Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
Volume 07 Issue 01
2024

Research Article
https://doi.org/10.56979/701/2024

Crypto Currency Price Prediction through Tweets Using NLP

Arsalan Khan[1], and Imran Ihsan[2*]

[1]FAST School of Management, NUCES, Islamabad, 44000, Pakistan.
[2]Department of Creative Technologies, Air University, Islamabad, 44000, Pakistan.
[*]Corresponding Author: Imran Ihsan. Email: iimranihsan@gmail.com

Received: January 12, 2024 **Accepted:** March 22, 2024 **Published:** June 01, 2024

Abstract: This research paper presents a comprehensive approach to sentiment analysis and stock price prediction using tweets from cryptocurrency influencers. The methodology involves several stages: environment setup, data loading, preprocessing, sentiment analysis using multiple machine learning models, and stock price prediction. The sentiment analysis leverages pre-trained models such as Roberta, VADER, XLNet, ALBERT, BERT, and BERTweet, each fine-tuned for aspect-based sentiment analysis. A majority voting mechanism determines the final sentiment, mitigating biases and enhancing reliability. Historical Bitcoin prices are fetched to analyze the relationship between tweet sentiments and price movements. The stock price prediction framework utilizes deep learning architectures, including LSTM, GRU, CNN-LSTM, Transformer, Bidirectional LSTM, and Simple CNN models. These models are trained on historical stock data, with features scaled and split into training, validation, and test sets. The models are evaluated using metrics such as RMSE, MSE, MAE, and R-squared, and their performance is visualized through plots of actual versus predicted prices and trends. The integration of sentiment analysis with stock price prediction provides a robust tool for understanding the impact of social media on financial markets. Using multiple models and majority voting ensures a comprehensive and reliable sentiment analysis, while the diverse deep-learning models offer insights into stock price trends and predictions. This research contributes to financial analytics by demonstrating the potential of combining NLP and deep learning techniques for market prediction.

Keywords: Crypto Currency; Tweets; Natural Language Processing.

1. Introduction

Cryptocurrencies are also known as digital money or virtual money an online currency that incorporates encryption techniques as a security means. Unlike conventional money, which is legally tender and issued by a government or central bank, it is decentralized and operated on blockchain technology. It is divided into two categories, the first one is Bitcoin which was started in 2009, and the second generations which include Litecoins and Ethereum. They are digital currencies. Spearhead the creation of P2P Payments; transactions that people can make directly without the involvement of a central controller such as a middleman bank. Cryptocurrencies are clear and safe because exchanges of the digital currency are registered on a public index. At the same time, the identities of the people making the transactions are also pseudonymous, so it affords some, the ability of anonymity.

Since the creation of bitcoins, cryptocurrencies have developed into different forms and types. Bitcoin was the first cryptocurrency introduced in 2009. The digital currencies can be categorized in the following broad categories: Bitcoin, Altcoins, Tokens, Stablecoins, Meme Coins, CBDCs and NFTs. Every kind of cryptocurrency has a particular use and is useful for people in various ways. Markets. Therefore, the criterion of the cryptocurrency to be invested in or used in any other endeavor depends on his or her objectives and attitudes to risk.

Twitter is one of the radical inventions that have impacted several aspects of life, including the economic sector. That is why, for example, a social network in the form of a microblogging service, Twitter with millions of users from different countries of the world, has become an important factor in forming the opinion of the population and, accordingly, the demand in the market. This influence is especially felt in the context of cryptocurrency trading. Some prominent and influential personalities in the business world today turn to the use of the popular site known as Twitter to convey their stance on given altcoins. Such tweets can influence the price and volume of the referred cryptocurrencies among the users of the app and the platform. For instance, remarks by Elon Musk, the chief executive officer of Tesla and SpaceX have been observed to have a significant floating point in the market. This is evident through an example of Musk tweeting about Dogecoin, a meme-based cryptocurrency, thus boosting its value by more than 800

Nevertheless, it's not just the posts that the market influencers make that cause shifts in the market. Another factor is the sentiment regarding cryptocurrency on Twitter in general. From the Twitter conversations that have been analyzed, it is evident that the sentiment toward cryptocurrencies is an inherent characteristic of the valuation of the said cryptocurrencies at any one time. It implies that the premise of a positive tweet is a positive price change or that a negative tweet has indicators of a negative price change. Twitter is one of the available channels through which the key opinion leaders manage to influence the markets, as well as through which the trends may be set, and ordinary users share their opinions on the given cryptocurrency or some other digital asset. Thus, the use of such hashtags as buy, sell, and other purchase/sale options can provoke enormous price changes and trading volumes in the sphere of cryptocurrencies.

There are various approaches to study such as cryptocurrency price prediction. The methodologies incorporate Natural Language Processing, (NLP), Machine Learning (ML), and Deep Learning (DL).

- **Natural Language Processing (NLP)** techniques are important for the analysis of textual information from Twitter, news, and others. Websites analyzed for market trends are one of the essential components of predicting cryptocurrency prices [6-7].
- **Machine Learning (ML)** techniques involve predictive modeling and statistical analysis to forecast market trends based on historical data using Regression Models [17-18], Ensemble Methods [18] and Hidden Markov Models [2, 5].
- **Deep Learning (DL)** techniques, particularly neural networks, are extensively used for their ability to process large sets of data and model complex non-linear relationships. These include Recurrent Neural Networks [7], Convolutional Neural Networks [39] and Transfer Models [22].
- **Hybrid Models** that combine multiple techniques have also been explored for improved accuracy by combining all above [6-7 [39].

The following are the objectives of this research:

1. Develop Robust NLP Models
2. Integrate Multiple Sentiment Analysis Techniques
3. Aspect-Based Sentiment Analysis
4. Correlation Analysis
5. Majority Voting Mechanism for Sentiment Determination
6. Historical Data Integration
7. Model Training and Evaluation

2. Related Work

2.1. Machine Learning Techniques

The studies on sentiment analysis have utilized classic ML techniques as presented in Table 2.1. Thus, machine learning remains significant even with the development of deep learning. Researchers have enhanced efficiency over time using machine learning in combination with deep learning methods or addressing the limitations of machine learning models [2].

2.2. Deep Learning Techniques

The current models of sentiment analysis used in cryptocurrency research are specifically preceded by deep learning models, based on the results of a survey shown in Table 2. 1. LSTM has been used mostly thanks to its appreciation for working with sequential data by most researchers. A novel approach employed entails the use of an 'Attention-based Hierarchical LSTM' to improve the process of sentiment analysis [45]. CNN, LSTM, and other approaches are compared or used one by one, and some works show that CNN appears to perform better than other models in some cases [17]. There is also literature that concerns the investigation into the effectiveness of various LSTM activation functions, indicating the possibility of using fewer common functions [1]. Furthermore, studies are carried out with mixed models using machine learning functions together with deep learning techniques namely CNN-LSTM for the context of sentiment analysis [3].

Table 1. Survey of Related Work

Paper	Year	Dataset	Algorithm(s)	Performance
[1]	2021	Energy Companies, CRIX	BSADF, Panel Models	Energy Risk Increase
[2]	2021	Google, Twitter, Market Data	HMM	Sentiment Drives Bulls
[3]	2021	Cryptocurrency Prices	LPPLM, Wavelet	Strong Market Correlation
[4]	2021	DeFi, NFTs, Bitcoin, Ethereum	Data Sampling PS	DeFi, NFT Bigger Bubbles
[6]	2023	Yahoo Bitcoin, Kaggle Tweets	FinBERT, GRU	Sentiment: 0.94 Error, Price: 0.036 Error
[7]	2023	Ethereum, Solana Market & Tweets	FinBERT, LSTM-GRU	Various MAEvalues Ethereum and Solana
[10]	2022	BTC & ETH Tweets, Price Data	SVM, NB, LSTM	Performance metrics for BTC and ETH
[34]	2023	CEV Model, Type 3 Bubbles	CEV Model	Bubble Detection
[36]	2023	Global Stock Indices	Phillips-Shi Bubble Test	
[37]	2023	Cryptocurrency Bubbles	SADF	103 pre-COVID 599 post-COVID
[38]	2023	NFT Data	LPPLS Models	NFT Forecasting
[39]	2022	Cryptocurrency Tweets	Ensemble LSTM-GRU	Sentiment: 0.99 Acc, Emotion: 0.92 Acc
[11]	2022	Crypto Tweets, Financial Data	MBHN, HGRU, TAM, GRU-DNMS	
[41]	2022	Social media, Crypto Community Data	TextBlob, GoogleNLP, BERT-models	Correlations: 0.33-0.57 Fin-BERT: 0.32
[13]	2022	NFTI, DPI, NFT & DeFi Assets	SADF, GSADF, LPPLS	NFT vs. DeFi Bubbles
[14]	2022	Bitcoin, Ethereum Prices	LPPLS, GSADF	Bitcoin (5), Ethereum (4)
[15]	2021	BTC, ETH Prices; NEO Tweets	Random Forest, BERT	Accuracy: 0.77
[16]	2021	Cryptocurrencies & COVID-19	Sentiment analysis	Lowered BTC, BNB, ADA
[17]	2021	Twitter & Reddit posts	SVM, LR, RNTN, GB	Various metrics

3. Methodology

3.1. Dataset

The dataset also has an important role in this research because it contains the data that is needed for analysis and further use in the modeling process. The totality of the following sections aims to describe the origin and organization of the dataset used in the current investigation as well as the unique characteristics of the data.

3.1.1. Source Collection Method

Data collection was intensive and entailed the use of the Twitter API to zoom in on the market analysis of the 40 cryptocurrencies. The systematic data acquisition process ensures that a focused as well as diverse data sample is developed which is ideal for analysis [40].

3.1.2. Content and Structure

- **Main Table (Tweets1):** The main data set is found in the "Tweets1," table which totals 11 columns. It gathers information known to each tweet; therefore it provides extensive information on the discussions about cryptocurrencies on the site Twitter [40].
- **Cryptocurrency-Specific Tables:** The dataset has topicality to base initial authentic individual tables on thirty money units besides, in addition to the main table, there are 40 tables for each cryptocurrency. These specialized tables are important for undertaking the analyses of the discussions about specific cryptocurrencies in detail [40].

3.1.3. Unique Features

- **Expertise Focus:** In contrast to most of the other databases, this one focuses on the popular figures on Twitter for their acknowledged knowledge of the cryptocurrency market. This focus enhances the suitability and quality of the data [40].
- **Reduced Irrelevance and Neutral Polarity:** Due to that, some irrelevant content and, or tweets with mostly neutral sentiments which are commonly observed while working with large tweet collections [40], were excluded from the dataset.

3.1.4. Usage and Applications

The given dataset can be applied to such machine learning programs as sentiment analysis, deep transfer learning, and forecasting the rates of cryptocurrencies. It is especially helpful for identifying trends of opinion influence on cryptocurrency markets and applying sentiment analysis to the modulization of financial markets [40].

3.1.5. Access and Citation

Experts can use the data set, though they need to acknowledge the source before using it in any research. Using Twitter data offers the identification of tendencies in the cryptocurrency market and is a valuable instrument for examining NLP's connection to crypto trading [40].

3.1.6. Financial PhraseBank Dataset

- **Description and Source:** The dataset of Financial PhraseBank contains sentences categorized according to the sentiments and intensity of the positive and negative sentiments expressed in financial news articles. It is conspicuous with the objective of training. as well as for cross-validation and testing of the models for sentiment analysis being specific to the financial domain[42]. Both this dataset and others are important and necessary for sentiment analysis models' training (XLNet, ALBERT, Interpreting the results of this study regarding the applied models (BERTweet, RoBERTa, and BERT).
- **Structure:** The other external dataset used is Financial PhraseBank which has a news sentence and its sentiment tag positive, negative, or neutral which assists in the training of models in handling finance-specific language.

3.1.6. Binance API Data

- **Description and Usage:** To perform a detailed analysis, historical data with regards to cryptocurrencies was collected via the Binance API where enough details regarding price changes for various cryptocurrencies in the past were provided. It is used to create a relation between the text, in this case, tweets, and the actual market value giving insight into the effects of social media in determining the price of cryptocurrencies.
- **Data Collection Method:** To collect the historical price data of the mentioned cryptocurrencies in the tweets conducted API requests to Binance. It is possible to specify the granularity to access data for getting the information by the desired time interval; however, for this research, data on the minutes was used to correspond with the tweet's time provided.
- **Integration with Tweet Data:** Twitter data and cryptos' historical price data from the Binance API were merged based on the timestamps of the tweets. Thus, it yielded a comprehensive data set that included both sentiment data and price data; the two are used for evaluating the correlation between sentiment and price fluctuations at different points in time.

3.2. Data Acquisition and Preprocessing

In this research, we have been keenly collecting and preparing Twitter data throughout this study for comprehensive sentiment analysis. The first steps involved in this phase of analysis were to set up a working environment for analysis of the text data, tools for handling and processing the text data, NLP, and machine learning. The main dataset was derived from a CSV file containing tweets of cryptocurrency influencers between the years 2021 and 2023. The preprocessing procedures conducted are outlined below:

- **Data Cleaning:** Tweets lacking 'full_text' were removed to ensure dataset completeness.
- **Text Standardization:** Text entries were standardized by filling in missing values and converting all text to a uniform case.
- **Text Refinement:** The 'full_text' field was processed to remove punctuation, tokenize text, eliminate stopwords, and apply lemmatization, thereby enhancing data quality.
- **Temporal Organization:** Tweets were organized chronologically by converting the 'created_at' field to date time format, facilitating time-series analysis.

3.3. Influential Data Identification and Stock Data Integration

Having cleaned the data, the focus was on the classification of tweets that generated noticeable resonance and, thus, could be considered impactful. We also determined the importance of the coefficient of the tweet and normalized it to find out how influential the tweet was. At the same time, we also got historical cryptocurrency data from Binance API and used the timestamp of the tweets to analyze the link between tweet sentiments and changes in the stock market.

3.4. Sentiment Analysis Implementation

To analyze the emotions conveyed in the tweets, I utilized a multi-model approach for sentiment analysis.

- **Hybrid Model Application:** We used spaCy for analyzing the language and Roberta for extracting sentiment from tweets using deep learning techniques.
- **VADER Utilization:** VADER was employed to evaluate the emotional sentiment of tweets due to its effectiveness in analyzing brief texts.
- **Advanced Model Integration:** Both BERTweet and XLNet were fine-tuned to capture the subtleties of language used on Twitter, allowing for a more thorough analysis of the emotions and situations expressed in tweets.

3.5. Influential Data Identification and Stock Data Integration

In this paper, composing extensive market data, we have assessed the endeavor of Twitter sentiments in affecting the price of cryptocurrencies. Cryptocurrency's historical prices were obtained from Yahoo Finance. To filter out useful information, major changes in the prices and trading volumes of these securities were obtained from this data. Furthermore, the current date formats were also transformed to improve consistency during analysis, while the dates were regarded as perfect to match the time stamps of the tweets.

3.6. Predictive Modeling and Evaluation

Inspired by the capability of some enriching tweets to sway the prices of cryptocurrencies, I have established highly accurate predicting models that employ sophisticated natural language processing. These models include Transformer, LSTM, GRU, and Electra and the training of the models was geared towards explaining compound sentiments captured within important tweets with relative price changes of cryptocurrencies.

One of the important steps of my assessment was a paper trading simulation. All the decisions made for trading were done with the models using historical data, and thus this simulation was a real-world simulation of the trading activities without actual money being used. This helped me in determining how the models would have been when used under real market situations considered from historical data which gave me a good indication of how practical the models can be.

3.7. Scaling Predictions and Paper Trading

To ensure more accurate and scalable predictions, attention was paid to the selection of more important 'tweets' – those tweets with a high 'importance coefficient' that define the shifts in public opinion and the stock market. Such an approach pointed to more effective and relevant predictions than the previous strategy. For paper trading, our decisions were made regarding the past trends of tweets that we had gathered. This simulation enabled an evaluation of the effectiveness of the models compared to actual results and the alignment and modification to make them more feasible in a real market environment.

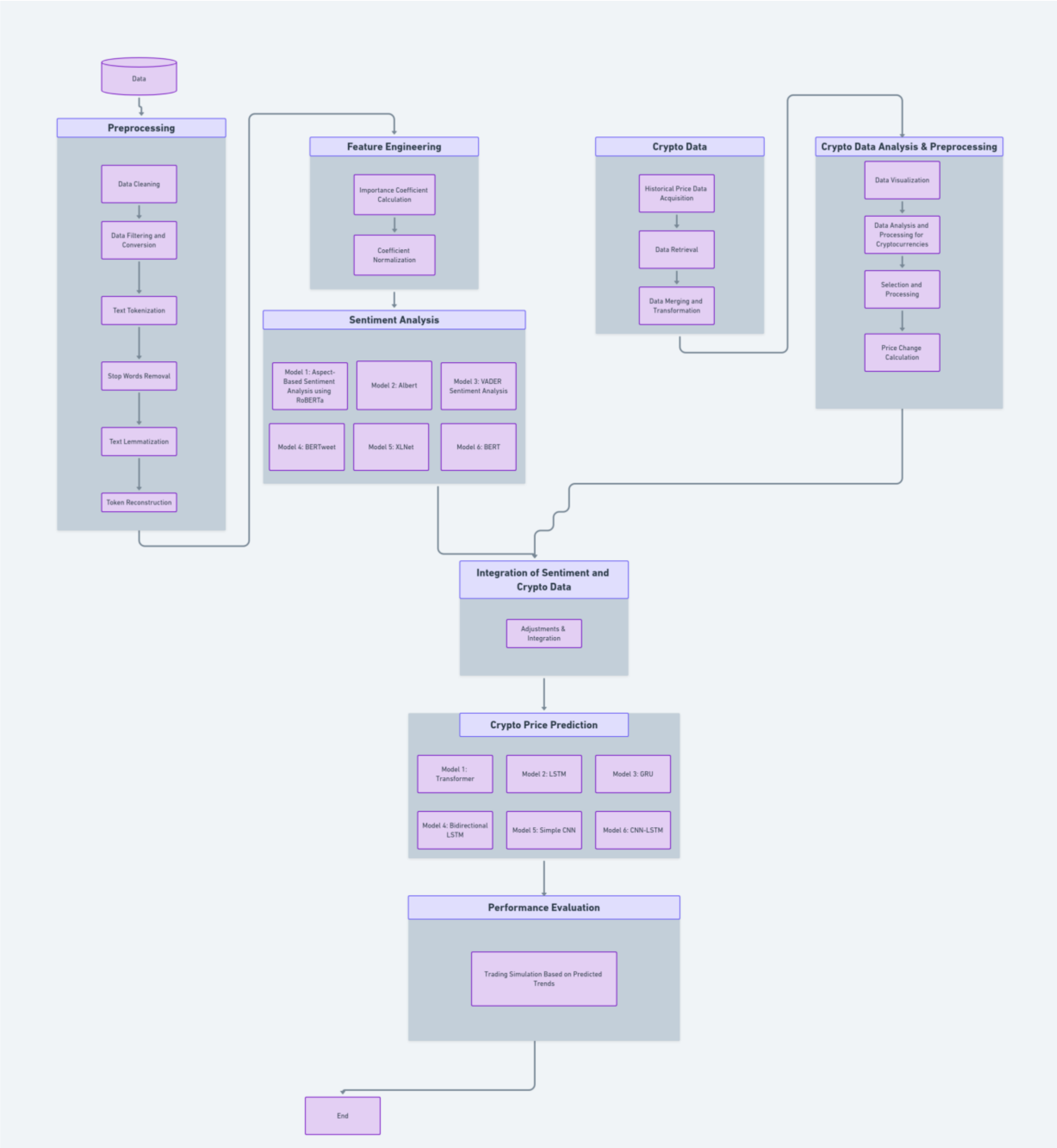

Figure 1. Flow Diagram of the Thesis Methodology

4. Experimental Design and Results

4.1. Model Training and Validation Loss

Table 2 shows the training and validation losses achieved by different models after training for 10 epochs on the merged dataset of tweets and sentiment and stock data. The training loss represents the model's performance on the training data, while the validation loss estimates the model's performance on unseen data. Lower values indicate better model performance.

Table 2. Training and Validation Loss for Different Models (10 epochs).

Model Name	Training Loss	Validation Loss
LSTM	0.0220	0.0817
GRU	0.0233	0.0826
CNN-LSTM	0.0257	0.0989
Transformer	0.0056	0.0773
Bidirectional LSTM	0.0185	0.0832

Simple CNN	0.0252	0.0910

4.2. Validation Accuracy on Financial PhraseBank Dataset

Table 3 presents the validation accuracy achieved by different pre-trained models on the Financial PhraseBank dataset. These models were trained on the labeled dataset and later used to predict sentiment in influential tweets. The validation accuracy measures the model's performance on unseen validation data from the Financial PhraseBank dataset. Higher accuracy indicates better model performance.

Table 3. Validation Accuracy of Different Models on the Financial PhraseBank Dataset.

Model Name	Validation Accuracy
BERTweet	90%
BERT	89%
ALBERT	88%
RoBERTa	91%
XLNet	87%

4.3. Sentiment Prediction Results and Majority Voting

Table 4 shows sample results of sentiment predictions made by different models on influential tweets, along with the final sentiment determined through majority voting. The sentiment predictions from a variety of models such as ALBERT, XLNet, Roberta, Vader, BERTweet, and BERT are displayed in the table. By aggregating the predictions of all models, the final sentiment is determined through majority voting. It harnesses the unique capabilities of multiple models to produce more reliable sentiment predictions.

Table 4. Sample Results of Sentiment Predictions and Final Result Based on Majority Voting.

Albert	XLNet	RoBERTa	Vader	BertTweet	BERT	Result
1	2	2	0.0	0.0	1	1
0	2	2	0.0	2.0	2	2
2	2	2	0.0	2.0	2	2
2	2	2	0.0	0.0	2	2
1	2	2	0.0	2.0	2	2
1	2	2	0.0	2.0	2	2
2	2	2	0.0	1.0	1	1
2	2	2	0.0	0.0	2	2
2	2	2	0.0	2.0	2	2
2	2	2	0.0	2.0	2	2

5. Discussions

5.1. Model Training and Validation Loss

From Table 2, we can note that the Training loss for the Transformer model is the smallest having a training loss of 0.0056 on the training data which signifies its high performance. Nevertheless, several caveats are in order in this regard and the first one is that despite the training losses on (D) decreasing over time, the rate of decrease is deceptive, and the magnitude of loss does not necessarily imply a better generalization on the unseen data. The validation loss gives a truer figure of how the model is going to perform on unseen data.

As for the validation loss, LSTM achieved the lowest with a value of 0.0817. The Transformer model was second, with 0. 0773. In the context of low validation losses observed, the models learned from the training phase have good generalization properties for unseen data. The CNN-LSTM and a Simple CNN had higher validation losses compared to the other models, suggesting an overfitting problem or inability to learn the underlying features in the dataset.

5.2. Validation Accuracy on Financial PhraseBank Dataset

Table 3 shows that the results confirmed by the trained Financial PhraseBank dataset's model validation reveal that the RoBERTa model obtained the highest percentage correctness score of 91 while BERTweet is in the second place with 90 correctness score. From these findings, this therefore shows that the use of these pretrained language models is efficient in the task of capturing the sentiment of financial phrases.

ALBERT model also gives reasonable validation accuracy of 88% while the BERT gave slightly better validation accuracy with 89%. However, the lowest validation accuracy, which is 87% was recorded by the XLNet model highlighting some possible weaknesses of the model in capturing the sentiment of financial texts. These validation accuracies are important to compare the results that derive from the different models when applied to the sentiment analysis on influential tweets.

5.3. Sentiment Prediction Results and Majority Voting

Table 4 provides the results of sample sentiment predictions of various models using influential tweets. From the table above, it is possible to conclude that the majority voting technique is beneficial because it is a way of combining the predictions not only of a single model but several models at once, which increases the reliability of the sentiment classification.

For some rows, for example, the first and seventh ones, it was possible to untangle different misreading produced by the sentiment analysis models, which is an indication of how the majority voting approach can minimize the drawbacks of different models while boosting the consolidated benefits. Also, one can observe the variations of the sentiment predictions given by various models, which highlight the need for ensemble procedures such as majority voting to achieve higher accuracy and stability of obtained results.

6. Conclusions

While pursuing the topic of this research study, I constructed an elaborate framework to assess sentiment originating from key tweets and their outcomes on cryptocurrencies' prices. For the sentiment analysis, I utilized advanced natural language processing tools and deep learning models and integrated the obtained sentiments with historical stock data for the prediction of prices and corresponding trends from the tweets. The key findings and contributions of this study are as follows:

1. **Ensemble Sentiment Analysis:** For sentiment analysis, I used an ensemble method using many pre-trained models including but not limited to BERT, RoBERTa, ALBERT, XL-Net, and BERTweet. As for the strategy of selecting the majority vote of the models as the final output, this strategy performed well in averting some of the shortcomings of these models and enhancing the precision of sentiment predictions of the models.
2. **Aspect-Based Sentiment Analysis:** Besides, we did an aspect-based analysis which enabled me to determine the sentiments on the aspects or topics mentioned in the tweets. It gave a breakdown analysis of the sentiment factors and how they would affect the cryptocurrency prices in case they got implemented.
3. **Model Training and Evaluation:** Hence, we experimented on and assessed different deep learning models such as LSTM, GRU, CNN-LSTM, Transformer, Bidirectional LSTM, and Simple CNN in the context of stock price forecasting and trend identification.
4. **Data Integration and Analysis:** When combining sentiment scores derived from the most influential tweets with the historical prices of cryptocurrencies, we managed to investigate the correlation between the two parameters. This integration proved useful in giving me an outlook of the effects of social media sentiment on financial market and useful in training of different trading strategies depending on predicted market trends.
5. **Visualizations and Interpretability:** We devised detailed graphs for the actual and expected close prices based on the model I created above as well as for the actual and expected trends. These representations proved helpful in the analysis since they assisted in the understanding of results on the model's performance and the correctness of the trends displayed.

The research work given showed the possibility of using social media sentiment analysis and deep learning for financial market analysis and prediction. The work adds to the knowledge pool around Sentiment Analysis and its application in the financial domain and can be used as a literature study for future investigations.

Funding: "This research received no external funding".

Conflicts of Interest: "The authors declare no conflict of interest."

References

1. Ji, Qiang, Ripple, Ronald D., Zhang, Dayong, Zhao, Yuqian, Cryptocurrency Bubble on the Systemic Risk in Global Energy Companies, 2021-2022.
2. Kwansoo Kim and Sang-Yong Tom Lee, The dynamics of cryptocurrency market behavior: sentiment analysis using Markov chains, 2021-2022.
3. Junhuan Zhang, Yunqing Xu, Haodong Wang, Cryptocurrency price bubble detection using log-periodic power law model and wavelet analysis, 2021-2022.
4. Youcef Maouchi, Lanouar Charfeddine, Ghassen El Montasser, Understanding digital bubbles amidst the COVID-19 pandemic: Evidence from DeFi and NFTs, 2021-2022.
5. Ross C. Phillips and Denise Gorse, Predicting cryptocurrency price bubbles using social media data and epidemic modeling, 2017-2018.
6. Haritha GB, Sahana N.B., CRYPTOCURRENCY PRICE PREDICTION USING TWITTER SENTIMENT ANALYSIS, 2023.
7. Abba Suganda Girsang and Stanley, Cryptocurrency Price Prediction Based Social Network Sentiment Analysis Using LSTM-GRU and FinBERT, 2023.
8. Julian Geuder, Harald Kinateder, Niklas F. Wagner, CRYPTOCURRENCIES AS SPECULATIVE BUBBLES: THE CASE OF BITCOIN, 2017-2018.
9. Abba Suganda Girsang and Stanley, Hybrid LSTM and GRU for Cryptocurrency Price Forecasting Based on Social Network Sentiment Analysis Using FinBERT, 2023.
10. Ali Raheman, Anton Kolonin, Igors Fridkins, Ikram Ansari, Mukul Vishwas, Cryptocurrency Price Prediction using Forecasting and Sentiment Analysis, 2022.
11. Ramit Sawhney, Shivam Agarwal, Vivek Mittal, Paolo Rosso, Vikram Nanda, Sudheer Chava, Cryptocurrency Bubble Detection: A New Stock Market Dataset, Financial Task & Hyperbolic Models, 2022.
12. Ozkan Haykir and Ibrahim Yagli, Speculative bubbles and herding in cryptocurrencies, 2022.
13. Yizhi Wang, Florian Horky, Lennart J. Baals, Brian M. Lucey, Samuel A. Vigne, Bubbles all the way down? Detecting and date-stamping bubble behaviours in NFT and DeFi markets, 2022.
14. Ramit Sawhney, Shivam Agarwal, Vivek Mittal, Paolo Rosso, Vikram Nanda, Sudheer Chava, Asset Bubble Detection in the Cryptocurrency Market, 2022.
15. Authors from the University of the Aegean, Department of Financial and Management Engineering, Tweet Sentiment Analysis for Cryptocurrencies, 2021.
16. Evangelos Vasileiou and Petros Koutrakos from the University of the Aegean, The Performance of Cryptocurrencies under a Sentiment Analysis Approach in the Time of COVID-19, 2021.
17. Jia Ming Low, Zi Jian Tan, Tiong Yew Tang, Narishah Mohamed Salleh, Using Sentiment Analysis of Social Media Posts in Cryptocurrency Price Prediction, 2021.
18. Helder Sebastião and Pedro Godinho, Forecasting and trading cryptocurrencies with machine learning under changing market conditions, 2021.
19. Min Shu, Ruiqiang Song, and Wei Zhu, The 2021 Bitcoin Bubbles and Crashes—Detection and Classification, 2021.
20. Authors from the University of the Aegean, Department of Financial and Management Engineering, Deep Learning and Sentiment Analysis-Based Cryptocurrency Price Prediction, 2023.
21. Raj Parekh, Nisarg P. Patel, Nihar Thakkar, Rajesh Gupta, Sudeep Tanwar, Gulshan Sharma, Innocent E. Davidson, and Ravi Sharma, DL-GuesS: Deep Learning and Sentiment Analysis-Based Cryptocurrency Price Prediction, 2022.
22. Markus Frohmann, Manuel Karner, Said Khudoyan, Robert Wagner, and Markus Schedl, Predicting the Price of Bitcoin Using Sentiment-Enriched Time Series Forecasting, 2023.
23. Harsh Sengar, Solving the Mystery of Crypto's Bubble, 2020.
24. F.A. Enoksen, Ch.J. Landsnes, K. Luˇcivjanská, and P. Molnár, Understanding risk of bubbles in cryptocurrencies, 2020.
25. Min Shu and Wei Zhu, Real-time Prediction of Bitcoin Bubble Crashes, 2020.
26. Marc Gronwald, How explosive are cryptocurrency prices?, 2020.
27. Integrate Deep Learning-based Sentiment Analysis on Cryptocurrencies into Predictive Regressions, 2021.
28. Predicting Cryptocurrency Value using Sentiment Analysis, 2019.
29. Franco Valencia, Alfonso Gómez-Espinosa, and Benjamín Valdés-Aguirre, Price Movement Prediction of Cryptocurrencies Using Sentiment Analysis and Machine Learning, 2019.

30. Esakki Elango, E. Manohar ME, and S. Vishnu Durga, Cryptocurrency Market Movement and Tendency Forecasting using Twitter Emotion and Information Quantity, 2023.
31. Cryptocurrency – Sentiment Analysis in Social Media, 2019.
32. What Can Predict Bubbles in Cryptocurrency Prices?, 2019.
33. Cathy Yi-Hsuan Chen and Christian M. Hafner, Sentiment-Induced Bubbles in the Cryptocurrency Market, 2019.
34. Michael Jacobs, The Detection of Asset Price Bubbles in the Cryptocurrency Markets with an Application to Risk Management and the Measurement of Model Risk, 2023.
35. M. Asım Karaömerlio˘ glu, The Cryptocurrency Bubble, Blockchain and Beyond, 2023.
36. Muhammad Asif Khan, Do Bubbles in the Bitcoin Market Impact Stock Markets? Evidence From 10 Major Stock Markets, 2023.
37. Çelik, Ali and Ulu, Ça˘ grı, Testing the Price Bubbles in Cryptocurrencies using Sequential Augmented Dickey-Fuller (SADF) Test Procedures: A Comparison for Before and After COVID-19, 2023.
38. Ikkou Okubo, Kensuke Ito, Kyohei Shibano, and Gento Mogi, NON-FUNGIBLE TOKEN BUBBLE PREDICTION USING EXTENDED LOG-PERIODIC POWER LAW MODEL, 2023.
39. Naila Aslam, Furqan Rustam, Ernesto Lee, Patrick Bernard Washington, Imran Ashraf, Sentiment Analysis and Emotion Detection on Cryptocurrency Related Tweets Using Ensemble LSTM-GRU Model, 2022.
40. jahanbin, kia; Zare Chahooki, Mohammad Ali; Rahmanian, Fereshte, Database of influencers' tweets in cryptocurrency (2021-2023), Mendeley Data, V5, doi: 10.17632/8fbdhh72gs.5, 2023.
41. Ali Raheman, Anton Kolonin, Igors Fridkins, Ikram Ansari, Mukul Vishwas, Social Media Sentiment Analysis for Cryptocurrency Market Prediction, 2022.
42. Pekka Malo, Ankush Sinha, Pekka Korhonen, Jyrki Wallenius, Pyry Takala, Good debt or bad debt: Detecting semantic orientations in economic texts, Journal of the Association for Information Science and Technology, 65(4), 782-796, 2014.