Cryptocurrencies investment framework using sentiment analysis of Twitter influencers

Mohammed Ali Zare Chahooki[1], Kia Jahanbin[1], Tole Sutikno[2]
[1]Department of Computer Engineering, Yazd University, Yazd, Iran
[2]Department of Electrical Engineering, Universitas Ahmad Dahlan, Yogyakarta, Indonesia

Article Info

Article history:

Received Sep 27, 2022
Revised Dec 9, 2022
Accepted Jan 1, 2023

Keywords:

Cryptocurrencies
Sentiment analysis
Trend price prediction
Twitter influencers
User importance factor

ABSTRACT

In recent years, cryptocurrency technology has become an attractive area for investment due to its transparency, independence, and non-transactional nature. Many analysts and researchers talk daily on social media about the future of various cryptocurrencies. These ideas can significantly impact whether or not people are willing to invest. This paper provides a framework to help traders learn about the opinions of influential people and organizations in the field. Over the course of six months, the sentiment of more than 90 significant Twitter users was extracted for the proposed framework. In this study, we used the Vader open-source tool for sentiment analysis. This paper provides an excellent opportunity for investment through sentiment analysis of lesser-known or emerging cryptocurrencies. Also in this paper, we introduce the user importance factor to calculate the value of each tweet based on the number of retweets and comments. This factor shows the importance of their opinions instead of considering the number of followers of the authors. This factor causes a lower coefficient to be assigned to an author's opinion if it decreases in importance over time. The results show that in the short and long term, users' opinions are very effective in the market for cryptocurrencies and in predicting its price trend.

Corresponding Author:

Kia Jahanbin
Department of Computer Engineering, Yazd University
Yazd, Iran
Email: kia.jahanbin@stu.yazd.ac.ir

1. INTRODUCTION

Digital assets such as Bitcoin and Ethereum are now known as cryptocurrencies. Through peer-to-peer networks, cryptocurrencies are exchanged decentralized and without the use of intermediaries. The main feature of cryptocurrencies is that, unlike conventional currencies, they can be traded without operating fees or imposed laws and regulations. The mentioned feature has caused the global acceptance of cryptocurrencies. Bitcoin and Ethereum are perhaps the best-known cryptocurrencies available, and both have seen price increases in recent years. Figure 1 shows the significant growth trend of the cryptocurrency market cap since mid-2017.

In addition to price growth, Bitcoin and Ethereum blockchain technology have been used to build other cryptocurrencies and applications, such as smart contracts [1]. A smart contract is a code on a blockchain like Ethereum that executes commands such as freeing up the contract amount if the contract terms are met (for example, building a website). Previous studies [2] show that cryptocurrencies fluctuate sharply against fiat currencies. One of the reasons for these fluctuations is the publication of negative or positive news and the opinions of social network users who are significant users of cryptocurrency. For example, Figure 2 shows the Bitcoin price chart and Elon Musk's famous tweet that caused it to plummet.

Figure 1. The upward trend of the market value of Cryptocurrencies-Source site Coinmarketcap.com

Figure 2. Elon Musk tweet with bitcoin price chart (from COIN360.COM)

The red rectangle in Figure 2 shows the amount of bitcoin falling from when the tweet was published until a few days later. According to the studies [3], four methods have been used to predict cryptocurrencies or shares, i.e., i) using only stock price data, ii) using technical or fundamental analysis, iii) using data from social networks and/or news, and iv) a combination of methods. Twitter is one of the most famous and oldest social networks. Twitter is a microblog with about 35 million daily active users in the United States. In this research, using the Twitter API, more than 350 thousand tweets from November 1, 2021, to April 6, 2022, related to the opinions of 90 individuals and organizations regarding 40 cryptocurrencies have been collected.

We have collected influencer feedback on many cryptocurrencies to help traders make decisions for the first time. The study conducted in sentiment analysis to predict or detect the trend of cryptocurrencies on social networks can be classified in two ways [4], i.e., i) focusing on extracting specific cryptocurrency comments such as Bitcoin using hashtags such as #Bitcoin or #BTC, regardless of the importance of the commenting user; and ii) extracting the opinions of influencers only for one or more famous cryptocurrencies. One of the reasons for this research is to suggest lesser-known cryptocurrencies to traders for review and investment. These cryptocurrencies can be very profitable quickly, but they are complicated and time-consuming to identify without the right tools. This paper extracts the opinions of influential individuals and organizations in cryptocurrency from Twitter. Then, text preprocessing is performed, such as deleting duplicate data with the same tweet ID. Then, a separate table with its own name is formed for each coin in the tweet. Separating cryptocurrencies causes the sentiment analysis to be taken out of the traditional model and to be considered for tweets talking about several cryptocurrencies. After the preprocessing steps, the polarity of the tweets is calculated using a role-based tool called Vader [5]. Finally, for each cryptocurrency, the percentage of polarity (positive, negative, or neutral) is calculated by assigning a normal coefficient called user significance.

The user importance factor is another innovation of this research, dynamically assigned to each tweet in terms of the total number of comments and retweets. The reason for dynamic allocation is that if the user's popularity or influence decreases over time, the system recognizes that his comments will have less impact on the final output. The proposed model's output indicates each coin's polarity and percentage. This paper focuses on using social network data (Twitter). The main challenge addressed in this paper is the purposeful selection of opinions. Also, considering the impact of opinions and news on the market price of cryptocurrencies, the use of standard analytical techniques in traditional financial literacy cannot correctly determine the trend of this market [6], [7]. The significant contributions of this paper are: i) we are reducing the impact of fake news and redundant information by extracting information from cryptocurrency news experts in this field; ii) we presented a dynamic model to find and recognize new cryptocurrencies and investment opportunities. Our model covers many cryptocurrencies, as opposed to other articles that only cover a few specific cryptocurrencies; iii) Our model is real-time usable. This model has a high speed in detecting and assigning labels due to the online retrieval of information, separating tweets related to each cryptocurrency in a separate table, and using Vader; iv) our model considers the information aspect of sentiment analysis because opinions related to each cryptocurrency are separated and stored in a separate table; v) our model helps to understand the market's behavior and make the investors' correct decisions in the stock portfolio's selection and management; and vi) the model assigns a coefficient of importance to each tweet instead of giving it based on the number of followers of the cryptocurrency's author. This approach will keep the value of the tweet if a user with fewer followers publishes a valuable tweet. The author's weight is also measured based on his opinions' impact. In general, the framework of the proposed model for predicting the cryptocurrency market is as follows: i) we find expert users in cryptocurrency and extract their opinions continuously; ii) necessary preprocessing is done on the extracted comments. Each comment is then saved in a table associated with its cryptocurrency; iii) sentiment analysis is performed by the Vader. Then, each tweet's coefficient of importance is calculated, and the Vader score is added. This score is saved as the final score for the tweet; and iv) finally, after entering the query, the user will see sentiment analysis results related to 40 cryptocurrencies. In the second part, the methods of predicting cryptocurrencies are reviewed. In the third part, the impact of sentiment analysis methods on buying and selling cryptocurrencies is discussed more precisely. The fourth part of the proposed method is fully described, and the fifth part presents the results of experimental experiments. Finally, conclusions and suggestions for future work are presented.

2. SOCIAL NETWORKS AND SENTIMENT ANALYSIS

In this section, we review the papers predicting the price of cryptocurrencies. According to the algorithms used in predicting the price of cryptocurrencies, articles can be divided into two categories [8], i.e., i) use of basic learning models such as support vector machines, decision trees, and shallow neural networks; and ii) use of deep learning models, reinforcement learning, or hybrid models (combining several deep learning methods). Also, during the review of previous articles [4], [9], the authors have found another classification that divides the papers into the following three categories according to the type of data used in predicting cryptocurrencies, i.e., i) use price data; ii) use price data along with social media and news data; and iii) use price data with technical analysis data, fundamental data, or blockchain data. Other classifications can also be provided based on the forecast period (long-term or short-term) or type of output (classification, regression, or both) considered in survey articles.

2.1. Use of price data

In this group of articles, only the price features of the cryptocurrency are used to predict the price. Price characteristics are abbreviated as OHLCV. These letters indicate the open price, the highest price, the lowest price, the closing price, and the volume of cryptocurrency transactions in a daily trading session. In this paper, machine learning algorithms aim to predict the closing price of cryptocurrencies for one or more subsequent trading sessions. In [6], three models based on recursive neural networks—long short-term memory (LSTM), gated recurrent unit (GRU), and bidirectional LSTM (Bi-LSTM)—are proposed to predict the price of Bitcoin, Litecoin, and Ethereum. The presented models are evaluated using the mean absolute percentage error (MAPE). The results show GRU performed better at predicting all three cryptocurrencies than LSTM and bi-LSTM models. In [8], a hybrid model of GRU and LSTM for Monroe and Litecoin prediction is presented. The authors have shown that using LSTM is very effective because of its ability to memorize data features. For future work, the use of public sentiment is suggested. Despite the high accuracy of deep learning models, due to the strong dependence of cryptocurrencies on news and opinions, these models cannot identify the risks of this type of stock.

2.2. Price data along with social media and news data for price prediction

One of the data sources that has recently received much attention from researchers is commentary related to cryptocurrencies on social networks and Google Trends. As stated in [9], [10], cryptocurrencies receive significant influence from the opinions published on social networks, especially the opinions of influencers. On the other hand, in new research [4], the authors are more focused on determining the price trend of cryptocurrencies because the actual price prediction has a high error rate due to their volatile nature. Therefore, instead of accurately predicting the price of cryptocurrencies, the articles reviewed in this group predict their future price trends.

In [10], a trading platform called KryptoOracle is presented for real-time digital currency price forecasting based on Twitter sentiment analysis. The proposed platform is based on Spark and online learning, where the model's weight is adapted to new prices and sentiments. This paper uses the Vader model for sentiment analysis, and the XGBoost algorithm is used for machine learning operations. Despite the coverage of news data and comments on social networks related to cryptocurrencies, these methods have not considered the value of opinions and users. For example, just by searching for hashtags like #BTC or #ETH, all comments have been collected regardless of the user's expertise. Also, if a user is checking, authors have only paid attention to the number of followers, and the value of individual opinions has not been taken into account according to the number of user experts and other criteria. The mentioned challenge has not been investigated in an article so far.

2.3. Use of price data with technical analysis, fundamental, or blockchain data

The authors have used economic and trading knowledge in this group of articles. One of the standard techniques in trading is the use of technical analysis. In economics, technical analysis is a method for forecasting prices through past market information [11]. On the other hand, fundamental sentiment analysis is the financial analysis of businesses, competitors, and related markets. This analysis also considers the general state of the economy and factors such as interest rates, production, income, employment, GDP, and other macroeconomic factors [11].

In [12], first, the XGBoost and Random Forest methods are used to select the appropriate features to predict the future price of Ethereum. Then, a set of deep learning models consisting of LSTM, GRU, and TCN is presented to predict the price and price trend of Ethereum in the short and long term (one and seven days, respectively). In [13], a classification tree-based model for predicting bitcoin returns has been developed using 124 technical analysis indicators. The results show that the proposed model has the desired predictive power for new samples at limited daily intervals. These methods have two weaknesses. One, fundamental analysis of cryptocurrencies is difficult and requires proof. For example, the relationship between the increase or decrease in the rate of the dollar, oil, gold, etc., and cryptocurrencies needs to be proven. Also, due to the decentralized nature of these stocks and their lack of dependence on the government or a specific institution, it is practically impossible to examine their physical assets and financial statements. The second weakness of these methods is the lack of attention paid to news and experts' opinions on social networks.

2.4. Social networks and sentiment analysis in predicting the price trend of cryptocurrencies

As mentioned in the previous sections, behavioral sciences and scientific research provide evidence of a close relationship between social media and cryptocurrency price fluctuations. There are two main sentiment analysis methods: machine-based and lexicon-based methods. Machine-learning-based approaches use classification algorithms to classify texts. On the other hand, Lexicon-based methods use the dictionary of sentiment and its adaptation to texts to determine its polarity [14]. These methods assign scores to words in the text that ultimately describe the text as having a positive, negative, or neutral sentiment charge. Figure 3 shows the classification of sentiment analysis methods (see [15] for further reading).

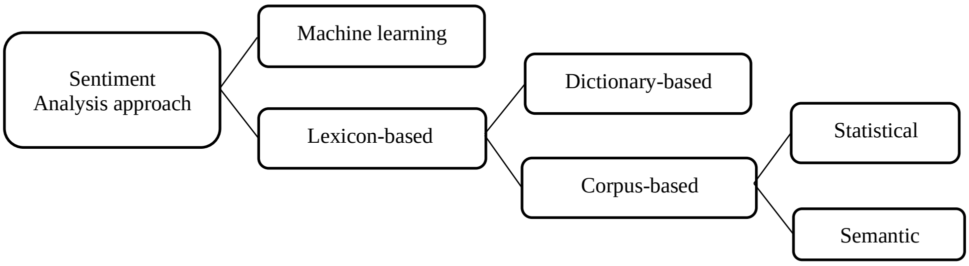


Figure 3. Classification of sentiment analysis methods [15]

In [16], the authors predicted the price of two cryptocurrencies, Bitcoin and Litecoin, in the short term for two hours. This paper predicts the price of these two cryptocurrencies based on the sentiment analysis of users' tweets and a multiple linear regression model. In future work, it is suggested to consider the credibility and popularity of the model to increase accuracy. One of the challenges on social media is the risk of a fake pump and dump of a cryptocurrency. In [17], it is stated that the purpose of "pump-and-dump" scams is to artificially increase the value of a cryptocurrency to encourage victims to invest. This study presents a computational approach to automatically identify pump-and-dump scams by combining social media information such as Twitter and Telegram. In [18] , we present a method for predicting the price of bitcoin using Twitter and historical prices. In this study, daily prices are combined with sentiment, and the next day's price is predicted using automatic regression models. In [19], social media data and basic machine learning algorithms like neural networks (NN), support vector machines (SVM), and random forests (RF) were used to predict the price trend of Bitcoin, Atrium, Ripple, and Litecoin.

3. METHOD

Social media literature has different definitions of influencers, but these people significantly impact their communities. The influence of famous people, also known as celebrities, has been proven in public health [20]. In this research, first, using Web Scraping, the IDs of influential individuals and organizations have been collected from the three websites, i.e., i) https://cryptoweekly.co/100, ii) https://lunarcrush.com/influencers and iii) https://coinculture.com/au/people/top-crypto-twitter-influencers. These users are considered seeds. After reviewing the seed, the authors selected the top 90 users with the richest comments. Selected users have fewer public comments. The general opinion refers to opinions that have little to do with cryptocurrencies and are often neutral. Figure 4 shows the effects of deleting these comments from the data.

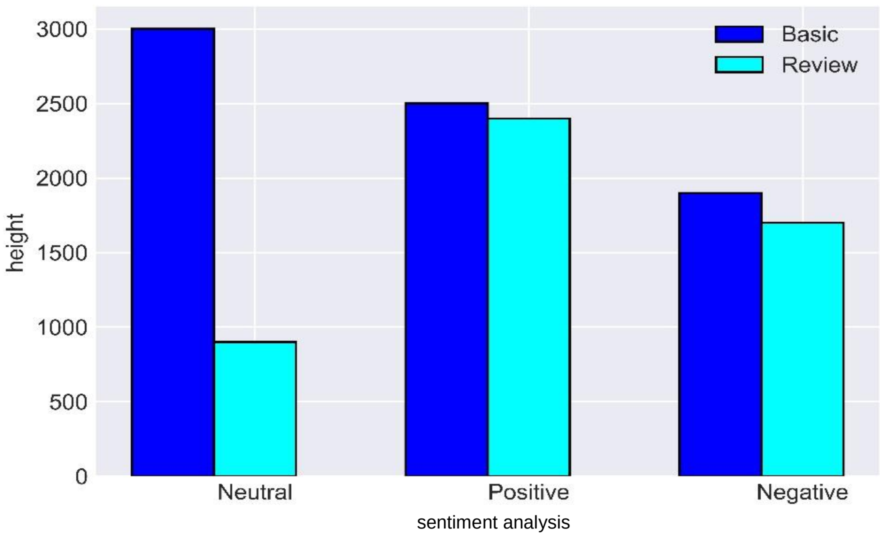


Figure 4. Comparison of initial primary data and post-clearance data

Figure 4 shows that if irrelevant comments are not removed, the model may make numerous errors in predicting the price trend. The collected data is then put through preprocessing. The text in this stage is first normalized by switching to lowercase letters. Using the preprocessing class of the Gensim package, the links in the duplicate tweets are then cleared, and duplicate tweets are then deleted in accordance with their tweet id. Added to a secondary list called Seed can be additional influencers. Retweeted by the original author, with their ids added, and then their ids removed from the tweet text. The text is then traced using the NLTK package's Snowball algorithm. Following the deletion of tweets with less than five characters, the tweets for each cryptocurrency are then stored in that cryptocurrency-specific table.

The Screener class of the Yahooquery package has been used to obtain the list of cryptocurrencies. Storing tweets related to each password in a separate table has the following advantages: i) ease of executing queries on the database in terms of runtime with increasing data volume; ii) suitability for online applications due to convenient runtime; and iii) ability to group comments according to the polarity and the author's name of each cryptocurrency. Then, the Vader tool was used for sentiment analysis of the extracted texts. Vader is a standard sentiment analysis tool in numerous articles [21], [22]. This tool is specially designed for texts published on social networks. VADER is created using a gold-standard sentiment lexicon based on valence-based and human curation.

As shown in [23], emoji are common in most texts published on Twitter. It can be said that emojis are standard literature in this type of writing. Most articles and tools like Vader [21], [22] do not cover this. Therefore, in the preprocessing phase, the emosent-py package converts emojis into points and detects them [24]. An example of these transformations is shown in Figure 5. The normalized score of the emojis is then added to the text sentiment score calculated by Vader. The last step in scoring and determining the polarity of comments is to consider the number of retweets and comments in each tweet. In this research, the normalized adaptive coefficient of importance is assigned to each tweet for the first time through (1).

```
from emosent import get_emoji_sentiment_rank
get_emoji_sentiment_rank('❤')

{'negative': 355.0,
 'neutral': 1334.0,
 'occurrences': 8050,
 'position': 0.746943086,
 'positive': 6361.0,
 'sentiment_score': 0.746,
 'unicode_block': 'Dingbats',
 'unicode_codepoint': '0x2764',
 'unicode_name': 'HEAVY BLACK HEART'}
```

Figure 5. Convert emoji to score

$$Positive_{Score} = \frac{\sum_{Row_i=1}^{n} \#Retwee(i) + \#Commen(i)}{MAX(\#Retweet_{Positive}.\#Comment_{Positive})}. \quad (1)$$
$$Row \in Positive\_Score\ belong\ Auther\ J$$

In (1), $\#Retwee$ and $\#Commen$ represent the number of comments on Tweet I, respectively. $MAX(\#Retweet_{Positive}.\#Comment_{Positive})$ indicates the maximum number of comments or retweets. The second part of the formula is to normalize the values. In (1) has the property that according to the popularity of each tweet, some value is added to the total score of positive comments of author J for the cryptocurrency I. In references [9] and [25], only the user's number of followers is considered, while another user's opinion with fewer followers may have a higher quality. Also, this relationship causes if the user decreases the quality of his comments over time, fewer points will be awarded to his comments. In (2), the total score equals the sum of positive observations and the coefficient of points calculated in (1). In (2), $Positive_{vader}$ represents the positive score by Vader about cryptocurrencies I, and Positive_Score represents the sum of the normalized scores. Negative and neutral comments are calculated from (1) and (2) by placing the corresponding numbers. The final user opinion of J for cryptocurrencies I is obtained through (3).

$$Total_{Score\_P} = Positive_{vader} + Positive_{Score} \quad (2)$$

$$Final_{Sentiment} = MAX(Total_{Score_P}, Total_{Score_{Ng}}, Total_{Score_{NU}}) \quad (3)$$

In (3), $Total_{Score_P}, Total_{Score_{Ng}}, Total_{Score_{NU}}$ they are the sum of the positive, negative, and neutral comment scores. Table 1 shows how to calculate (1) to (3).

Table 1. Pseudocode calculates comments score

Line	Sintax
1	*For Crypto I=1 to N*
2	$Positive_{Score} = \sum_{Row_i=1}^{n} \#Retwee(i) + \#Commen(i) / MAX(\#Retweet_{Positive}. \#Comment_{Positive})$
3	$Total_{Score\_P} = Positive_{Vader} + Positive_{Score}$
4	$Final_{Sentiment} = MAX(Total_{Score_P}. Total_{Score_{Ng}}. Total_{Score_{NU}})$
5	$return\ rank\ \left(Total_{Score_P}. Total_{Score_{Ng}}. Total_{Score_{NU}}\right) and\ Final_{Sentiment}$
6	*End For*

The framework of the proposed method is shown in Figure 6. This architecture consists of five blocks: information extraction, preprocessing, sentiment analysis, final score calculation, and the display block. In the information extraction block, the comments of influential users of cryptocurrency are extracted from Twitter. The information from this block is sent to the preprocessing block. This block performs operations such as lowering letters, deleting duplicate tweets, extracting other influential users, removing hyperlinks, and other preprocessing functions.

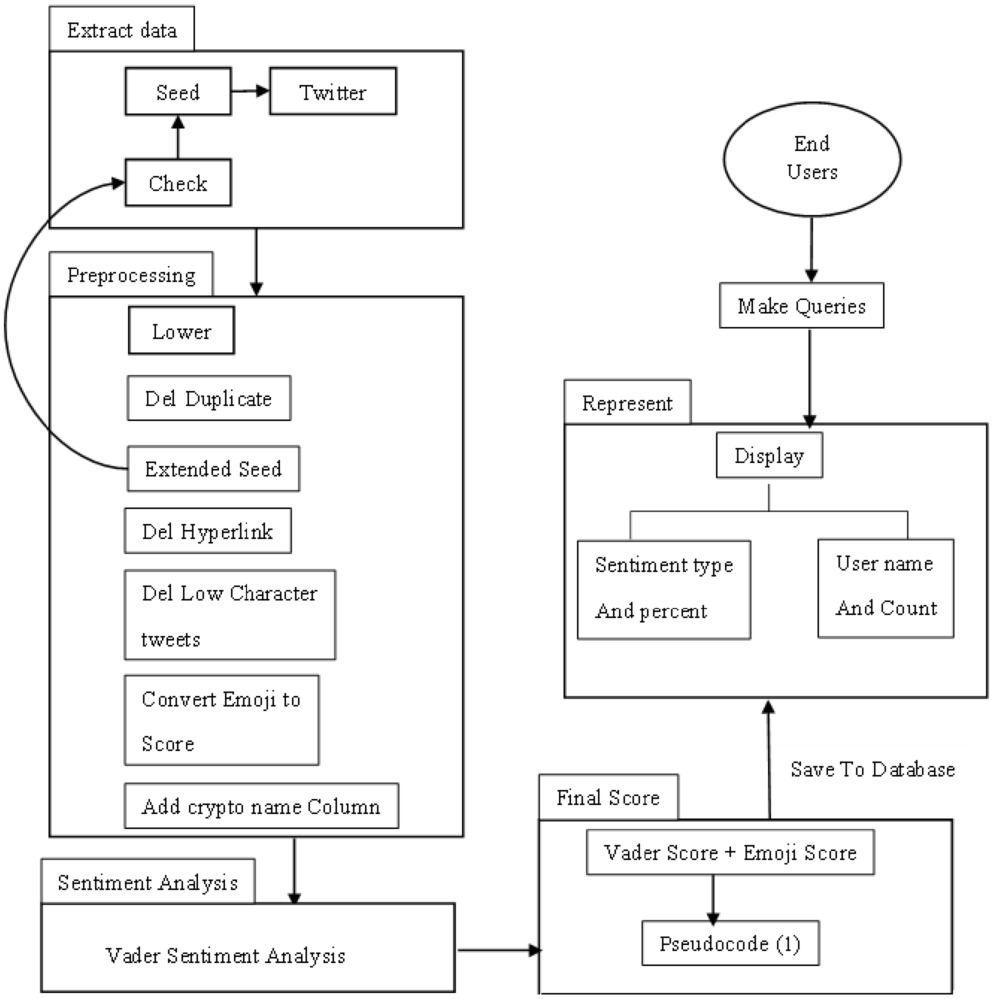


Figure 6. Proposed model architecture

In the final score block, the Vader score is first summed and normalized with the emojis, and then the final score is calculated with the help of Table 1. The output of the display block includes the type and percentage of sentiment, user name, and the number of user comments for 40 cryptocurrencies. Users are added to the Add to Seed section of influential users whose profiles and tweets are cryptocurrency and whose tweets are republished by the original Seed users.

4. RESULTS AND DISCUSSION

In this study, the Twitter API extracted information related to 90 influential personalities. Also, users who are direct advertisers or stakeholders in specific cryptocurrencies are ignored. Tweets are collected from November 1, 2021, to April 6, 2022, and include 350 thousand tweets. Figure 7 compares the number of tweets for Bitcoin and Ethereum with those for 19 other cryptocurrencies.

Unlike previous articles, the model's output is intended to assist traders in making the best decisions in seven modes [16], [26]. The three main polarities are positive, negative, and neutral, and the other four states are positive-negative, positive-neutral, negative-neutral, and equal. Four additional modes are created due to the 3% protective margin on the output. This margin helps to display both types of sentiment when, for example, positive and negative or positive and neutral opinions differ by less than 3%. Figure 7 shows that most tweets are dedicated to Bitcoin and Ethereum. Table 2 shows the percentage and polarity of the extracted sentiment related to 40 cryptocurrencies from July 1, 2022, to August 1, 2022.

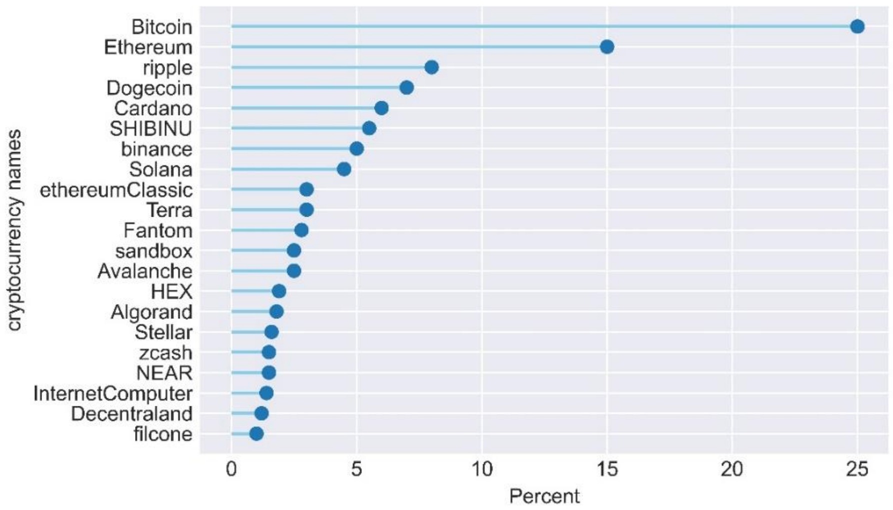
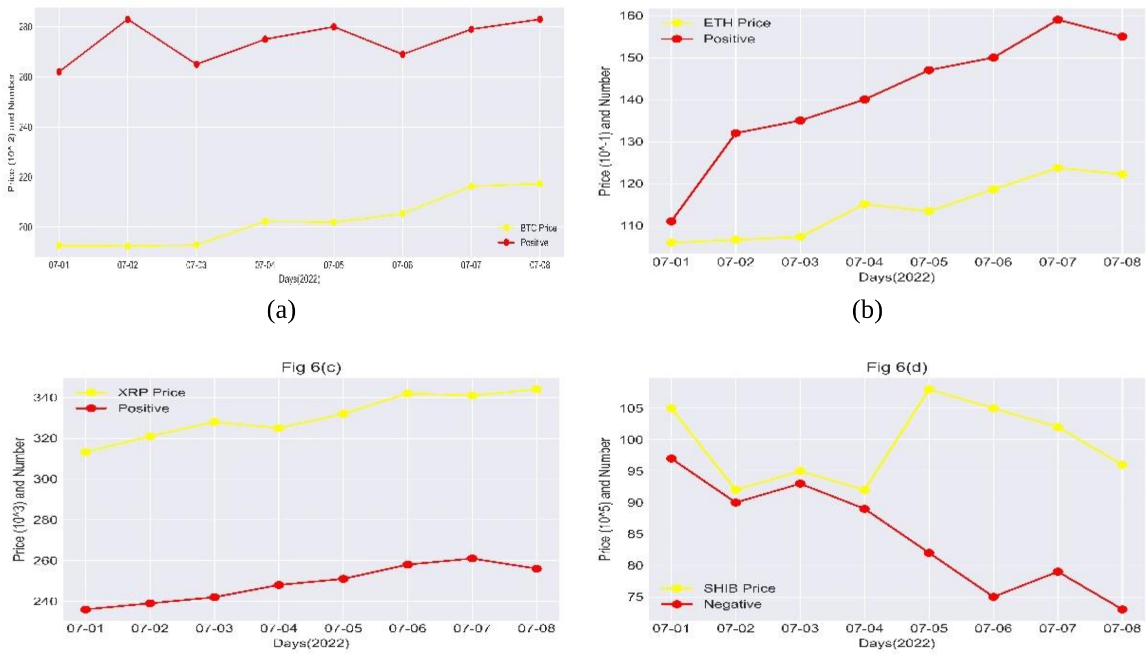


Figure 7. Compare the number of extracted Bitcoin and Ethereum tweets with other cryptocurrencies

Table 2. Predicting the price trend of cryptocurrencies by sentiment analysis

Row	Cryptocurrency	Percentage and Polarity	Row	Cryptocurrency	Percentage and Polarity
1	Bitcoin	Positive:57.6	21	Zcash	* Positive -Neural~ 35.8
2	Ethereum	Positive:61.9	22	Filcoin	Positive:47.4
3	Binance	Positive:52.5	23	Bitcoin Cash	Neutral:36.3
4	Shibinu	Positive:49.9	24	Polkadot	Positive:43.8
5	Ripple	Positive:40.7	25	Polygon	Positive:53.7
6	Cardano	Positive: 41.6	26	Dai	Neutral:47.9
7	Solana	Positive: 42.8	27	Lite Coin	Positive:48.7
8	Terra	Negative: 40.9	28	Cosmos	Positive:49.9
9	NEO	Neural:41.8	29	TRON	Neutral:39.2
10	ROSE	Positive:48.6	30	FTX Token	Positive:38.2
11	Avalanche	Positive:43.8	31	Unus Sed Leo	Negative:47.9
12	Dogecoin	*Negative-Neural~34.2	32	Monero	Positive:46.9
13	NEAR Protocol	Positive:43.7	33	BitTorrent	Positive:36.2
14	Algorand	*Negative-Positive~33.6	34	Aave	Positive:41.4
15	Decentraland	Positive:38.5	35	Waves	Neural:49.1
16	Stellar	Neural:39.6	36	Chainlink	Positive:39.7
17	Fantom	Positive:53.9	37	Uniswap	Positive:38.8
18	Ethereum Classic	Positive:49.3	38	VeChain	Positive:46.3
19	Internet computer	Positive:39.1	39	Helium	*Negative-Neural~33.5
21	Sandbox	Positive:38.6	40	Wrapped Bitcoin	Neural:37.3

Table 2 shows the price trend forecast of some cryptocurrencies with two types of sentiment. These results are due to the similarity of the number of views between the two types of polarity. For example, the count of negative and positive comments about Phantom cryptocurrencies has been very close. Due to the 3% margin in the model's output, the trader shows that the experts' positive and negative sentiments are equivalent. Figure 8 shows the four cryptocurrencies' price fluctuations: Bitcoin, Ethereum, Ripple, and Shiba, with positive or negative comments (displayed in the output). In Figure 8(a), the price of bitcoin has normalized in the sub-thousand-dollar range with two decimal places due to better understanding. Figure 8(a) shows a significant relationship between bitcoin prices' downward and upward trend and the day's sentiment. Figures 8 (b), (c), and (d), as well as Figure 8(a), depict the prices of ETH, XRP, and SHIB, respectively. Table 3 shows the impact of user comments on long-term investment and identifies investment opportunities in the four lesser-known cryptocurrencies between July 4, 2022, and August 6, 2022. Figure 9 depicts protocol cryptocurrency diagrams, with Figures 9(a) and (b) representing Cosmos and Near, respectively, extracted from Coin360.com to support the data in Table 3.



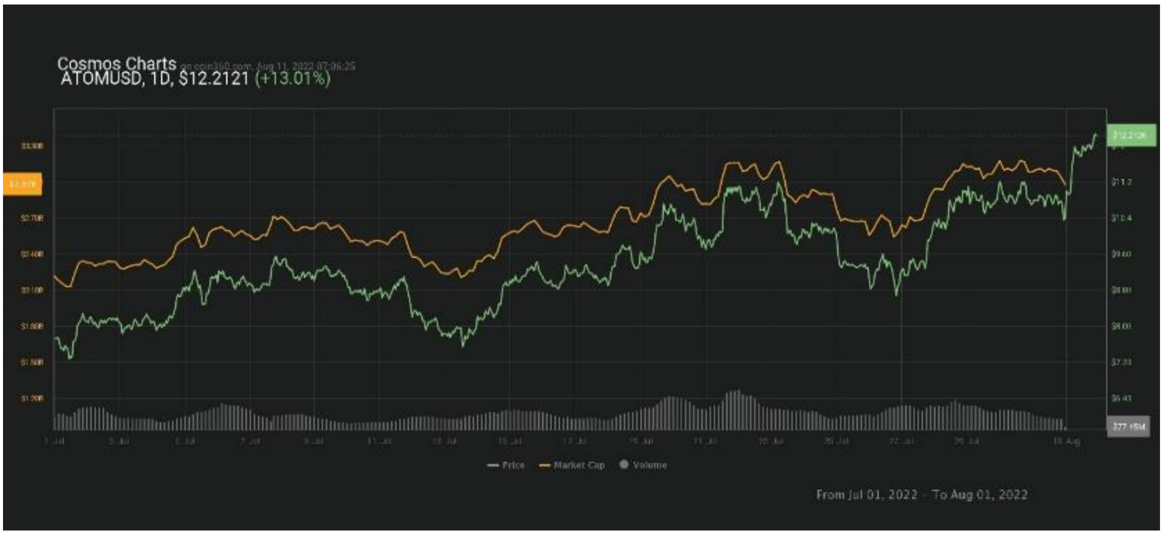
(a) (b) (c) (d)

Figure 8. Price fluctuations and sentiment analysis results of (a) Bitcoin, (b) Ethereum, (c) Ripple, and (d) Shiba

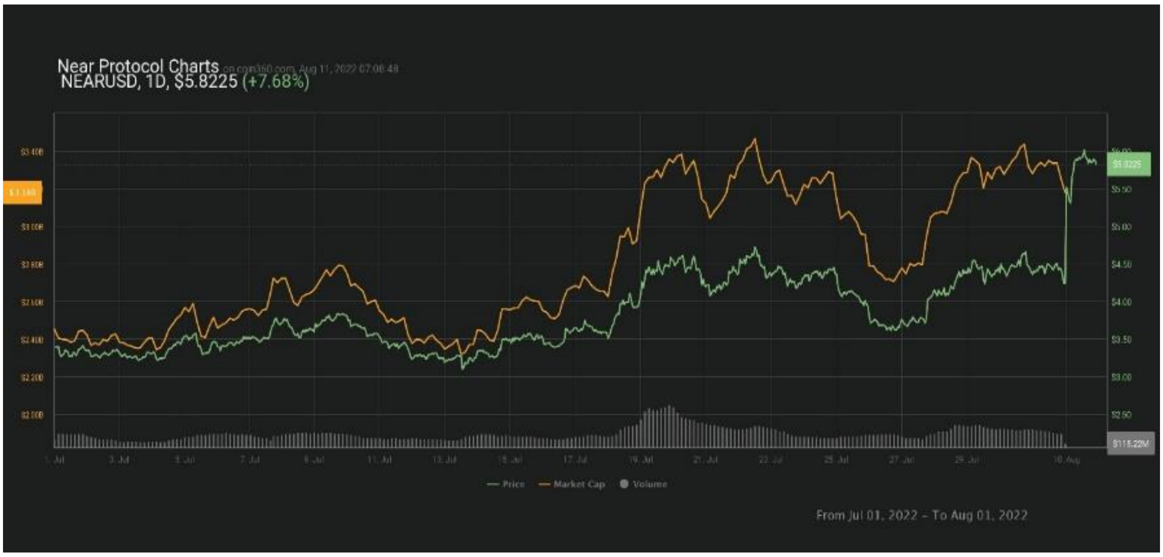
Table 3. Impact of user comments on identifying long-term investment opportunities

Row	Cryptocurrency	Positive	Negative	Neural	Profit / Loss
1	Uniswap	45.6%	29.9%	24.5%	+59.13%
2	Unus Sed Leo	25.4%	46.3%	28.3%	-1%
3	NEAR Protocol	46.5%	18.1%	35.4%	+47.1%
4	Cosmos	47.3%	16.9%	35.8%	+17%

As shown in Figure 9 and Table 3, the proposed framework has been able to accurately predict the price trend of cryptocurrencies by extracting the opinions of influential users. In Table 4, we compare our proposed model with the methods presented [16], [26] for predicting sentiment analysis's polarity (from July 1, 2022, to August 1, 2022). The results show that [16] or [26] neutral polarity is the dominant polarity or is very close to positive and negative polarity. Therefore, the authors have omitted this polarity, which could indicate a reluctance to trade in cryptocurrency. The large number of neutral comments mentioned in [16] and [26] is the lack of attention paid to cleaning texts unrelated to cryptocurrency transactions. We have solved this challenge effectively in the proposed model. In Table 4, the efficiency of the proposed model is compared with these articles.



(a)



(b)

Figure 9. Protocol cryptocurrency diagrams: (a) Cosmos and (b) Near (source: Coin360.com)

Table 4. Comparison of the proposed model with other methods

Model name	Cryptocurrency	Polarity Percent
Proposed model	BTC	Pos:57.6
		Neg:17.6
		Neu:24.8
	ETH	Pos:61.9
		Neg:15.3
		Neu:22.8
[26]	BTC	Pos:39.9
		Neg:23.1
		Neu:37.6
	ETH	Pos:41.2
		Neg:20.1
		Neu:38.7
[16]	BTC	Pos:40.6
		Neg:20
		Neu:35.4
	LTC	Pos:39.4
		Neg:26.7
		Neu:33.9

5. CONCLUSION

In this paper, we have extracted the tweets of more than 90 influential Twitter users in cryptocurrency. The polarity of the tweets is calculated using the sum of the Vader open-source tool points, the normalized score of the emojis, and the adaptive importance coefficient (pseudocode 1). The output of the proposed method includes the percentage and polarity of the tweets, the user's name, and the number of tweets for more than 40 cryptocurrencies. The adaptive importance coefficient gives more weight to tweets that get more readers' attention. Also, this coefficient is independent of the user's number of followers since a user's tweet with a smaller number of followers may have received more attention from readers. In addition, if the quality of a user's comments declines over time, this coefficient can give them less weight. The normalized score of the emojis, due to their widespread use in the Twitter literature, causes the polarity of the tweet to be calculated more accurately. The results show that the proposed method can, with acceptable accuracy, predict cryptocurrency price trends by analyzing the sentiment of expert users. This method can also offer attractive investment opportunities to traders in the short and long term. Another advantage of the proposed method is the study of lesser-known cryptocurrencies. These cryptocurrencies can generate a good profit for traders. The proposed model can detect cryptocurrency polarity with greater accuracy and percentage. The higher accuracy of the model of this paper is due to the purposeful extraction of opinions, proper cleaning, and the use of a normalized adaptive coefficient of importance. These three factors have caused us to minimize the challenge of neutralizing most of the opinions reported in other articles in this research. For future work, the authors will examine the impact of other social media, such as Reddit and Telegram, and political news in predicting the price trend of cryptocurrencies.

REFERENCES

[1] U. Mukhopadhyay, A. Skjellum, O. Hambolu, J. Oakley, L. Yu, and R. Brooks, "A brief survey of cryptocurrency systems," in *2016 14th annual conference on privacy, security and trust (PST)*, 2016, pp. 745-752: IEEE. doi: 10.1109/PST.2016.7906988.

[2] K. Delmolino, M. Arnett, A. Kosba, A. Miller, and E. Shi, "Step by step towards creating a safe smart contract: Lessons and insights from a cryptocurrency lab," in *International conference on financial cryptography and data security*, 2016, pp. 79-94: Springer. doi: 10.1007/978-3-662-5335746.

[3] A. M. Khedr, I. Arif, M. El-Bannany, S. M. Alhashmi, M. J. I. S. i. A. Sreedharan, Finance, and Management, "Cryptocurrency price prediction using traditional statistical and machine-learning techniques: A survey," vol. 28, no. 1, pp. 3-34, 2021. doi: 10.1002/isaf.1488.

[4] M. Birjali, M. Kasri, and A. J. K.-B. S. Beni-Hssane, "A comprehensive survey on sentiment analysis: Approaches, challenges and trends," vol. 226, p. 107134, 2021. doi: 10.1016/j.knosys.2021.107134.

[5] C. Hutto and E. Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text," in *Proceedings of the international AAAI conference on web and social media*, 2014, vol. 8, no. 1, pp. 216-225. doi: 10.1609/icwsm.v8i1.14550.

[6] M. J. Hamayel and A. Y. J. A. Owda, "A Novel Cryptocurrency Price Prediction Model Using GRU, LSTM and bi-LSTM Machine Learning Algorithms," vol. 2, no. 4, pp. 477-496, 2021. doi: 10.3390/ai2040030.

[7] S. Mohan, S. Mullapudi, S. Sammeta, P. Vijayvergia, and D. C. Anastasiu, "Stock price prediction using news sentiment analysis," in *2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService)*, 2019, pp. 205-208: IEEE. doi: 10.1109/BigDataService.2019.00035.

[8] S. Biswas, M. Pawar, S. Badole, N. Galande, and S. Rathod, "Cryptocurrency price prediction using neural networks and deep learning," in *2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS)*, 2021, vol. 1, pp. 408-413: IEEE. doi: 10.1109/Confluence52989.2022.9734139.

[9] S. Wooley, A. Edmonds, A. Bagavathi, and S. Krishnan, "Extracting cryptocurrency price movements from the reddit network sentiment," in *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, 2019, pp. 500-505: IEEE. doi: 10.1109/ICMLA.2019.00093.

[10] S. Mohapatra, N. Ahmed, and P. Alencar, "KryptoOracle: a real-time cryptocurrency price prediction platform using twitter sentiments," in *2019 IEEE international conference on big data (Big Data)*, 2019, pp. 5544-5551: IEEE. doi: 10.1109/BigData47090.2019.9006554.

[11] I. K. Nti, A. F. Adekoya, and B. A. J. A. I. R. Weyori, "A systematic review of fundamental and technical analysis of stock market predictions," vol. 53, no. 4, pp. 3007-3057, 2020. doi: 10.1007/s10462-019-09754-z.

[12] A. Politis, K. Doka, and N. Koziris, "Ether price prediction using advanced deep learning models," in *2021 IEEE International Conference on Blockchain and Cryptocurrency (ICBC)*, 2021, pp. 1-3: IEEE. doi: 10.1109/ICBC51069.2021.9461061.

[13] J.-Z. Huang, W. Huang, J. J. T. J. o. F. Ni, and D. Science, "Predicting bitcoin returns using high-dimensional technical indicators," vol. 5, no. 3, pp. 140-155, 2019. doi: 10.1016/j.jfds.2018.10.001.

[14] P. Chakriswaran, D. R. Vincent, K. Srinivasan, V. Sharma, C.-Y. Chang, and D. G. J. A. S. Reina, "Emotion AI-driven sentiment analysis: A survey, future research directions, and open issues," vol. 9, no. 24, p. 5462, 2019. doi: 10.3390/app9245462.

[15] N. Gupta and R. Agrawal, "Application and techniques of opinion mining," in *Hybrid Computational Intelligence*: Elsevier, 2020, pp. 1-23. doi: 10.1016/B978-0-12-818699-2.00001-9.

[16] A. Jain, S. Tripathi, H. D. Dwivedi, and P. Saxena, "Forecasting price of cryptocurrencies using tweets sentiment analysis," in *2018 eleventh international conference on contemporary computing (IC3)*, 2018, pp. 1-7: IEEE. doi: 10.1109/IC3.2018.8530659.

[17] M. Mirtaheri, S. Abu-El-Haija, F. Morstatter, G. Ver Steeg, and A. J. I. T. o. C. S. S. Galstyan, "Identifying and analyzing cryptocurrency manipulations in social media," vol. 8, no. 3, pp. 607-617, 2021. doi: 10.1109/TCSS.2021.3059286.

[18] J. Davchev, K. Mishev, I. Vodenska, L. Chitkushev, and D. Trajanov, "Bitcoin price prediction using transfer learning on financial micro-blogs," in *The 16th Annual International Conference on Computer Science and Education in Computer Science*, 2020.

[19] F. Valencia, A. Gómez-Espinosa, and B. J. E. Valdés-Aguirre, "Price movement prediction of cryptocurrencies using sentiment analysis and machine learning," vol. 21, no. 6, p. 589, 2019. doi: 10.3390/e21060589.

[20] K. Jahanbin, M. Jokar, and V. J. A. P. J. o. T. M. Rahmanian, "Using twitter and web news mining to predict the monkeypox outbreak," vol. 15, no. 5, p. 236, 2022.
[21] A. Ibrahim, "Forecasting the early market movement in bitcoin using twitter's sentiment analysis: An ensemble-based prediction model," in *2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS)*, 2021, pp. 1-5: IEEE. doi: 10.1109/IEMTRONICS52119.2021.9422647.
[22] O. Kraaijeveld, J. J. J. o. I. F. M. De Smedt, Institutions, and Money, "The predictive power of public Twitter sentiment for forecasting cryptocurrency prices," vol. 65, p. 101188, 2020. doi: 10.1016/j.intfin.2020.101188.
[23] Y. Chen, J. Yuan, Q. You, and J. Luo, "Twitter sentiment analysis via bi-sense emoji embedding and attention-based LSTM," in *Proceedings of the 26th ACM international conference on Multimedia*, 2018, pp. 117-125.
[24] P. Kralj Novak, J. Smailović, B. Sluban, and I. J. P. o. Mozetič, "Sentiment of emojis," vol. 10, no. 12, p. e0144296, 2015. doi: 10.1371/journal.pone.0144296.
[25] A. A. Chandawarkar, D. J. Gould, and W. J. A. s. j. Grant Stevens, "The top 100 social media influencers in plastic surgery on Twitter: who should you be following?," vol. 38, no. 8, pp. 913-917, 2018. doi: 10.1093/asj/sjy024.
[26] J. Abraham, D. Higdon, J. Nelson, and J. J. S. D. S. R. Ibarra, "Cryptocurrency price prediction using tweet volumes and sentiment analysis," *SMU Data Science Review*, vol. 1, no. 3, p. 1, 2018.

BIOGRAPHIES OF AUTHORS

Mohammed Ali Zare Chahooki is an Associate Professor in the Computer Engineering Department of Software Engineering at Yazd University, Iran; he received a B.Sc. degree in computer science from Shahid Beheshti University, an M.Sc., and a Ph.D. degree in software engineering from the Tarbiat Modares University, Iran. His research interests include machine vision, software engineering, machine learning, and social networks analysis. You may contact him by sending an email chahooki@yazd.ac.ir.

Kia Jahanbin received a B.Sc. degree in software engineering from the Shahid Bahonar Shiraz University, Iran, an M.Sc. degree in Informa Technology from the Azad University of Kerman, Iran, and now his Ph.D. candidate at Yazd University in soft engineering, Iran. His research interests include text mining, data mining, web mining, machine learning, and deep learning. You may contact him by sending an email to kia.jahanbin@stu.yazd.ac.ir.

Tole Sutikno is currently employed as a lecturer in the Electrical Engineering Department at Universitas Ahmad Dahlan (UAD), which is located in Yogyakarta, Indonesia. In 1999, 2004, and 2016, he graduated with a Bachelor of Engineering from Universitas Diponegoro, a Master of Engineering from Universitas Gadjah Mada, and a Doctor of Philosophy in Electrical Engineering from Universiti Teknologi Malaysia. All three degrees are in the field of electrical engineering. Since the year 2008, he has held the position of Associate Professor at the Universitas Ahmad Dahlan in Yogyakarta, Indonesia. He is among the top 2% of researchers named by Stanford University and Elsevier BV as the most influential scientists in the world for 2021–present. His research interests include the areas of digital design, industrial applications, industrial electronics, industrial informatics, power electronics, motor drives, renewable energy, FPGA applications, embedded systems, artificial intelligence, intelligent control, digital libraries, and information technology. You may contact him by sending an email to tole@te.uad.ac.id.