International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Bitcoin Price Prediction Using Sentiment Analysis and Long Short-Term Memory (LSTM)

[1]Ajay Kumar, [2] Varun Srivastava, [3] Mahesh Kumar Chaubey, [4] Dr. Megha Sehgal

Submitted: 21/03/2023 **Revised**: 23/05/2023 **Accepted**: 12/06/2023

Abstract*:* Bitcoin is steadily gaining popularity online and expanding its use in a wide range of transactions. The sentiments are the main driver of its pricing, which is very sensitive. Gains would increase as forecasting accuracy improves. Although there are many statistical methods for predicting prices, accuracy still needs improvement. This research aims to increase the forecast precision of Bitcoin price movements. The volume, polarity, and price variables of the data set have been added by including a new field called sentiment in order to achieve this goal. Twitter is used to determine sentiment, which is then included in the data collection. This work has suggested a dynamic sentiment analysis strategy that uses the long short-term memory (LSTM) to forecast prices with greater accuracy. Prior to and after the sentiment components were incorporated into the data set, prices were forecasted. The incorporation of mood has a beneficial effect on forecast accuracy, according to the findings. As a result, our approach can significantly reduce risk that is brought on by the high fluctuation of the price of bitcoin.

Keywords: *Twitter Sentiment analysis, LSTM, Bitcoin, forecast*

1. Introduction

Cryptocurrencies like bitcoin, which was first presented in 2008, are very popular online. 2017 saw a substantial increase in interest in cryptocurrencies as the price of the most popular cryptocurrency, bitcoin, rocketed to almost $20,000. In comparison to other currencies in use, this currency touches new highs and lows quite quickly. In November, the price of bitcoin reached an all-time high of $68000 before falling back to under $20000 in the months that followed. As a result, bitcoin prices are incredibly unstable, and experts are quite interested in figuring out its pricing pattern. In addition to the straightforward regression and correlation approach, more sophisticated AI algorithms are also available that can be used to forecast and monitor the Bitcoin price.

Bitcoin introduced in 2008 is a type of crypto currency and widely popular on internet. Crypto currencies have gained significant attention in year 2017 with the price of the major cryptocurrency that is Bitcoin skyrocketing to nearly $20,000.

[1] *Bharati Vidyapeeth (Deemed to be university) Institute of management and research, New Delhi*
ORCID ID : 0000-0003-0919-1395
[2] *Computer science department, Thapar Institute of Engineering and Technology, Patiala*
ORCID ID : 0000-0003-4695-8575
[3] *Bharati Vidyapeeth (Deemed to be university) Institute of management and research, New Delhi*
ORCID ID : 0000-0002-8750-9963
[4] *Bharati Vidyapeeth (Deemed to be university) Institute of management and research, New Delhi*
** Corresponding Author Email: kumar.ajai@yahoo.co.in*

New high and low are touched by this currency in a very short time relative to other currency prevailing. Bitcoin hits all time high value of $68000 in November and dropped back to below $20000 in subsequent months. Therefore, Bitcoin prices are extremely volatile and determining its pricing trend is drawing a lot of attention from researchers. Other than the simple regression and correlation method, advance AI algorithms also exist that can be employed to forecast and track the Bitcoin price. Tweets from the microblogging site Twitter are one example of a social media post that can influence the price in the near term. The degree to which sentiments are reflected and the amount of change they can make to price movement are being measured and agreed upon by researchers in an ongoing effort to evolve this method.

In order to predict the price of bitcoin, neural network based deep learning is increasingly used. This technology uses Long Short-Term Memory (LSTM), convolutional neural network (CNN) , or a hybrid model that combines these models. The majority of models use volume and pricing to forecast prices. But clarity is lacking, especially in the short term.

Many well-known works have recommended using social media posts or Twitter posts to ascertain the feelings that are dominating. Twitter can be used to analyse sentiment. The researchers do, however, reflect a variety of viewpoints about the capacity for sentiment. For instance, focusing simply on a few criticised accounts might not accurately depict the situation. Similar to this, a select few influential people might be able to influence tweets to support their story. As a result, sentiment analysis calls for a cautious methodology that is both effective and reliable. Additionally, a promising sentiment analysis tool called Valence Aware Dictionary and Sentiment Reasoner (VADER) has been evolved. When the outcome of an event is to be given to the VADER, it uses both good and negative tweets as well as T-1 day tweets. All of the compiled tweets aid in organising and reaching a conclusion on positive, negative, and organic sentiment. Although LSTM and RNN are analogous, LSTM maintains its usefulness over a lengthy period of time. Additionally, it fixes the RNN's gradient descent problems. For both time series and sequential data, LSTM has produced a positive result.

The remaining text is structured as follows: The literature review is discussed in section 3, and section 4 focuses on the research methodology used and the methods used for data collecting. Section 5 has presented the results and comments. Finally, section 6 presents the conclusion.

2. Bitcoin Background

Bitcoin is a type of crypto currency that is widely used on the internet. It operates on a peer network, and the government or other parties have no say over pricing. Although it is not legally recognised by many governments, it is a widely accepted medium of business transaction. Bitcoin is decentralised, which means it is not controlled by a single authority or administrator. Bitcoin, which is based on block-chain technology, operates through peer-to-peer networks and cryptographic protocols. Because there is no regulating authority for cryptocurrencies, their exchange rates are primarily determined by public perception. The world has become increasingly connected in the twenty-first century. It doesn't take long for news to spread like wildfire [1], [2].

With the evolution of machine learning as an AI domain and corresponding evolution of high-performance processing systems, it is possible to analyze the perception among masses prevailing in real time. Same can be integrated and used for intraday trading of a single currency or portfolio management.

3. Literature Survey

Text classification and sentiment analysis techniques are used to detect digital currency market movement using the Twitter data set [3][4]. Each algorithm's forecast whether the price of Bitcoin will rise or fall during a specific time period. The implementations of Naive Bayes logistic regression and support vector machines in the Scikit Python library were used for the text categorization strategy [5]–[7]. Training and testing on sentiment analysis data requires the same implementation of support vector machines and logistic regression [8]. Although both types of algorithms are trained on the same data set, the fundamental approaches to formatting each model's feature vector is quite different.

Around 2.5 million Tweets regarding S&P 500 companies were run through the authors' sentiment classifier and compared to stock returns in the study trading on Twitter: Using Social Media Sentiment to Predict Stock Returns by Sul et al. [9]. Fast-spreading feeling is more likely to be reflected in a stock price on the same trading day, whereas slower-spreading feeling is more likely to show up on future trading days. Trading tactics are expected to generate annual profit of 11-15 percent based on these forecasts[10], [11].

[12], [13], in their paper Algorithmic Trading of Cryptocurrency Based on Twitter Sentiment Analysis, examined how sentiment in tweets could be utilised to influence investing decisions, with a focus on Bitcoin. The researchers employed supervised machine learning approaches to achieve an hour-by-hour and day-by-day accuracy of above 90%. The authors note out that the 90 percent accuracy was achieved by doing robust error analysis on the input data, which resulted in a twenty five percent improvement in accuracy on average. Both Colianni et al. and Hutto and Gilbert mentioned the presence of noise in their datasets, and the former team saw a significant reduction in error rates after removing the noise[12].

The authors [14] has proposed to include the implicit and explicit knowledge namely common sense that cannot be included in the LSTM, accordingly authors have proposed a new model termed that was termed sentic LSTM. In addition, hybrid LSTM was designed to lay the specific attention to other knowledge aspect of data. In another work, authors [15] proposed the hybrid approach that combines the LSTM and CNN to predict the Arabic sentiment. Authors [15] explored the impact of sentiment on bitcoin price movement by including the emotion theory and lexicon sentiments analysis. Model was applied on image recognition, voice recognition and sentiment analysis from twitter. Further authors have concluded the need of further advancement to deal with high degree of morphology that is inherent in the language such as Arabic.

4. Research Methodology

In order to conduct the experiment for this undertaken work, research methodology adopted can be explained with the help of figure 1. On one side, bitcoin price data were collected, and sentiments detail such as polarity and subjectivity were collected. Finally, Both the datasets were merged.

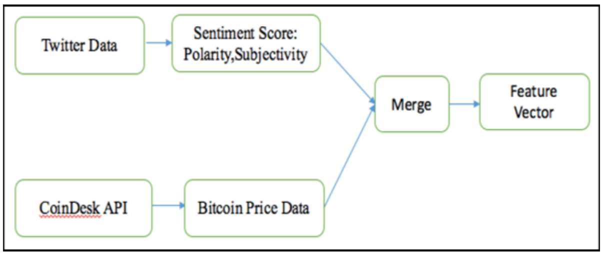

Fig 1: Pre-processing phase

4.1 Data Procurement

To procure the data, two datasets were used: Twitter dataset: contains tweets dated from 27/10/2022 to 22/02/2023 containing the hashtag "bitcoin". Bitcoin pricing dataset: This dataset contains hourly bitcoin pricing data. It includes information such as price, volume, change, timestamp.

For sentiment analysis, data from micro blogging website Twitter has been collected [10]. Twitter enables users to share facts and provide their point of view on topics of their interest ranging from the release of a new mobile phone model to price movement in stocks. In fact, corporations today use this platform to interact with users and to gain insights about the perception of the brand among the masses. Twitter data related to a particular topic can be collected using the tags people attach to them, generally called hashtags. For this research, tags used were bitcoin, altcoin etc.

Tweets can be collected using 3rd party API's like Tweepy. Generally, these API's are run on cloud due to high resource usage need. The data for prediction was obtained from one such source reference. To capture the sentiments on social media, we have used twitter data. The Coindesk's API namely, Bitcoin Price Index API was used to collect hourly prices of bitcoin.

4.2 Data cleaning

The aim of this research was to focus on the people's reaction on the bitcoin and corresponding to that measurement of price movement. This change in pricing might be triggered by market conditions for bitcoins be it in terms of cryptocurrency regulation or the decisions for forking of various currencies, and the mapping of this reaction to bitcoin value [16], [17]. For this, the tweets that do not give any information about sentiment were rooted out. The major sources of such tweets were the twitter handles run by bots. They merely pushed an update about bitcoin value at specific intervals of say one hour. Other sources were twitter handles pushing suspicious tweets such as promise of "free bitcoin".

After rooting out such tweets, the individual tweets were made to go through a series of cleaning processes. These were:

1. Removal of hashtags and hyperlinks

2. Removal of numeric data and punctuations

3. Removal of stop words

4.3 Sentiment analysis

After cleaning of relevant tweets, sentiment analysis was done using Textblob library.

Sentiment was judged using two parameters:

1. Polarity:

This parameter denotes if the sentiment of the tweet was negative or positive. Its value ranged from -1.0 to 1.0 where -1.0 denotes very negative and 1.0 denotes very positive [18], [19].

2. Subjectivity score:

This parameter denotes if the tweet was merely a factor was influenced by one's opinion, and to what degree. Its value ranged from 0.0 to 1.0 where 0.0 denoted very objective and 1.0 denoted very subjective.

4.4 Sentiment summarization

Once the sentiment for individual tweets was known, to infer perception of the masses, the sentiment scores were summarized for all tweets in a particular time frame of 1 hour. Summarization was done by adding up the individual sentiment scores, i.e., all the tweets were given equal importance. To avoid the trap of inferring positive sentiment in an hour of high activity on twitter where most of the tweets denoted neutral sentiment, the sentiment of whole-time frame was divided by the number of tweets in that period.

4.5 Merging

The hourly polarity and subjectivity scores for the tweets were merged with the hourly Bitcoin pricing dataset, containing information about price, volume, change of bitcoin price per hour, corresponding to time frame of the tweets.

4.6 Accuracy computation

In order to compute the accuracy, well known method popularly known as Root mean square error (RMSE) was chosen. The key reason for choosing this method over other was since the target field was numeric, hence value can easily

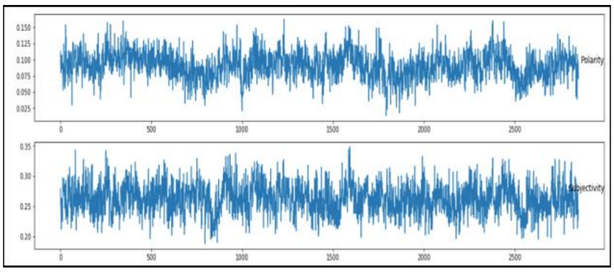

Fig 2: Polarity and subjectivity

be computed with the help of The Root mean square error (RMSE). Other methods such as F1, confusion matrix, accuracy score will be ill suited.

4.7 Model induction

LSTM (Long short-term memory), which a type of RNN (Recurrent Neural Network) model that was used for this work. First the proposed model was employed for the learning of dataset that helps on formulation of rule under which the target values are governed [20]–[21]. The merged dataset has the following attributes: Polarity, Subjectivity, Volume, sentiments and Price. The LSTM model accepts the aforementioned attributes for (t-1) hour and predicts the "Price" attribute for (t) hours.

Table 1: RMSE results

Attributes used	**RMSE Value**
Price (Without sentiment information)	6.1
Polarity, Subjectivity, Volume, Price (With sentiment information)	0.55

5. Results

To evaluate the efficiency of the proposed model, LSTM was used to predict the price. RMSE value obtained is 6.1 and shown in table 1, LSTM model with and without the aid of sentiment score were presented. Afterwards, a new field sentiments prevailing collected from the twitter is included. Target value and other dimensions remained untouched. Result revealed that the RMSE value has dropped substantially and touched the value of .55. Hence, it was concluded that the RMSE has improved when sentiment score added into the dataset.

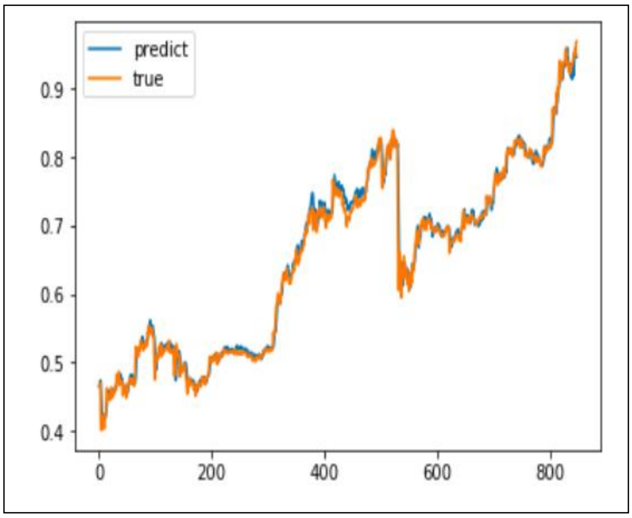

Fig 3: Testing Plot

6. Conclusion & Future Scope

Degree of volatility deeply impact the one's income incurred. Existing techniques in which majority of them are dominated by statistical techniques or machine learning approach are considering the fixed dimensions of the dataset to forecast the bitcoin price. Real-time sentiments of the public is lacking. Owing to the non-involvement of current trend, precision lacks. Proposed system has included the sentiment into the bitcoin database. Thereby, accuracy has been improved. Additionally, forecast will turn more accurate. However, the present approach will not able to reach closely to the price forecast if all the tweets are assigned same weightage. Extensive research is required to investigate the factors that can be detrimental to the regulation of the bitcoin price, which should begin with the parsing of words from the tweet. Including the aforementioned method will not only help to improve accuracy but will also prove to be highly robust. Simultaneously, the frequency of tweet collection must be determined.

7. Acknowledgement

The authors are grateful to the anonymous reviewers for their insightful comments, which helped this article reach its current state.

8. Author contributions

1Ajay Kumar: writing, data curation, 2 Varun Srivastava: methodology ,data curation and analysis, 3 Mahesh Kumar Chaubey: visualization and editing, 4 Dr. Megha Sehgal reviewing and editing.

9. Conflict of Interest

We further clarify that there is no conflict of interest.

References

[1] 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), 339–343. https://doi.org/10.1109/PDP2018.2018.00060

[2] J. Guo, C. Li, G. Zhang, Y. Sun, and R. Bie, "Blockchain-enabled digital rights management for multimedia resources of online education," Multimed. Tools Appl., vol. 79, no. 15–16, 2020, doi: 10.1007/s11042-019-08059-1.

[3] E. Kouloumpis, T. Wilson, and J. Moore, "Twitter Sentiment Analysis: The Good the Bad and the OMG!", ICWSM, vol. 5, no. 1, pp. 538-541, Aug. 2021.

[4] Wang, Q., & Chong, T. T. L. (2021). Factor pricing of cryptocurrencies. The North American Journal of Economics and Finance, 57, 101348. https://doi.org/10.1016/J.NAJEF.2020.101348.

[5] Mukiri, R. kumari, & Vijaya Babu, B. (2021). Prediction of rumour source identification through spam detection on social Networks- A survey. Materials Today: Proceedings. https://doi.org/10.1016/J.MATPR.2021.03.367

[6] L. Cocco, R. Tonelli, and M. Marchesi, "Predictions of bitcoin prices through machine learning based frameworks," PeerJ Comput. Sci., vol. 7, pp. 1–23, 2021, doi: 10.7717/PEERJ-CS.413.

[7] B. Pang and L. Lee, "Opinion mining and sentiment analysis," Foundations and Trends in Information Retrieval, vol. 2, no. 2. pp. 1–135, 2008.

[8] Chhabra, G. (2023). Comparison of Imputation Methods for Univariate Time Series. International Journal on Recent and Innovation Trends in Computing and Communication, 11(2s), 286–292. https://doi.org/10.17762/ijritcc.v11i2s.6148

[9] P. Nakov, A. Ritter, S. Rosenthal, F. Sebastiani, and V. Stoyanov, "SemEval-2016 Task 4: Sentiment Analysis in Twitter," in Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016), 2016, pp. 1–18. doi: 10.18653/v1/S16-1001.

[10] G. P. Dwyer, "The economics of Bitcoin and similar private digital currencies," J. Financ. Stab., vol. 17, pp. 81–91, Apr. 2015, doi: 10.1016/j.jfs.2014.11.006.

[11] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, and R. Passonneau, "Sentiment Analysis of Twitter Data." Association for Computational Linguistics, pp. 30–38, 2011. [Online]. Available: http://www.webconfs.com/stop-words.php

[12] E. Stenqvist and J. Lönnö, "Predicting Bitcoin price fluctuation with Twitter sentiment analysis," degree project technology. 2017.

[13] S. Colianni, S. Rosales, M. S.-C. Project, and undefined 2015, "Algorithmic trading of cryptocurrency based on Twitter sentiment analysis," cs229.stanford.edu, Accessed: Sep. 28, 2022. [Online]. Available: http://cs229.stanford.edu/proj2015/029_report.pdf

[14] P. Paroubek and A. Pak, "Twitter as a Corpus for Sentiment Analysis and Opinion Mining." 2010. [Online]. Available: http://tumblr.com

[15] Y. Ma, H. Peng, T. Khan, E. Cambria, and A. Hussain, "Sentic LSTM: a Hybrid Network for Targeted Aspect-Based Sentiment Analysis," Cognit. Comput., vol. 10, no. 4, pp. 639–650, Aug. 2018, doi: 10.1007/s12559-018-9549-x.

[16] A. M. Alayba, V. Palade, M. England, and R. Iqbal, "A combined CNN and LSTM model for Arabic sentiment analysis," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2018, vol. 11015 LNCS, pp. 179–191. doi: 10.1007/978-3-319-99740-7_12.

[17] Mokni, K., Bouteska, A., & Nakhli, M. S. (2022). Investor sentiment and Bitcoin relationship: A quantile-based analysis. The North American Journal of Economics and Finance, 60, 101657. https://doi.org/10.1016/J.NAJEF.2022.101657

[18] M. Bhuvana and S. Vasantha, "Mediating effect of demonetization of currency notes towards adopting cashless payment system," Int. J. Civ. Eng. Technol., vol. 8, no. 6, pp. 699–707, 2017.

[19] M. Giatsoglou, M. G. Vozalis, K. Diamantaras, A. Vakali, G. Sarigiannidis, and K. C. Chatzisavvas, "Sentiment analysis leveraging emotions and word embeddings," Expert Syst. Appl., vol. 69, pp. 214–224, Mar. 2017, doi: 10.1016/j.eswa.2016.10.043.

[20] Ms. Madhuri Zambre. (2012). Performance Analysis of Positive Lift LUO Converter . International Journal of New Practices in Management and Engineering, 1(01), 09 - 14. Retrieved from http://ijnpme.org/index.php/IJNPME/article/view/3

[21] D. Can, A. Kazemzadeh, F. Bar, H. Wang, and S. Narayanan, "A System for Real-time Twitter Sentiment Analysis of 2012 US Presidential Election Cycle Singing View project Emotion Prediction From Movies View project A System for Real-time Twitter Sentiment Analysis of 2012 U.S. Presidential Election Cycle." Association for Computational Linguistics, pp. 8–14, 2012. [Online]. Available: https://www.researchgate.net/publication/262326668

[22] L. Di Persio and O. Honchar, "Artificial neural networks approach to the forecast of stock market price movements." [Online]. Available: http://iaras.org/iaras/journals/ijems

[23] D. Shadrin, A. Menshchikov, A. Somov, G. Bornemann, J. Hauslage, and M. Fedorov, "Enabling Precision Agriculture through Embedded Sensing with Artificial Intelligence," IEEE Trans. Instrum. Meas., vol. 69, no. 7, pp. 4103–4113, Jul. 2020, doi: 10.1109/TIM.2019.2947125.