular focus on the significant influence of emotions, perceptions, and crowd behavior on the dynamics of financial markets. Fundamentally, this theory posits that the sentiment of investors, as opposed to just relying on logical evaluations of fundamental facts, can exert a substantial influence on asset prices. One of the fundamental elements of Market Sentiment Theory is emotional influence. Conventional economic theories assume that investors engage in rational decision-making processes wherein they make choices based on a comprehensive examination of the information at their disposal. The Market Sentiment Theory acknowledges that investment decisions are frequently influenced by human emotions, including fear, greed, optimism, and pessimism. These emotional responses have the potential to induce investors to exhibit either excessive or insufficient reactions to news and events, resulting in amplified price fluctuations that diverge from the underlying basic valuations.

Market Sentiment Theory recognizes the inclination of investors to imitate the behaviors of their peers, a phenomenon sometimes referred to as herd behavior. The amplification of market movements and the formation of trends and bubbles can occur when a substantial proportion of market players adopt a common investment strategy or exhibit similar reactions to news. Herd behavior frequently arises from the apprehension of being left out (FOMO) or the apprehension of experiencing a loss, leading investors to make decisions based on the actions of others rather than conducting independent analysis. Market emotion has the potential to establish self-reinforcing feedback loops that sustain patterns in asset prices. Positive feeling has the potential to attract a greater number of investors, so resulting in an increase in prices and further strengthening bullish sentiment. On the other hand, the presence of negative sentiment has the potential to initiate selling pressure, resulting in more decreases in prices and strengthening the prevailing pessimistic sentiment. Feedback loops have the potential to intensify market volatility and contribute to the phenomenon of price momentum in the immediate term.

When applying in the case of Twitter new and Bitcoin, the Market Sentiment Theory provides significant observations regarding the intricacies of the cryptocurrency market. Bitcoin is known for its decentralized structure and speculative allure, making it highly vulnerable to fluctuations in market sentiment. Twitter, being a widely used site for engaging in conversations and deliberations over Bitcoin, provides an ideal environment for the articulation of investor sentiment.

### **2.1.2. Information Cascade theory**

Information cascade theory describes a scenario in which individuals' decision-making processes are influenced by the behavior of others, as opposed to being solely based on their own knowledge. There are two core conditions suggested in this theory. Firstly, individuals navigate choices sequentially, encountering analogous situations one after the other. Secondly, while possessing some private information, individuals heavily lean on observing the decisions and outcomes of those preceding them, indicating a reliance on external cues to inform their own choices.

In the context of Tweet sentiment influencing Bitcoin price, the theory could be applied within the following process. Consider a scenario where individuals are contemplating whether to buy or sell Bitcoin. This decision-making process aligns with the first component of information cascades: a decision to be made. In this case, it's whether to engage in Bitcoin transactions.

Suppose influential figures or large numbers of users on social media platforms like Twitter start expressing overwhelmingly positive sentiments about Bitcoin's prospects. This sentiment, whether genuine or exaggerated, can influence the decisions of others, echoing the second step of an information cascade where outside factors shape decisions. As more people observe these positive sentiments on Twitter, they may be inclined to follow suit and invest in Bitcoin, mirroring the behavior of those they observe, which is characteristic of the sequential decision-making process outlined in information cascades. Moreover, individuals considering Bitcoin investment may not have access to all the relevant information influencing others' decisions, including the sentiments expressed in tweets. However, they can infer the prevailing sentiment from the observable actions of those around them, aligning with the fourth and fifth components of information cascades. Thus, within this context, the sentiment expressed in tweets acts as a catalyst for information cascades, potentially driving waves of buying or selling activity in the Bitcoin market as individuals react to the perceived consensus expressed on social media.

### **2.1.3. Social Learning theory**

The Social Learning Theory, proposed by Albert Bandura in 1977, suggests that individuals learn from observing, imitating, and modeling the behavior of others within a social context. Applied to financial markets, this theory implies that investors may adjust their trading decisions based on the sentiments expressed by others, particularly in online social networks like Twitter. This phenomenon is often referred to as social learning in financial markets.

The theory was built upon the foundation of social learning mechanism by emphasizing the role of sentiment in driving market dynamics. According to this theory, investor sentiment can have a profound impact on asset prices, often leading to periods of irrational exuberance or pessimism. When investors exhibit positive sentiment towards an asset, it tends to drive demand higher, resulting in upward price movements. Conversely, negative sentiment can trigger selling pressure and drive prices lower.

Given the decentralized nature and largely unregulated environment of Bitcoin, traditional fundamental analysis methods may have limitations in accurately predicting its price movements. However, sentiment analysis of Twitter data offers a complementary approach by capturing the collective sentiment of the cryptocurrency community in real-time. By analyzing the sentiment of tweets related to Bitcoin in real time, researchers can gain insights into investor sentiment and market dynamics, which may help inform trading decisions and predict short-term price movements.

## **2.2. Literature review**

### **2.2.1. Sentiment analysis**

For a computer, raw natural language text lacks meaning and is simply represented as encoded bytes. In recent times, notable advancements have been achieved in the domain of Natural Language Processing (NLP) with the aim of augmenting the capacity of computational systems to comprehend and engage in logical reasoning using human language. Analysis of sentiment, as its name implies, involves the examination and extraction of sentiment, opinion, subjectivity, and polarity from textual data. Sentiment analysis has various applications, including analyzing product markets and automatically detecting good, negative, and potentially hazardous remarks on internet pages and social media platforms. Since its introduction to the natural language processing (NLP) field by Pang and Lee in 2008, Sentiment Analysis has been the subject of various proposed strategies aimed at establishing a connection between polarity and a given text. One potential methodology for addressing this issue is conceptualizing it as a classification problem, wherein a given text section is classified into positive, negative, or neutral categories. Additionally, certain methods include a metric that indicates the degree of certainty associated with the relevant mood. Lexicon-based methodologies employ lexicon and sentiment scores to engage in textual analysis and ascertain its comprehensive polarity score. VADER, a widely employed lexicon-based solution, employs rule-matching to ascertain polarity by examining linguistic patterns in the input text (Hutto and Gilbert 2015). In their study, Stenqvist and Lonno (2017) derived criteria for tweet intensity by analyzing the valence values, syntax, and grammar of 800 tweets. They then integrated these rules with specific features from three lexicons that had been validated. In the last ten years, sentiment analysis has gained significant popularity for examining Twitter data, posing distinct difficulties in comparison to conventional sentiment analysis applications. Several methodologies have been devised to address these difficulties. Furthermore, the "International Workshop on Semantic Evaluation" convention significantly contributed to the advancement of research by presenting a set of common challenges for the community to address.

### **2.2.2. Previous researches**

The 800 daily observations of thirteen distinct cryptocurrencies between January 1st 2018, and March 1st, 2020 were analyzed in a study by Caferra (2020). Finance herding and the potential convergence of opinion were the focal points of the study. A convergence of opinion and price development is pertinent to comprehending the correlation between opinion and trading price in crypto market sector, despite the fact that an examination of financial herding was not crucial to this investigation. Twitter is scraped for sentiment analysis purposes in the research of Pano and Kashef (2020)**.** To correlate the emotional evaluations of Tweet text with Bitcoin prices throughout the COVID-19 pandemic, the paper develops a variety of text preprocessing strategies. The objective is to ascertain the tweet preprocessing method that produces the most favorable correlation between sentiment scores and closing BTC prices. A tweet is categorized as negative, positive, or neutral by means of VADER, a sentiment analysis tool. The sentiment outcomes are assessed in relation to the BTC prices subsequent to the preprocessing stage. The objective of this paper is to make a scholarly contribution to the extant literature on BTC price forecasting. The following describes the process of preprocessing data: Utilizing Python, the analysis is conducted on the data collected via the Twitter API. A list of keywords was utilized to manually select the messages. This study gathered a cumulative sum of 4.169.709 tweets. The study's extensive data set (comprising more than four million tweets) is a virtue of the paper in comparison to others. The data was cleansed, and words were tokenized using various Python functions during the preprocessing phase, enabling VADER to more accurately capture distinct sentiment attributes. In order to conduct a more comprehensive comparison between the outlook for Bitcoin prices and the outcomes of sentiment analysis, the Pearson's correlation method was employed to examine the relationship between the development of Bitcoin prices and the scores of data from Twitter. During the two-month period from June 4th 2018 to August 4th 2018, Kraaijeveld and De Smedt (2020) gathered a total of 24 million tweets pertaining to the nine most prominent cryptocurrencies in 2018. The mean score per interval, which was calculated daily and hourly, was also applied using the sentiment assessment tool VADER. The Granger Causality test by Toda and Yamamoto was used to assess the correlation between sentiment and cryptocurrency on an hourly and daily basis.

To determine whether tweets affect the price of bitcoin and trading volume, Shen et al. (2019) investigate whether the frequency of tweets has an impact rather than sentiment analysis. The objective of the study was to examine the potential of Twitter data in conjunction with Bitcoin price forecasting. Employing sentiment analysis, the authors categorized Bitcoin-related tweets into positive, negative, or neutral categories between January 2014 and August 2018. After controlling for additional variables including fluctuation, trading volume, and the news sentiment, they employed a Granger causality evaluate to investigate the potential predictive power of Twitter sentiment on Bitcoin prices.

As an indication of the convergence of price expectations, Caferra (2020) research revealed that positive news had a substantial effect on diminishing the dispersion of returns. This finding indicates that investor sentiment significantly influences cryptocurrency prices, and that price fluctuations will eventually conform to market expectations. The Efficient Market Hypothesis (EMH), a widely recognized theory in the financial industry, posits that market prices will ultimately stabilize at levels that incorporate all pertinent information (Fama, 1970).

Caferra (2020) asserts that the impact could have been significantly influenced by the media's focus. The convergence of values and price expectations was facilitated by this informative signal. Price increases resulted when the media extensively reported positive developments concerning cryptocurrencies. According to the findings, in order to make informed choices regarding investments, it is critical that traders and investors remain current on the sentiment and media coverage surrounding cryptocurrency markets. Short-term results were also discovered by Pano and Kashef (2020). The cost of cryptocurrencies fluctuates in response to market sentiment. Precisely, the VADER model may be employed in practical Bitcoin price prediction for the immediate future. In contrast to the actual prices, machine learning models for forecasting could attain greater precision by employing the most effective preprocessing technique (for the data). Unpredictable correlation polarity renders the results inconclusive over the long term. Furthermore, the article poses the hypothesis that a significant increase in the correlation between Bitcoin prices and sentiment could be more pronounced when examined over a brief period of time (minutes as opposed to days). Even the quantity of tweets exhibits a correlation with the price movement of Bitcoin, Shen et al. (2019) discovered, rendering it sentiment independent. Statistically speaking, Twitter sentiment predicted Bitcoin returns for a maximum of two days in the short term, according to the authors. Nevertheless, with an extended prediction horizon, the prognostic efficacy of Twitter sentiment waned, suggesting that alternative variables might exert a more permanent impact on Bitcoin prices.

From June 4th, 2018 to June 9th, 2018, Kraaijeveld and De Smedt (2020) assess the eight largest cryptocurrencies besides Bitcoin. Distinct conclusions were derived at various intervals for each coin. In this discussion, solely the pertinent conclusions pertaining to Bitcoin is addressed. Twitter users are more likely to express positive or negative sentiments regarding Bitcoin when its daily price returns are high, according to the research. Nevertheless, there is no discernible indication of predictive capability in either direction during hourly intervals. This implies that Twitter is only responding to the price fluctuations of Bitcoin, and not inducing any Granger-Causal influence on the returns. Conversely, the daily analysis of Ethereum price returns indicated that sentiment expressed on Twitter lacks predictive capability. Nevertheless, a correlation was found between price returns and the volume of posts on the Twitter platform. As indicated by the lagged evaluation of the VAR model, price returns for specific delays during intraday and, in some instances, daily intervals are proportional to Granger-Cause message volume. This suggests that Twitter is merely responding to the price fluctuations of Ethereum, and not exerting a Granger-Causal influence on the returns. In the context of Cardano, the research revealed that Twitter sentiment exhibits a transient predictive impact on the volume of daily trading (lag 1). Kraaijeveld and De Smedt (2020) did not identify any intraday causal relationships.

John & Stantic (2022)utilized Granger Causality analysis in an attempt to predict cryptocurrency price fluctuations. To forecast the value of the cryptocurrency Dogecoin, the authors employ a time series of public opinion conveyed as the quantification of a vast collection of daily tweets (over 5 million in total from May 5 to May 31, 2021). An examination is conducted on the textual content of every tweet that makes reference to Dogecoin. This analysis is conducted utilizing a modified iteration of VADER, a sentiment polarity evaluation technique based on lexicons. A comparison is then made between the sentiment on Twitter that is generated and a time series representing the daily closing prices of Dogecoin. John and Stantic (2022) identified a unidirectional Granger-Causal relationship between Twitter sentiment and Dogecoin prices, with delays of two to four days. However, a bivariate connection between the variables was not observed. As confirmed by the Pearson correlation coefficient, Twitter and Dogecoin are correlated. Although the study's results appear convincing, we must approach their application with caution and skepticism. The data set encompasses a comparatively brief duration of 21 days, and the authors restrict their analysis to a single token, specifically Dogecoin. It is well-known that this token is highly sentiment-sensitive (Dablander, 2021). In addition, stationarity and autocorrelation tests for the variables have not been performed, which may result in erroneous regressions.

In conclusion, each paper tries to establish a correlation between sentiment and cryptocurrency prices. Whether cryptocurrency prices are correlated with sentiment or tweets volume is correlated with sentiment. Despite proving the existence of a correlation, the findings are uncertain with respect to the relationship's direction. In the next section, we will address the data used for the research, its content, issues related to the data, and their remedies, identifying methods and requirements for data preprocessing.

# **3. METHODOLOGY**

## **3.1. Data**

In this research, two main datasets will be utilized: (i) The dataset containing the daily price history and trading volume of Bitcoin, and (ii) the dataset containing detailed information of Tweets on Twitter. Regarding the Bitcoin price and volume dataset, the data is directly extracted from Yahoo Finance via the yfinance module integrated into Python. The data fields include information on Opening Price, Closing Price, Adjusted Closing Price, High Price, Low Price, and Trading Volume obtained on a daily basis. As for the dataset concerning the information of Tweets on Twitter, it is sourced from Kaggle. This dataset comprises raw data with 4,865,604 tweets extracted directly from Twitter using the static API connection method and the Tweepy library. The dataset is extracted based on the basic rule that the content of the tweet contains at least one of the 4 hashtags #BTC, #Bitcoin, #btc, #bitcoin or one of the 4 keywords "BTC", "Bitcoin", "btc", "bitcoin". The basic rule set during the querying process ensures that the extracted data is comprehensive, raw, and relevant to the referenced subject. Both datasets were collected within the time frame from February 2021 to March 2023, as there are three distinct trends during this period: an uptrend (February 2021 - November 2021, with a 239% increase), a downtrend (November 2021 - June 2022, with a 74% decrease), and a sideways movement (June 2022 - February 2023). However, due to hardware computational limitations, we will only utilize 1% of the collected dataset following the random sampling approach. And because this is raw data, there are still many issues to be addressed before processing them with the model. The next section will outline the steps needed to process the data cleanly for analysis and model identification purposes.

## **3.2. Data cleaning and preprocessing**

The vast amount of data on Twitter is uploaded by a diverse range of users and is used to serve various requests and purposes, such as advertising, propaganda, entertainment, etc. Additionally, a significant portion of posts on this platform are also posted automatically by algorithms, commonly known as “bot tweets”. Identifying, classifying, and filtering such tweets remains a challenging task for both the Twitter development team and researchers in our field. Currently, there is no reliable model developed and integrated into a library for classifying such tweets, other than models developed by individual researchers separately. Therefore, for this research, with the aim of preliminary filtering out tweets that are less likely to affect the psychology and behavior of investors, the author employs several data processing steps as follows:

* Eliminate tweets containing phrases or hashtags: Giveaway, Referral, as these tweets primarily serve communication and marketing purposes.
* Exclude tweets with similar content posted by the same user, reflecting the characteristics of a bot tweet.
* Remove tweets posted by users with fewer than 100 followers, which is a threshold determined by the author to eliminate tweets with low potential influence on the community and to partially filter out users who may be bots due to their low follower count.
* Elimination of non-English tweets.
* Elimination of characters that are not in the Latin alphabet.
* Elimination of URLs from messages
* Hash tag processing: removing words that were not present within English word following the hash sign. In all other cases, the word was retained with the '#' symbol omitted.
* User mention: Substituting the text 'USER' for user mentions (similar to labeling, which consists of '@' follows by a username)
* Tokenization: breaks down the stream of text into smaller units, it can be words, punctuation marks, numbers, or phrases depending.
* Elimination of punctuation
* Elimination of stop words, figure 1 illustrates the content and corresponding volume relative to the size of stop words contained within the tweet content.
* Lemmatization: converts words to their base or dictionary form.

A close up of words

Description automatically generated

*Figure 1: The stop words and their respective frequencies within the tweet content correspond to the display size.*

**A graph of a number of words

Description automatically generated**

*Figure 2: Top 20 most frequently used words within the tweet content.*

Figure 2 presents a statistical analysis of the top 20 most frequently used words within the tweet content. After the filtering and preprocessing stages, 35% of the total tweets have been removed due to not meeting the required conditions. For the processed data, it can be directly used as input for the machine learning models that we will implement in the subsequent sections.

## **3.3. Steps of conducting the study**

The content of this research will be structured as follows: In Section 4, we will conduct a exploratory data analysis to identify some basic characteristics of the dataset and a preliminary data analysis using the cross-correlation method to observe the correlation between Tweet sentiment and Bitcoin price. Section 5 will apply specialized data processing models for text, namely LSTM and BiLSTM, to assess the effectiveness of content classification of Tweets concerning Tweet sentiment and Bitcoin price trends. In Section 6, we will apply specialized Natural Language Processing models (VADER and TextBlob) to analyze Tweet sentiment, thereby utilizing the data obtained as an independent variable, and add some technical indicator data as additional input, to evaluate effectiveness through machine learning models. Section 7 will present the results obtained and discuss them. Section 8 will comment on the limitations of the research and propose directions for future studies.

# **4. EXPLORATORY DATA ANALYSIS**

Firstly, to gain a general understanding of the user sentiment in the tweets, we will use the VADER module to calculate the compound score for the content of each tweet and utilize it as a baseline. The compound score is a measure that computes the aggregate of every lexicon assessment, normalized within the range of -1 (indicating the most negative rating) to +1 (indicating the most positive rating). The acronym VADER module is short for "Valence Aware Dictionary and Sentiment Reasoner." It is a sentiment analysis application developed in particular of nltk python toolkit. VADER was developed with the purpose of sentiment analysis of text that is frequently encountered on social networking platforms. Informal speech, slang, and emoticons are included. VADER is also a Lexicon-Based methodology algorithm, meaning it comprises an extensive compilation of words pre-assigned scores for sentiment (positive, negative, or neutral). By analyzing the content for these Tweets and their combinations, the overarching sentiment can be determined.

A graph showing a graph of growth

Description automatically generated with medium confidence

*Figure 3: The comparison chart between the compound score and the price of bitcoin.*

The comparison chart between the compound score and the Bitcoin price shows in figure 3 has no clear correlation in either the medium or long-term cycles of the two indices. Regarding the compound score index, the integrated data, calculated as the sum per day for all three states (positive, negative, and neutral), does not reveal any period where the total compound sentiment of each tweet related to Bitcoin shows signs of negativity.

Based on the compound score just calculated, we can classify and label the tweets into three sentiment labels: positive, negative, and neutral. Tweets will be evaluated along one dimension and categorized into one of three main types based on the classification criteria outlined in the original documentation of the VADER model (Hutto & Gilbert, 2014):

* Positive: Score 0.05 or higher.
* Neutral: Score between -0.05 and 0.05 (essentially zero).
* Negative: Score lower than -0.05.

We obtain the proportions for the labels as depicted in figure 4, along with the distribution for the trend of Bitcoin price during the observed period.

A comparison of pie charts

Description automatically generated

*Figure 4: The distribution ratio of user sentiment (left) and the distribution ratio of bitcoin price movement (right).*

Based on figure 4, the classification results are as follows: negative: 11.94%, neutral: 45.72%, and positive: 42.34%. More than 87% of the tweets are classified as positive and neutral, while just over 11% show negativity towards Bitcoin. This observation aligns with the study by Critien et al., 2022, which utilized an LSTM model (different from VADER) to classify tweets from the period 2016-2019, contrasting with the timeframe of this study (2021-2023), yielding corresponding results of negative: 16.4%, neutral: 37.4%, and positive: 46.2%. As for the Bitcoin price, it fluctuates, and the lack of alignment between the closing prices of two consecutive days is evident, reflecting only upward and downward trends.

A graph of a bitcoin price

Description automatically generated

*Figure 5: The volume of each tweet sentiment and the price of bitcoin.*

Displaying the number of tweets by sentiment over time on figure 5, we can observe a consistent trend appeared in the pie chart. Tweets with negative sentiment only occupy a small percentage compared to a large overwhelming number of tweets with positive or neutral sentiment, even during periods of significant Bitcoin price declines. This is consistent with previous studies by Kraaijeveld et al., 2020, Critien et al., 2022, and Edgari et al., 2022. Although this imbalance in the data can be observed, it reflects the reality of the tweet dataset, which is a raw primary dataset directly extracted from the source.

However, the compound score and sentiment label only evaluate the sentiment of each individual tweet without considering the level of influence and impact of that tweet on other users. Therefore, we will utilize a new integrated metric that incorporates both factors: tweet sentiment and level of influence. To calculate the level of influence of a tweet, various factors come into play, such as the credibility of the poster, the number of followers, the level of engagement of the tweet, the number of views, shares, etc. After considering the factors affecting the level of influence of a tweet, the author proposes to use a formula referenced from Hasan et al., 2024, and simplified to compute the new metric below, termed as the "score":

Annotate for formula:

* score: The total score obtained, reflecting the sentiment and level of influence of the tweet.
* compound: The compound classification score from the VADER module.
* userfavourite: The number of users who have interacted with the post by favoring it.
* userview: The number of users who have viewed the post.
* isretweet: A binary value indicating whether the tweet is a retweet (1: retweeted tweet, 0: not a retweeted tweet).
* userfollower: The number of users following the posting account.
* userverified: A binary value indicating whether the account is verified (1: verified account, 0: unverified account).

Explanation for the coefficients used:

The compound value is the amplified angle coefficient based on the following multiplicative numbers. “userfavourite” represents users who interact with the original post, mostly sharing the same viewpoint and weighted with a magnification factor of 0.8 for the original viewpoint (Hasan et al., 2024). “userview” is often ten times the interaction rate (Riquelme & González-Cantergiani, 2016), (Rezaie et al., 2020), but the magnification weight is only a quarter for individual interactions (Hasan et al., 2024). The coefficient for retweets is a rather difficult measure to evaluate due to data on retweet counts being only available internally within Twitter, so for simplification, we set the amplification coefficient to 1 (binary: 1 or 0). Regarding “userfollower”, Hasan et al., 2024 also measured the level of influence as 1/1000 when analyzing accounts with high influence. The ratio of 1/1000 is also a basic measure for the Reach/Sales coefficient in marketing. For verified accounts with the blue tick (proving that the account holder is the legitimate owner and pays for the blue tick periodically to enjoy privileges for content creators), the coefficient is chosen to be amplified by over 30% compared to unverified accounts.

The data for the score variable for each tweet has been calculated, aggregated, and normalized by day along with the Bitcoin price, displayed in the following chart with the derivative additional for two variables below:

A graph showing a line graph

Description automatically generated with medium confidence

*Figure 6: The comparison chart between the sentiment score and the price of bitcoin.*

For figure 6, the “score” coefficient has displayed negative values within the normalized range from -1 to 1. However, due to the predominance of positive tweets, the community consistently expresses market optimism.

**A graph with blue lines

Description automatically generated**

*Figure 7: Compare the correlation between bitcoin price movement and sentiment score over lag. (from left to right: Pearson, Kendall, and Spearman)*

On the other hand, the cross-correlation test charts using the Pearson, Kendall, and Spearman methods indicate a positive correlation between the two indices after a cycle starting from day -5 onwards. This implies that if the price of Bitcoin shows positive or negative signs, then in the subsequent period from 0 to 5 days, the sentiment of Bitcoin on Twitter will similarly trend in the same direction and increase over that time span. This reflects investor psychology correlated with the rise or fall of the Bitcoin price. While not a causal relationship in this test, the results partly reinforce the existence of all three foundational theories: Market sentiment, information cascade, and social learning theory mentioned earlier. Observing the reverse trend, the chart shows a fixed trough at day -1, indicating a negative correlation in investor sentiment on Twitter before any change in the Bitcoin price. This means that if investors have positive or negative sentiment, the price fluctuation of Bitcoin will move in the opposite direction one day later. While the linear correlation measurement method of Pearson shows a decreasing trend from the preceding days up to day 1, the Kendall and Spearman methods with normalized data indicate a reversal in investor sentiment towards the Bitcoin price more than 10 days before a change in sentiment direction.

Based on the cross-correlation test results from figure 7, tweet sentiment shows a correlation one day before the fluctuation of the Bitcoin price. This will also be the fixed lag we choose for this study. Partly due to the short-term predictive nature of tweet sentiment and partly due to consistency with previous research, when compared across various time cycles, the one-day lag of tweet influence appears to be the most effective.

# **5. IMPLEMENT LSTM MODEL**

Based on the literature review and the preliminary data analysis conducted in the previous sections, we can observe that using a single aggregate index such as “compound” or “score” with one-dimensional analysis to determine the trend of Bitcoin price fluctuations is relatively challenging, although there exists a certain level of correlation between them. Therefore, in this section, we will delve into investigating and analyzing whether the content of tweets, specifically delving into the details of each word used in those tweets, has any correlation with the fluctuations of Bitcoin price. The specialized model for text data processing, LSTM, will be employed in this section.

## **5.1. Construct LSTM model architecture and determine parameters**

In this section, we will use the LSTM model to train and predict classifications for the content of tweets into 3 sentiment labels: positive, neutral, and negative. The goal of classifying tweet content at the word level that labelled with the baseline of VADER sentiment classification is to demonstrate that the model architecture is suitable for fully exploiting the content of each word in the tweet. If the model demonstrates good classification performance for sentiment labels, it indicates that tweets classified under the same label exhibit high linguistic similarity and achieve an optimal model architecture, therefore ready to classify tweet content regarding Bitcoin price fluctuations.

The LSTM deep learning model are highly suitable for analyzing text data due to their capability to handle sequential information and capture long-term dependencies within text. Text inherently possesses sequential properties, where words in a sentence follow a specific order, and the meaning of a sentence can heavily rely on the arrangement of words. Moreover, the long-term dependency characteristic of LSTM models distinguishes them from simpler models that only focus on nearby words. LSTMs have the ability to learn and retain information from earlier parts of the text, which is crucial because the meaning of a word can be influenced by words that appeared much earlier in the sentence. Therefore, LSTM models are a go-to choice for various Natural Language Processing (NLP) applications that rely on text analysis.

The architecture of the model is designed and described according to figure 8 below:

A screenshot of a computer

Description automatically generated

*Figure 8: The architecture of the LSTM model*

This model architecture is a neural network designed for sequential data processing, specifically for text classification, the components of the model are as follows:

Sequential Model: The model is constructed using the Sequential API in Keras, which allows stacking layers one after the other. The data is splited into 85% for trainning and 15% for testing.

Parameters:

* Batch Size (128): Balances computational efficiency and model stability. Smaller batch sizes may cause noisy updates. Larger batch sizes may require more memory and computational resources.
* Epochs (10): Allows model iterative learning without overfitting. Depending on dataset complexity and model convergence.
* Max Features (20,000): Limits vocabulary size to focus on relevant words. Prevents overfitting and reduces computational overhead.
* Embedding Dimension (100): Balances semantic information capture and computational efficiency. Higher dimensions may increase model complexity.

Layers:

* Embedding Layer: This layer is used to convert integer-encoded representations of words into dense vectors of fixed size. It transforms input sequences of length 30 into dense vectors of size 100.
* Conv1D Layer (with MaxPooling1D): This layers with 1D filters are used to capture local patterns in the input sequences. The conv1d layer applies convolutional filters with a kernel size of 3, resulting in 32 output channels. The subsequent max\_pooling1d layer performs max pooling operation to reduce the spatial dimensions of the output, resulting in sequences of length 15.
* Second Conv1D Layer (with MaxPooling1D): Similar to the previous convolutional layer, but applied to the output of the first max pooling layer. It further captures higher-level features from the input sequence, resulting in sequences of length 7.
* LSTM Layer: This is the main recurrent neural network (RNN) layer designed to capture long-term dependencies in sequential data. In this model, it processes the input sequences and outputs a vector of size 100.
* Dense Layer: Finally, a dense layer with 3 units and softmax activation is used for classification. This layer converts the output of the LSTM layer into a probability distribution over the three classes (positive, neutral and negative)

## **5.2. Result of the LSTM model for sentiment classify**

A graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of

Description automatically generated

*Figure 9: Result of the LSTM model for sentiment classify.*

The model output demonstrates strong performance in sentiment classification with an accuracy of 95.4%. The F1-scores for the positive, neutral, and negative classes are 87%, 97%, and 96%, respectively, indicating that the model achieves a good balance between precision and recall for each sentiment category. The model proves effective in capturing the nuances of sentiment expressed in tweets. Since the model has successfully demonstrated its ability to process and classify at the word level within each tweet, in the following section, we can utilize this model with the aim of classifying tweet content concerning Bitcoin price fluctuations to evaluate whether there is any correlation between these two variables at the word level.

## **5.3. Result of the LSTM model for crypto trend classify**

The model architecture, along with the parameters used, is similar to sentiment classification, with the only difference being the labeled data portion dedicated to classifying Bitcoin price fluctuations. The results of the model are obtained as shown in figure 10 below: A graph of a graph of a graph

Description automatically generated with medium confidence

*Figure 10: Result of the LSTM model for crypto trend classify.*

Although the model has achieved an accuracy of up to 95.4% for sentiment classification, it only achieved an accuracy of 53.7% for Bitcoin price fluctuations, slightly exceeding the rate of random guessing. The precision is 50% for positive and 57% for negative, with the F1-score at 55% and 52%, respectively. Furthermore, for the training dataset, the model has progressively become more proficient, with the loss approaching zero through each training iteration and the labeling accuracy reaching nearly 95%. However, for the test dataset, there appears to be a lack of textual similarity regarding cases of Bitcoin price increases or decreases. Therefore, the model demonstrates higher loss values over epochs and accuracy fluctuating around 53%.

## **5.4. Discussion on the results of the LSTM model application**

Based on the results of the LSTM model above, we can conclude that while sentiment classification is performed very well based on learning from labeled data, with an accuracy of 93.57% for the first test set, despite the model not fully learning from the training data (77.94%). With such high accuracy, the relationship between the vocabulary in the content of the tweets is shown to be similarly high, and the significant F1-scores for each group (87%, 97%, and 96%) indicate that the data imbalance does not significantly affect the model's prediction results. The model's loss index for the training set tends to increase with each epoch, but it does not affect the prediction capability. Limiting the number of epochs for the training process helps reduce computational costs while achieving equivalent results.

On the other hand, for tweet classification regarding Bitcoin price fluctuations, the model struggles to classify which word pairs are associated with which trends. With very few common patterns for such classification, the model predicts with a slightly higher probability than random guessing, despite having optimized training data over each epoch in contrast of the model's loss rate for the test set increasing rapidly. This suggests that there is indeed not much similarity in establishing which words will influence Bitcoin price fluctuations.

# **6. INTRODUCING ADDITIONAL VARIABLES AND IMPLEMENTING MACHINE LEARNING MODELS**

## **6.1. Introducing additional variables**

After examining the influence of Twitter user sentiment on the price of Bitcoin using two methods: (i) preliminary correlation analysis using three methods Pearson, Kendall, Spearman, and (ii) building an LSTM deep learning model to process raw text data from tweets and evaluate the affect on Bitcoin prices. Both methods have not yet demonstrated a significant impact of Twitter sentiment on Bitcoin prices in the short term. Therefore, in this section, we will supplement 3 new methods to determine whether Twitter sentiment truly affects the price of Bitcoin or not. The methods to be discussed include: (i) add 3 different variables, positive, neutral, and negative, from VADER; (ii) add 2 variables, subjectivity and polarity, from the TextBlob; (iii) add 3 technical indicators, Aroon Up, Aroon Down, and Accumulation/Distribution Indicator.

### **6.1.1. Separated sentiment variables from VADER**

While the VADER compound score provides a single measure of sentiment, it can be valuable to delve deeper and analyze the distribution of positive, neutral, and negative sentiment within a text. This allows for a more nuanced understanding of the emotional landscape expressed in tweets.

In this part, we decompose the VADER compound score into three separate variables: positive score, neutral score, and negative score. These scores represent the relative proportions of positive, neutral, and negative sentiment within a tweet. This division will allow us to analyze tweets from multiple dimensions rather than just one dimension of the compound index. A tweet can have from 1 to 3 indices with different weights. The compound index is removed due to multicollinearity problem concern, but the score index is still utilized as a variable, albeit calculated from the compound angle coefficient, it is combined with other data.

### **6.1.2. Variables from TextBlob**

Subjectivity and polarity, 2 more variables, are added from the TextBlob NLP module. The TextBlob text processing model allows for the computation of two indices: subjectivity, which is used to determine the level of subjectivity of a tweet (A score closer to 1 indicates the text is subjective and likely expresses personal opinions, feelings, or judgments. And a score closer to 0 suggests the text is objective and presents facts or factual information). The subjectivity score could potentially become a new variable to measure the influence of tweets sentiment on bitcoin price fluctuations, given the differences in classification rules and scoring of the TextBlob module. The polarity index is similar to the compound index of the VADER module (with a score closer to -1 indicates the text expresses a negative sentiment and a score closer to +1 suggests a positive sentiment), but for a different module, the classification principles will lead to differences result in determining the index compared to VADER.

### **6.1.3. Technical indicator variables**

The inclusion of 3 technical indicators, namely Aroon Up, Aroon Down, and Accumulation/ Distribution Indicator, is proposed. These three indicators are selected for different purposes. The Aroon indicator is chosen to determine the upward or downward trend of Bitcoin over time, although it has several limitations and is not widely used by investors because it focuses only on identifying direction (up or down) and time periods without considering amplitude fluctuations. However, the Aroon indicator is suitable for selection because it aligns with the research objectives. The secaond indicator, ADL (Accumulation/Distribution Indicator), can overcome the weaknesses of the Aroon indicator. Below are the formulas for the technical indicators Aroon Up and Down, along with the figure 11 computed results for bitcoin price data over the specified time period.

A graph with red and blue lines

Description automatically generated

*Figure 11: The calculated values of the Aroon Up and Down technical indicators.*

While Aroon focuses only on either increasing or decreasing trends, without considering amplitude fluctuations between cycles, ADL adds amplitude factors to trends and trading volume of Bitcoin in the market. Trends are also reflected in the ADL indicator, but similar to Aroon, ADL only confirms the trends that have occurred in specific time periods without providing future forecasts. In markets where trading volume and frequency are unpredictable, such as the crypto market, the ADL indicator is less commonly used. However, in this study, the selected indicators do not serve as primary forecasts but rather confirm the trend of Bitcoin within a specified time period, serving as a supporter for other variables to demonstrate their impact. Below are the formulas for the technical indicators ADL Indicator, along with the figure 12 computed results for bitcoin price data over the specified time period, where t is the mentioned month and all the others values is price of Bitcoin of the 1st of t month.

A graph with blue lines

Description automatically generated

*Figure 12: The calculated values of the Accumulation/Distribution Line technical indicators.*

These two technical indicators will be calculated over a time period of 30 days. For Aroon, this is the timeframe to determine whether new highs or lows are formed during this period. As for ADL, the value of the line is calculated on the first day of the month and applied to all subsequent days of the month. The reason we do not calculate the value of ADL on a daily basis is because the calculated result of ADL already encompasses information about the daily fluctuations of bitcoin, which would lead to the loss of support values and create a perfectly overfitted model. Instead, the initial values of each month are calculated, aiding the model in identifying the overall trend line of bitcoin prices, and providing a baseline for predicting the trend of bitcoin prices within that month.

## **6.2. Implementing machine learning models**

### **6.2.1. The aggregated datasets**

In this section, we will apply various machine learning models to train and make predictions based on the aggregated dataset. Table 1 displays some samples and fields of the aggregated dataset, with timestamp data for the index, columns for user sentiment metrics computed from the models, technical indicators, and finally the price trend of Bitcoin for the following day label.

Table 1: The aggregated datasets including data from score index, sentiment coefficients from models, and technical indicators

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Time stamp** | **score** | **subjectivity** | **polarity** | **Vader Pos** | **Vader Neg** | **Vader Neu** | **ADL** | **Aroon up** | **Aroon down** | **BTC trend** |
| 2021-02-05 | 0.000000 | 0.000000 | 0.000000 | 0.000 | 0.000 | 1.000 | 2.517522e+12 | 0.965517 | 0.482759 | positive |
| 2021-02-05 | 4717.234531 | 1.000000 | 0.300000 | 0.135 | 0.000 | 0.865 | 2.517522e+12 | 0.965517 | 0.482759 | positive |
| 2021-02-05 | 0.000000 | 0.000000 | 0.000000 | 0.000 | 0.000 | 1.000 | 2.517522e+12 | 0.965517 | 0.482759 | positive |
| 2021-02-05 | 0.000000 | 0.300000 | 0.200000 | 0.000 | 0.000 | 1.000 | 2.517522e+12 | 0.965517 | 0.482759 | positive |
| 2021-02-05 | 0.000000 | 0.466667 | 0.466667 | 0.000 | 0.000 | 1.000 | 2.517522e+12 | 0.965517 | 0.482759 | positive |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 2023-03-05 | -664304.990937 | 0.352778 | 0.120833 | 0.000 | 0.141 | 0.859 | 2.126499e+12 | 0.586207 | 0.689655 | negative |
| 2023-03-05 | 4408.622539 | 0.000000 | 0.000000 | 0.108 | 0.000 | 0.892 | 2.126499e+12 | 0.586207 | 0.689655 | negative |
| 2023-03-05 | 200.748913 | 0.142857 | -0.071429 | 0.133 | 0.072 | 0.795 | 2.126499e+12 | 0.586207 | 0.689655 | negative |
| 2023-03-05 | -47.916123 | 0.416667 | 0.166667 | 0.167 | 0.063 | 0.770 | 2.126499e+12 | 0.586207 | 0.689655 | negative |

*Soure: by author*

These 13 models include Dummy Model, Perceptron, Bernoulli Naive Bayes, Gaussian Naive Bayes, Decision Tree, Random Forest, K-Nearest Neighbors, Linear Support Vector Classifier (LinearSVC), Passive Aggressive, Ridge Classifier with Cross-Validation (RidgeClassifierCV), Stochastic Gradient Descent Classifier (SGDClassifier), AdaBoost Classifier và XGBoost. Each machine learning model has its own advantages and limitations, as well as different operating methods. We will test each model to evaluate their performance in predicting the price trend of Bitcoin based on the sentiment expressed in tweets. In this way, we will have an overview of the effectiveness of the machine learning models and identify the model that best suits our specific problem. The objective is to make evaluation and comparison of these machine learning models, while identifying the best-performing model for predicting the price trend of Bitcoin based on sentiment in tweets on Twitter.

However, to compare and evaluate the effectiveness of each set of variables separately, we will divide the process of training and testing machine learning models into two datasets: (i) the sentiment dataset, comprising a total of 6 sentiment metrics from NLP modules (score index, positive, neutral, and negative scores from VADER, and subjectivity and polarity scores from TextBlob); and the second dataset (ii) will include the first dataset (i) along with technical indicators (Aroon up, down, and ADL).

### **6.2.2. Result of machine learning models**

Figures 13 and 14 below illustrate the training and prediction results of the models on the sentiment dataset, comparing the accuracy and f1-score:

A graph with blue and white lines

Description automatically generated with medium confidence

*Figure 13: Results of the accuracy on the sentiment dataset.*

A graph of a bar graph

Description automatically generated with medium confidence

*Figure 14: Results of the F1-score on the sentiment dataset.*

Observing Figures 13 and 14, the average accuracy of the models is 51%, even though the models have been fine-tuned to find the optimal parameters. However, among the top 9 models with the highest accuracy, the accuracy rate still only reaches a maximum of 53%, slightly higher than random guessing. Compared to the LSTM model in the previous section, which achieved 53.7% accuracy, it seems that this number also reflects the optimization achieved in terms of architectural design and parameters of the LSTM model compared to standard sentiment extraction models like VADER and TextBlob. Once again, the model results reaffirm that using pure sentiment scores from tweets, without support or integration from external data sources, only leads to prediction capabilities slightly surpassing random chance, whether processing the model from raw primary data or applying available NLP modules. This conclusion aligns with the findings of Rezaie et al., 2020, Sattarov et al., 2020, Kraaijeveld & De Smedt, 2020, and Edgari et al., 2022.

Regarding the K-Nearest Neighbor model, it shows a prediction accuracy of only 41%. The algorithm of the K-Nearest Neighbor model requires referencing surrounding values to determine the prediction value (in this case, the model chooses n\_neighbors=10, which is the best coefficient when the author tried values of 1, 3, 5, 7, and higher coefficients), indicating that for an asset with strong and continuous volatility like bitcoin, predicting values based on neighboring values will be a poor trading strategy.

Figures 15 and 16 below illustrate the training and prediction results of the models on the sentiment dataset combined with technical indicators, comparing the accuracy and f1-score:

A graph of a bar graph

Description automatically generated with medium confidence

*Figure 15: Results of the accuracy on the sentiment dataset combined with technical indicators.*

A graph with blue and red lines

Description automatically generated

*Figure 16: Results of the F1-score on the sentiment dataset combined with technical indicators.*

Table 2: Summarizes the comparison of scores among the models and datasets

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Average Accuracy** | **Average F1-score** | **Max Accuracy** | **Max F1-score** |
| **Sentiments** | 51% | 41% | 52.9% | 49.5% |
| **Sentiments & Technical Indicators** | 59%  (SGD Classifier) | 57%  (Perceptron) | 76.1%  (XGBoost) | 81.5%  (XGBoost) |

*Source: by author*

For the integrated dataset combining sentiment and technical indicators, the training and prediction results are notably better. The average accuracy of all models increases to 59%, with AdaBoost and XGBoost achieving superior results compared to other models at 73.1% and 76.1%, respectively. Both XGBoost and AdaBoost are boosting algorithms, falling under the category of ensemble methods in machine learning. Which mean they could optimize the predictions from multiple weaker models to create a single, stronger prediction. In contrast, other models are trained independently and don't leverage the errors of previous models.

Regarding the technical indicator data, as mentioned earlier, they serve the purpose of identifying the overall long-term trend of Bitcoin, with a time cycle of 30 days. Therefore, the data from technical indicators do not reflect short-term fluctuations in Bitcoin price (1 day). Instead, they act as a baseline support threshold for the model to optimize the sentiment variables and predict the fluctuations during that time period.

# **7. CONCLUSION**

The aim of this study was to delve into understanding whether the sentiment of a post (tweet) on the social media platform Twitter truly impacts the fluctuation of Bitcoin price trends. The study delved into analyzing the textual content of tweets in three main ways. Firstly, it tested and measured the degree of correlation between user sentiment and Bitcoin price movement using three methods: cross-correlation, Pearson, Kendall, and Spearman, based on general sentiment scores calculated as "compound" and adjusted scores as "score". Secondly, it designed a model to identify, process, and classify text based on the LSTM deep learning model, with optimized architecture and parameters for identifying and classifying tweet content from raw data, aiming to find the relationship between tweet content, down to the level of detail of each vocabulary, and Bitcoin price movement. And thirdly, it applied specialized sentiment classification models, VADER and TextBlob, to evaluate whether these specialized classification models could generate a correlated data source with Bitcoin price movement, along with the support of technical indicator variables, to optimize the accuracy of predictions when inputting the dataset into 13 different machine learning models. After going through the three proposed methods in this study, the author draws the following conclusions:

(i) There exists a correlation between user sentiment and Bitcoin price movement; however, the impact on the trend of the two variables on each other is still not entirely clear.

(ii) The LSTM deep learning model has the ability to classify sentiment down to the level of individual words for tweet content effectively, demonstrating the model's quality in text processing. However, when training and predicting for price trend classification of Bitcoin, the model did not identify many common patterns, resulting in an accuracy level (53.7%) just slightly above random guessing. This suggests that there is no similarity between the content of the tweets posted and the Bitcoin price movement, consistent with the findings of (B & B, 2023) and (Sekhar et al., 2022) in the Literature Review.

(iii) The specialized models used to analyze sentiment in social media content also returned a similar level of accuracy (max accuracy: 52.9) when compared to the deep learning model. Once again, this indicates that there is no added advantage in using these NLP models to classify tweets without additional support from input data, consistent with the content of the studies by (Rezaie et al., 2020), (Sattarov et al., 2020), (Kraaijeveld & De Smedt, 2020), (McNally et al., 2018) and (Edgari et al., 2022).

(iv) With the new input data source being historical Bitcoin price data, describing general trends and trading volume conventions in the medium term, the machine learning models demonstrated prediction capabilities equivalent to previous studies (ranging from 55% to 65% with an average accuracy of 59%). The ensemble machine learning models AdaBoost and XGBoost exhibited the highest accuracy levels (73.1% and 76.1%) among all models. Several other studies have also yielded optimal results with the XGBoost model, such as Sekhar et al., 2022, Edgari et al., 2022.

# **8.** **LIMITATION AND FURTHER RESEARCH**

While the research provides valuable insights into the relationship between tweet sentiment and Bitcoin price trends, several limitations exist that warrant further investigation and refinement in future research endeavors:

Text-based analysis only: Our analysis primarily focuses on textual data extracted from tweets, neglecting the potential impact of multimedia content such as images, videos, and audio. Incorporating multimedia analysis techniques could provide a more comprehensive understanding of sentiment expressed on social media platforms like Twitter.

Language limitations: Our study currently analyzes sentiment in English-language tweets, overlooking sentiment expressed in other languages. Expanding the analysis to include multilingual sentiment analysis may offer a more inclusive perspective, especially considering the global nature of Bitcoin trading and social media usage.

Limited scope: While we concentrate on Twitter data, other sources of information such as news outlets, closed-group communities (e.g., pump-and-dump groups in the cryptocurrency space), and alternative social media platforms remain untapped. Future research could explore the sentiment expressed in these additional outlets to gain a more holistic view of market sentiment.

Time lag considerations: Our study adopts a fixed time lag of one day to analyze the impact of tweet sentiment on Bitcoin prices. However, the time lag between sentiment expression on social media and its influence on market prices may vary and could be influenced by factors such as the virality of tweets and market reaction times. Investigating different time lag intervals could provide insights into the dynamics of sentiment-driven market movements.

Narrow scope of influencers: Our analysis focuses on general sentiment expressed by Twitter users and does not specifically address the influence of key opinion leaders (KOLs), public figures, politicians, or tech CEOs. Future research could delve deeper into the impact of influential individuals and entities on market sentiment and price trends within the cryptocurrency space.

In light of these limitations, future research endeavors could address these gaps to enhance the robustness and applicability of sentiment analysis in predicting Bitcoin price trends. By broadening the scope of analysis to encompass multimedia content, multilingual sentiment, alternative information sources, varied time lags, and specific influencer effects, researchers can develop more nuanced models for understanding and forecasting cryptocurrency market dynamics.

# **REFERENCE**

1. Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. J. (2011, June). Sentiment analysis of twitter data. In Proceedings of the workshop on language in social media (LSM 2011) (pp. 30-38).
2. B, H. G., & B, S. N. (2023). Cryptocurrency Price Prediction using Twitter Sentiment Analysis. *Natural Language Processing, Information Retrieval and AI*. <https://doi.org/10.5121/csit.2023.130302>
3. Bandura, A. (1977). *Social Learning Theory*. Englewood Cliffs, N.J. : Prentice Hall.
4. Bikhchandani, S. (2021). *Information Cascades and Social Learning*.
5. Caferra, R. (2020). Good vibes only: The crypto-optimistic behavior. *Journal of Behavioral and Experimental Finance*, *28*, 100407. <https://doi.org/10.1016/j.jbef.2020.100407>
6. Critien, J. V., Gatt, A., & Ellul, J. (2022). Bitcoin price change and trend prediction through twitter sentiment and data volume. *Financial Innovation*, *8*(1). <https://doi.org/10.1186/s40854-022-00352-7>
7. Edgari, E., Thiojaya, J., & Qomariyah, N. N. (2022). The Impact of Twitter Sentiment Analysis on Bitcoin Price during COVID-19 with XGBoost. *2022 5th International Conference on Computing and Informatics (ICCI)*. <https://doi.org/10.1109/icci54321.2022.9756123>
8. Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, *25*(2), 383. <https://doi.org/10.2307/2325486>
9. Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, *25*(2), 383. <https://doi.org/10.2307/2325486>
10. Hasan, M. A. U., Bakar, A. A., & Yaakub, M. R. (2024). Measuring user influence in real-time on twitter using behavioural features. *Physica A: Statistical Mechanics and Its Applications*, *639*, 129662. <https://doi.org/10.1016/j.physa.2024.129662>
11. Hutto, C., & Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *Proceedings of the International AAAI Conference on Web and Social Media*, *8*(1), 216–225. <https://doi.org/10.1609/icwsm.v8i1.14550>
12. John, D. L., & Stantic, B. (2022). Forecasting Cryptocurrency Price Fluctuations with Granger Causality Analysis. *Recent Challenges in Intelligent Information and Database Systems*, 201–213. <https://doi.org/10.1007/978-981-19-8234-7_16>
13. Kahn, C. M. N. (1900). *Sentiment Market Analysis*.
14. Kraaijeveld, O., & De Smedt, J. (2020). The predictive power of public Twitter sentiment for forecasting cryptocurrency prices. *Journal of International Financial Markets, Institutions and Money*, *65*, 101188. <https://doi.org/10.1016/j.intfin.2020.101188>
15. Livieris, I. E., Kiriakidou, N., Stavroyiannis, S., & Pintelas, P. (2021). An Advanced CNN-LSTM Model for Cryptocurrency Forecasting. *Electronics*, *10*(3), 287. <https://doi.org/10.3390/electronics10030287>
16. McNally, S., Roche, J., & Caton, S. (2018). Predicting the Price of Bitcoin Using Machine Learning. *2018 26th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP)*. <https://doi.org/10.1109/pdp2018.2018.00060>
17. Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, *5*(4), 1093–1113. <https://doi.org/10.1016/j.asej.2014.04.011>
18. Nakamoto. (2008). Bitcoin: A peer-to-peer electronic cash system. Retrieved April 1, 2024, from <https://assets.pubpub.org/d8wct41f/31611263538139.pdf>
19. Paladines, J., & Ramirez, J. (2020). A Systematic Literature Review of Intelligent Tutoring Systems With Dialogue in Natural Language. *IEEE Access*, *8*, 164246–164267. <https://doi.org/10.1109/access.2020.3021383>
20. Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. *Foundations and Trends® in Information Retrieval*, *2*(1–2), 1–135. <https://doi.org/10.1561/1500000011>
21. Pano, T., & Kashef, R. (2020). A Complete VADER-Based Sentiment Analysis of Bitcoin (BTC) Tweets during the Era of COVID-19. *Big Data and Cognitive Computing*, *4*(4), 33. <https://doi.org/10.3390/bdcc4040033>
22. Rezaie, B., Zahedi, M., & Mashayekhi, H. (2020). Measuring time-sensitive user influence in Twitter. *Knowledge and Information Systems*, *62*(9), 3481–3508. <https://doi.org/10.1007/s10115-020-01459-y>
23. Riquelme, F., & González-Cantergiani, P. (2016). Measuring user influence on Twitter: A survey. *Information Processing & Management*, *52*(5), 949–975. <https://doi.org/10.1016/j.ipm.2016.04.003>
24. Sattarov, O., Jeon, H. S., Oh, R., & Lee, J. D. (2020). Forecasting Bitcoin Price Fluctuation by Twitter Sentiment Analysis. *2020 International Conference on Information Science and Communications Technologies (ICISCT)*. <https://doi.org/10.1109/icisct50599.2020.9351527>
25. Sekhar, P. C., Padmaja, M., Sarangi, B., & Aditya. (2022). Prediction of Cryptocurrency using LSTM and XGBoost. *2022 IEEE International Conference on Blockchain and Distributed Systems Security (ICBDS)*. <https://doi.org/10.1109/icbds53701.2022.9935871>
26. Shen, D., Urquhart, A., & Wang, P. (2019). Does twitter predict Bitcoin? *Economics Letters*, *174*, 118–122. <https://doi.org/10.1016/j.econlet.2018.11.007>
27. Stenqvist, & Lönnö. (2017). Predicting Bitcoin price fluctuation with Twitter sentiment analysis. Retrieved April 1, 2024, from <https://www.diva-portal.org/smash/get/diva2:1762598/FULLTEXT01.pdf>