Predicting the Price Direction of Bitcoin Using Twitter Data and Machine Learning

Abdul Mannan Kanji
PES Center for Pattern Recognition and Dept. of Computer Science and Engineering
PES University, Bengaluru, India
kanjimannan@gmail.com

Ishita Chaudhary
PES Center for Pattern Recognition and Dept. of Computer Science and Engineering
PES University, Bengaluru, India
ishita1001@gmail.com

Rithika Lakshmi Shankar
PES Center for Pattern Recognition and Dept. of Computer Science and Engineering
PES University, Bengaluru, India
shankar.rithika@gmail.com

Gowri Srinivasa
PES Center for Pattern Recognition and Dept. of Computer Science and Engineering
PES University, Bengaluru, India
gsrinivasa@pes.edu

2022 IEEE 2nd International Conference on Data Science and Computer Application (ICDSCA) | 978-1-6654-7200-5/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICDSCA56264.2022.9988558

Abstract—**Bitcoin is a decentralized digital currency that was intro- duced in 2009 and since then, has become increasingly popular as one of the most known and highly valued currencies. Contributing factors to its rise include crypto Twitter influencers. An engaged audience on Twitter seems to have an influence on the cryptocurrency market. In this paper, we analyze the impact that tweets have on the price of Bitcoin. Using word-clouds and candlestick plots, we gain insight into the factors that affect Bitcoin prices. We also use various machine learning techniques to automatically classify the sentiment in Tweets related to cryptocurrencies. We incorporate these and other relevant features to build and compare the performance of multiple machine learning models to predict the direction (increase or decrease) of the price of Bitcoin.**

Keywords—Cryptocurrency, Price direction prediction, Candlestick plots, Machine learning models

I. Introduction

Cryptocurrency is a digital or virtual currency that is secured by cryptography in which transactions are verified and recorded in a decentralized system, which makes it almost impossible to commit any sort of fraudulent transactions [1]. As of 2021, Bitcoin, the first cryptocurrency, has a market cap that has surged to $1 trillion [2]. As a result, investor interest in digital currencies, both retail and institutional has increased dramatically, prompting more companies in India and around the world to rush in and acquire a stake in this $1 trillion industry by coordinating their operations with the cryptocurrency industry and developing a platform that allows them to profit from this ever-growing digital economy.

Twitter is best known as a microblogging and social networking service on which users post status messages, called "tweets". These microblogs express opinions about multitudes of topics, but it is quickly evolving into something that is potentially more influential and profitable with the rise of cryptocurrencies. Elon Musk, bil- lionaire and the CEO of Tesla, has frequently been tweeting about cryptocurrencies, causing significant price movements in Bitcoin and Dogecoin, all in less than 280 characters. These tweets, while not intended to profit him financially, have a substantial impact on other cryptocurrency investors. The exponential growth in the popularity of Bitcoin has brought about an extensive analysis of Bitcoin prices, which are prone to frequent and extreme fluctuations [3]. Building models that can accurately and reliably predict the direction of cryptocurrency prices is thus an important task, made increasingly challenging by these sudden fluctuations.

The traditional analysis of Bitcoin prices was based on finance and economics; the methods used were the technical analysis method and the fundamental analysis method. However, according to [4], standard economic theories cannot fully explain how Bitcoin prices are formed because there are various characteristics of foreign currency supply and demand involved. The combination of a number of endogenous and exogenous elements are what drives the price of Bitcoin [5]. As a result, considerable research into possible internal and external factors influencing the price of cryptocurrencies was conducted. According to [6], major predictors of both short-term and long-term cryptocurrency values include market-related variables such as trading volume, market beta, and volatility. [7] examined data from the Bitcoin blockchain, including many wallets and unique addresses, block mining difficulty, hash rate, etc., and selected those characteristics that are strongly connected with the price of Bitcoin to create prediction models.

The inability of traditionally used financial models to predict Bitcoin prices led to machine learning models and artificial neural network-based approaches, in particular, becoming increasingly pop- ular [8]. A fundamental problem brought on by the unpredictability of Bitcoin prices is the extraction of features that are informative enough to be used for forecasting. Deep Learning approaches have been helpful in this regard, due to their ability to automatically extract features from data. The prediction accuracy of these models was higher than traditional time series models [9], [10]. Creating hybrid models by combining the best features of different machine learning models has proven to be the most effective technique yet [11].

Hence, in this paper, we present an extensive analysis of the various factors known to affect Bitcoin prices, driven by machine learning. We then predict the direction of price movement using a CNN-GRU model that outperforms other baseline models.

II. Related work

In order to build a successful model to predict the direction of the price movement, our analysis begins by looking into the factors affecting the price of Bitcoin. The work by [12] tests three hypotheses believed to affect this price. They noticed that the demand variables (such as the number of transactions and addresses) had a greater influence on the price of Bitcoin than the supply variables (e.g. number of Bitcoins, etc). The second statistically significant hypothesis confirmed from the tests performed was that an increasing acceptance and trust in Bitcoin (i.e., the investment attractiveness) drove up the price of Bitcoin.

We then proceed to look into the social media aspect by determining the sentiment of the tweets. Sentiment analysis is a subfield of natural language processing study that, at its most basic level, is the process of determining the positive, negative, and neutral opinions of given text. Using a lexicon-based approach to identify sentiments from the headlines of crypto news articles, [13] analyses the overall changes in different cryptocurrencies with multiple new platforms. An interesting aspect of this paper is that they identify the most frequent words, and generate a word cloud to visualize the keywords found using their frequencies. It is also observed that the TF-IDF model input to the SVM performs better than the general model (using entire sentiment words, instead of a weighted approach as in TF-IDF) for identifying the sentiment in terms of accuracy.

According to [14], several distinct modern deep learning meth- ods are compared such as deep neural networks (DNN), LSTM, CNN as well as their combinations to predict the price of Bitcoin. They checked the Spearman correlation coefficient between various features and the Bitcoin prices. The most important feature was found to be the market price followed by the market cap. Upon inspection, although models using LSTM perform better than the models used for regression problems, the DNN-based prediction models achieved outstanding results for classification problems. This consecutively indicated that classification models were more beneficial than regression models. The insensitivity of forecasting methods to time-series data fluctuations is demonstrated by [15]. In the case of insufficient data points or bad data conditioning, erroneous forecasting, such as creating random values that turn out to be outliers, is very prevalent. The classification model, on the other hand, presents a probabilistic perspective of the predictive analysis and hence is a safer bet as it predicts the trend's direction, instead of the price. The authors thus present a novel approach to extracting features from the exponentially smoothed closing price of stocks, to predict the direction of movement of price. They use an ensemble of random forest classifiers, and gradient boosted trees and display their advantages over non-ensemble techniques.

To compare the prediction of Bitcoin prices, [16] use different ma- chine learning and statistical methods. The statistical methods include ARIMA, moving averages, and simple moving averages whereas the ML methods consist of support vector machines, linear regression, RNN, CNN, etc. On comparing the results of both machine learning and statistical methods, it is found that the ML methods perform better in general. It was also determined that seasonality does not play a role in the Bitcoin dataset, adding to the disadvantage of using statistical methods. Finally, it was ascertained that the most successful methods over statistical and machine learning methods are deep learning methods for datasets that contain hidden and nonlinear relationships. In order to combine the benefits of several methods, [17] suggest a method to find the stock closing price of the next day using a stock price forecasting method based on CNN-LSTM. They combine the benefits of CNN which can extract fruitful features from the data, and LSTM which automates the process of finding the best mode for the data and in turn improves the accuracy of the forecasting successfully. [18] also propose a CNN- LSTM network to predict the prices of Bitcoin and also predict the direction of price. They use different macroeconomic variables that affect Bitcoin prices as input. Using an ADAM optimizer, they note that the prediction power of the CNN-LSTM architecture is stronger than either of them individually.

[19] explore a deep learning framework applied on candlestick pat- terns to forecast stock prices. Candlestick charts have been the most popular visual representations of daily prices, currency movements, volumes, or values of technical indices. A CNN and GRU model is used to break down the candlestick chart into sub-charts of 'm' days and predict the price movement using the extracted features of these charts. It aims to predict the price movement more accurately than the actual price itself. It clearly outperformed these state-of-the- art models in predicting the price movement. [20] use two different CNNPred architectures, a 2D and 3D CNNPred to predict the prices of various stocks like Google, Apple, etc using 82 indicators. Despite the usual success of these models in price prediction, the results here were not found to be promising with an F1 score of slightly more than 50%. This shows that just choosing a large number of variables/indicators is not good enough to give an accurate prediction. In [21], they investigate a collection of modern machine learning forecasting methods in order to predict the daily prices of Bitcoin. They compare the different methods using root mean squared error (RMSE). The results prove to show that the gated recurrent unit model with recurrent dropout achieves significantly better results when compared to prevailing models.

Hence, we propose a robust model to predict the direction of Bitcoin's price movement, applying the learnings we have obtained from studying existing models.

III. Data used in this study

The dataset - labelled "Bitcoin Tweets" consists of 1 lakh+ records with 13 columns. The tweets have been collected using the #Bitcoin and #btc hashtag and started on 6th February 2021. They have been restricted to English tweets only, to prevent having a mixed-language dataset. The second dataset comprises "Bitcoin Historical Data". It provides daily opening price, closing price, high, low and various other indicators such as volume traded, hash rate, etc. We also augment this data with the moon phase as an additional technical indicator and a sentiment score for each day. Each row is uniquely identified by the attribute "Date". The list of attributes is shown in Table I below.

TABLE I DEFINITION OF VARIABLES, DATA SOURCE AND CORRELATION SCORE

Feature	Description	Correlation	Source
Price	Daily price in USD	1.000	Investing.com
Open	Price as the exchange opens in USD	0.980	Investing.com
High	Highest price reached by the coin throughout the day in USD	0.992	Investing.com
Low	Lowest price reached by the coin throughout the day in USD	0.990	Investing.com
Volume	Total amount of trading of coin on a particular exchange	-0.337	Investing.com
Change %	Difference between the price 24 hours ago and the current price	0.040	Investing.com
AVBLS	Average block size of Bitcoin	0.074	NASDAQ
CPTRA	Cost per transaction- miners revenue divided by the number of transactions	0.767	NASDAQ
CPTRV	Cost % of transaction value	0.105	NASDAQ
DIFF	Difficulty of finding a block	0.396	NASDAQ
HRATE	Total number of tera hashes per second	0.529	NASDAQ
MIREV	Total value of coinbase block rewards and transaction fees paid to miners	0.887	NASDAQ
MKTCP	Total USD value of the Bitcoin supply in circulation	0.997	NASDAQ
MWNTD	Number of transactions My Wallet Users made per day	0.997	NASDAQ
MWNUS	Number of wallets hosts using our My Wallet Service	0.325	NASDAQ
MWTRV	24hr Transaction Volume of web wallet service	-0.161	NASDAQ
NADDU	Total number of distinct addresses used on the Bitcoin blockchain	0.337	NASDAQ
NTRAN	The number of confirmed Bitcoin transactions daily	0.246	NASDAQ
NTBRL	The average number of transactions per block	0.062	NASDAQ
TOUTV	Total value of all transaction outputs in a day	0.225	NASDAQ
TRFEE	Total value of transaction fees paid to miners	-0.003	NASDAQ
neg count	Total number of negative sentiment tweets on a day	0.024	Curated for this study
Tot count	Total number of tweets on a day	0.049	Curated for this study
Iscore	Sentiment score given to a day	0.024	Curated for this study
PhaseNo	Phase of the moon (0- Waning, 1- Waxing)	-0.005	Curated for this study
Target	Whether price increases or decreases the next day	-0.097	Curated for this study

IV. METHODOLOGY

The pipeline for the study is shown in Figure 1. Initially, we start by preprocessing the data and feature engineering. This is followed by an exploratory data analysis (EDA) to gain a deeper understanding. We then proceed to model the data to find the most accurate predictor for the sentiment of the tweets. Our study continues by finding the words that impact the prices; high and low. Finally, a machine learning model is used to predict the direction of price.

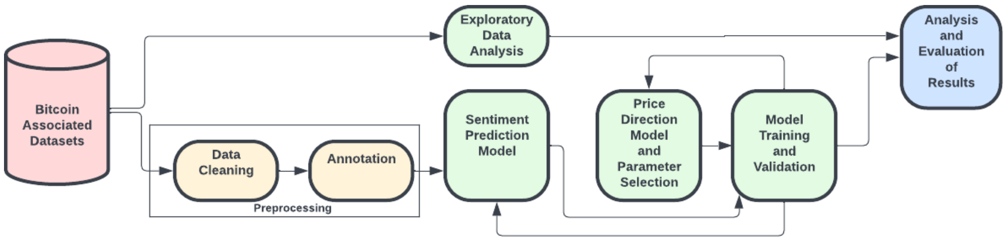

Fig. 1 A schematic diagram of the solution approach to predict the Bitcoin price direction

A. Preprocessing

Before running an analysis, the representation and quality of the data are most important. The process to make it suitable for use is called data preprocessing. There can be many activities for data preprocessing which are used to convert the rough data to quality data. The activities followed by us include the following steps:

- Dropping duplicate values and removing special characters
- For the column "source", the missing values and the values containing the term "bot" are removed, to improve the validity of our model
- Natural Language Processing techniques: We processed the raw text through a series of steps which included tokenization, stop word removal, stemming and lemmatization
- Annotation: Using TextBlob, a Python library for processing textual data, we classified each tweet into one of the three categories- positive, negative and neutral indicated by 1, -1 and 0 respectively

B. Exploratory Data Analysis (EDA)

Sentiment Analysis: The tweets, annotated for sentiment, are then observed for an insight to the distribution of sentiments. To obtain a better visualization of the polarity values, we remove the polarity values that are equal to zero and create a break in the histogram at zero. Figure 2 presents the histogram plots of the polarity values before and after we eliminate tweets that are considered neutral.

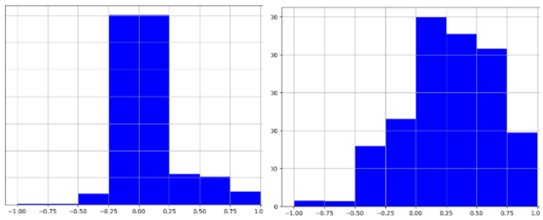

Fig. 2 Distribution of the polarity values of tweets associated with the Bitcoin

Wordcloud: To see what words are most frequently associated with an increasing price, we decided to create a word-cloud. A word-cloud is a visual representation of the most popular words in a document, which help to gain quick insight into the topic for a thorough analysis. To this end, we first separate the tweets created on the days that have a positive price increase. After cleaning and preprocessing these tweets, we build a vocabulary from these words following which the frequency of each word in this list is found. Next, the first 20 words from this list are removed as they are stop words related to this domain, such as cryptocurrency, Bitcoin, blockchain, etc. We find that the most popular words are indicative of the factors previously known to drive up Bitcoin prices, such as increasing demand and the attractiveness of Bitcoin. The words suggesting an increasing demand were 'invest', 'bought', 'buying. Similarly, we looked into building a word-cloud with tweets from days when the price dropped. The most frequent words here were associated with Bitcoin's competitor coins, Ethereum and Dogecoin. Interestingly, 'NFT' and 'metaverse' were also popular terms in these tweets. The wordclouds associated with the increase and decrease in price have been illustrated in Figure 3 and Figure 4 respectively.

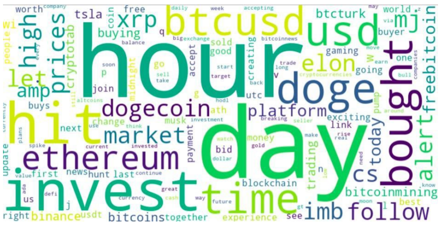

Fig. 3 Wordcloud for tweets associated with an increase in Bitcoin price

Fig. 4 Wordcloud for tweets associated with a decrease in Bitcoin price

C. Factors affecting the price

Correlation: Digging deeper into finding out the impact of various factors on the price of Bitcoin, we come across the work of [22], which studies Bitcoin price formation by considering both the tra- ditional determinants of currency price, such as market forces of supply and demand, and digital currency-specific factors, e.g., Bitcoin attractiveness for investors, as well as interactions between different Bitcoin price determinants. They observed that since supply is ex- ogenous, demand played a higher role than supply in driving Bitcoin prices and that this impact would likely only increase in the future.

[23] found that news related to potential or implemented regulatory policies increased the volatility of Bitcoin prices. This result was consistent with the findings of [24], who state that regulation is a significant factor in determining the price of cryptocurrencies.

To analyze the effect of regulation and competition from other coins on the price, we first created a dictionary of words associated with each of these factors. We then filter out the tweets that contain these terms and find there to be a negative correlation between these and the price. Values of the correlation obtained for these have been mentioned in Table II.

TABLE II CORRELATION TABLE

Factor	Correlation
Regulation	-0.287232
Competition from other coins	-0.318032
Demand	-0.036038

This hypothesis is consistent with the real-world scenario [25], where the price of Bitcoin is seen to be negatively correlated with the price of other coins like Litecoin and Peercoin, and with regulation speculation as well. However, when we performed the procedure on the demand factor, we got a conflicting result. There seemed to be a negative correlation between the price and words. Exploring further as to why this could be, we decided to do a bi-gram and tri-gram analysis to gain more context into these tweets. The initial dictionary consisted of words such as 'want', 'demand', 'hold', 'strong', etc. We observed that most of these 'demand' words were followed by the names of competitor coins–Ethereum, Doge, etc. This explained the negative correlation we found as positive sentiment around competitor coins is known to drive down Bitcoin prices.

Lunar Phases: The lunar buy-sell strategy, simply refers to buying on a new moon and selling on the next full moon (usually 14-16 days). The lunar effect appears to be highest for the smallest-cap equities, which strengthens the researchers' psychological interpretation. Only a small percentage of institutional investors own such stocks, while individuals are more likely to be influenced by the lunar cycle - in part because institutions typically have lengthy decision-making processes, often involving committees, that shield them from their employees' moods.

Fig. 5 Candlestick plot relating Bitcoin price and lunar phases

To showcase the effectiveness of this strategy, we plot a candlestick graph of the Bitcoin prices for the year 2021 along with the positions of the New Moons and Full Moons as shown in Figure 5. We predict that after a New Moon the price will go up and after a Full Moon, the price will go down. The results obtained are summarized in Table III.

TABLE III PREDICTION OF PRICE DIRECTION USING CANDLESTICK PLOT

Trend	F1-score
Down	0.58
Up	0.62
Average F1 score	0.60

Thus, the lunar buy-sell strategy can be used as a decent indicator for investors in making decisions along with our model as an extra attribute since it performs better than a random guess.

D. Models for Sentiment Analysis

We developed the models such as Logistic Regression (LR), Support Vector Classifier (SVC), Complement Naive Bayes (CNB), Random Forest Classifier (RFC) and Decision Tree Classifier (DTC) using the python scikit-sklearn module to predict the sentiment of the tweets and compared the results to identify the best classifier.

TABLE IV SENTIMENT ANALYSIS RESULTS

Model	Accuracy	Recall	Precision
LR	91.7	84.8	93.0
SVC	92.0	85.6	92.9
CNB	80.5	77.2	75.5
RFC	91.7	85.0	92.6
DTC	87.6	82.3	83.4

As shown in Table IV, after the 10-fold cross validation, the sentiment analysis model based on the Support Vector Classifier has the highest F1-score, Accuracy and Recall among these five models. This can be attributed to the fact that SVC works when there is a clear margin of separation and high dimensional space.

E. Models for predicting the direction of price

In order to predict the direction of the price, we compare the following models:

Convolutional Neural Network (CNN)

CNN is a network model proposed by [26] that involves convolution and pooling operations to obtain an activation map that is robust and permutation invariant, hence making it a reliable and efficient model for computer vision and time-series problems. Convolution layer provides a feature map and pooling layer helps in reducing dimensions and helps in making the model robust.

Suppose that we have some N×N matrix as input to our convolutional layer. If we use an m×m filter ω , our convolutional layer output will be of size (N-m+1)×(N-m+1). To find the pre-nonlinearity input to a unit x_{ij}^{l} in our layer, we add up the contributions (weighted by the filter components) from the preceding layer cells using the formula below:

$$x_{ij}^{l} = \sum_{a=0}^{m-l} \sum_{b=0}^{m-l} \omega_{ab} y_{(i+a)(j+b)}$$

Then, the convolutional layer applies its nonlinearity:

$$y_{ij}^{l} = \sigma(x_{ij}^{l})$$

Recurrent Neural Network (RNN)

Recurrent Neural Network [27] is a class of artificial recurrent networks where each network can be thought of as multiple repeating units of networks that are connected successively and the output of one step is input to the next step.

Long Short-Term Memory (LSTM)

LSTM [28] is a modified version of an RNN that is modified to address the vanishing gradient problem of RNNs using special structures called gates. Gates learn what information to keep and what information to discard from memory. It has 3 gates- input, forget and output gate. These gates control how the memory cell remembers information, what it adds, deletes and outputs from the cell.

Gated Recurrent Unit (GRU)

GRU [29] is another variant of an RNN, very similar to an LSTM. The only difference lies in the fact that GRU has only 2 gates, reset and update gates. This results in lesser learnable parameters and faster training times with comparable performance to an LSTM.

CNN-GRU

A CNN-GRU model [30] is known to be effective in sequence learning problems. The CNN is employed to extract feature maps for sequence representation which is followed by the GRU layers to learn this sequence. The GRU models the time-dependency of features. The dense layers, with dropout to prevent overfitting, are used to predict the direction of stock price. The parameter setting of our model is shown in Table V and the architecture of our hybrid model is presented in Figure 6.

TABLE V CONFIGURATION OF CNN-GRU

Layer	Parameter1
Input Layer	input shape = (282,26)
Conv1D + Tanh	units = 64, kernel = 9
MaxPooling1D	pool size = 4
Conv1D + Tanh	units = 32 , kernel = 5
MaxPooling1D	pool size = 2
GRU	units = 64
GRU	units = 64
Flatten	none
Dense + Tanh	units = 32
Dropout	rate = 0.3
Dense + Tanh	units = 16
Dropout	rate = 0.1
Dense + Sigmoid	units = 1

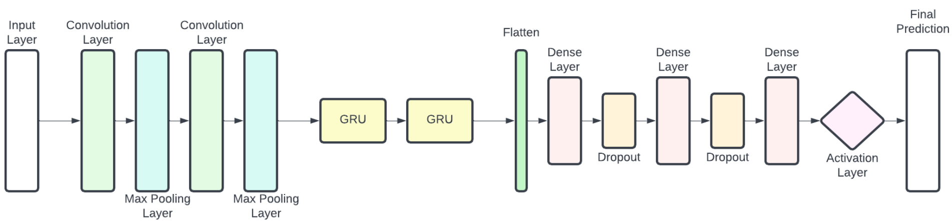

Fig. 6 Components of the CNN-GRU hybrid model architecture

V. RESULTS

A comparison of the performance of different models, viz., the CNN, LSTM, GRU, CNN-LSTM, CNN-GRU, SVC and Logistic Regression for the prediction of the direction of the stock price is presented in Table VI. Models were trained multiple times for 100 *epochs* with validation accuracy as the criteria for selecting the best model for each classifier. We tuned the parameters for each model to get the best results from each model. A 14-day window was used for predicting the direction of price the next day. A dataset split of 80, 10, and 10 (in percentages) for training, validation and test respectively were used for training and testing the models. Recall is the fraction of times an increase in price was predicted of the total number of times the price increased (based on ground truth). Precision is the fraction of times the price actually increased of all the times the model predicted an increase in price. Accuracy is the average of the times the model correctly predicted a rise or fall in price.

TABLE VI PREDICTION OF THE DIRECTION OF PRICE CHANGE

Model	Accuracy	Recall	Precision
Logistic Regression	50.9	51.4	51.9
Support Vector Classifier	51.9	52.1	52.7
CNN	43.48	60	42.9
LSTM	34.78	50	31.2
GRU	47.83	50	45.5
CNN-LSTM	52.17	60	40
CNN-GRU	60.86	54.5	60

It is interesting to note that small-data machine learning models, such as the logistic regressor or support vector classifier, perform better than deep learning models, such as the CNN, GRU and LSTM, trained individually. However, the hybrid models of CNN- LSTM and CNN-GRU outperform their individual counterparts and the small-data machine learning models. In particular, CNN-LSTM demonstrates a higher recall (ability to predict a rise in price the next day) whereas the CNN-GRU presents with a higher precision (greater fraction of its predictions for a rise are true) and accuracy (average of the correct predictions of a rise or fall in price the next day). The results show that through indicators like accuracy (fraction of times the model correctly predicted a rise or fall of price), precision (fraction of times model correctly predicted rise of price out of all its predictions) and recall (fraction of rises predicted out of all the actual rises), the hybrid model of CNN-GRU/LSTM perform better than either of them individually and are superior compared to small data machine learning methods such as the Logistic Regression and Support Vector Classifier.

VI. CONCLUSION

We have presented an analysis and integration of various technical indicators of Bitcoin price with the lunar cycle and daily sentiments of the coin on Twitter to predict the direction of price change.

Experimental results demonstrate that our premise that sentiments of tweets on Bitcoin have a bearing on its price appears to have credence. Further, a comparison of hybrid models such as the CNN- GRU and CNN-LSTM to other deep learning and machine learning models to gain insight into the efficacy of these methods in predicting the direction of price movement for the next day, demonstrates that hybrid models, and in particular, the CNN-GRU, is more effective than using the CNN or GRU individually and better than other models such as the LSTM, support vector classifier and logistic regressor.

SUPPLEMENTARY INFORMATION

In the spirit of reproducible research, the code that accompanies this work, along with a small subset of the data has been made available online [31].

REFERENCES

[1] C. Rose et al., "The evolution of digital currencies: Bitcoin, a cryptocurrency causing a monetary revolution," International Business & Economics Research Journal (IBER), vol. 14, no. 4, pp. 617–622, 2015.

[2] M. C. Pero, "Understanding bitcoin and its utility for special operations forces," Ph.D. dissertation, Monterey, CA; Naval Postgraduate School, 2022.

[3] O. Poyser, "Exploring the determinants of bitcoin's price: an application of bayesian structural time series," arXiv preprint arXiv:1706.01437, 2017.

[4] L. Kristoufek, "Bitcoin meets google trends and wikipedia: Quantifying the relationship between phenomena of the internet era," Scientific reports, vol. 3, no. 1, pp. 1–7, 2013.

[5] E. Bouri, N. Jalkh, P. Molna´r, and D. Roubaud, "Bitcoin for energy commodities before and after the december 2013 crash: diversifier, hedge or safe haven?" Applied Economics, vol. 49, no. 50, pp. 5063–5073, 2017.

[6] Y. Sovbetov, "Factors influencing cryptocurrency prices: Evidence from bitcoin, ethereum, dash, litcoin, and monero," Journal of Economics and Financial Analysis, vol. 2, no. 2, pp. 1–27, 2018.

[7] M. Saad, A. Anwar, S. Ravi, and D. Mohaisen, "Revisiting nakamoto consensus in asynchronous networks: A comprehensive analysis of bitcoin safety and chain quality," in Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security, 2021, pp. 988–1005.

[8] J. Wang and J. Wang, "Forecasting stock market indexes using principle component analysis and stochastic time effective neural networks," Neurocomputing, vol. 156, pp. 68–78, 2015.

[9] S. McNally, J. Roche, and S. Caton, "Predicting the price of bitcoin using machine learning," in 2018 26th euromicro international conference on parallel, distributed and network-based processing (PDP). IEEE, 2018, pp. 339–343.

[10] S. Siami-Namini and A. S. Namin, "Forecasting economics and financial time series: Arima vs. lstm," arXiv preprint arXiv:1803.06386, 2018.

[11] M. Kyoung-Sook and K. Hongjoong, "Performance of deep learning in prediction of stock market volatility." Economic Computation & Economic Cybernetics Studies & Research, vol. 53, no. 2, 2019.

[12] P. Ciaian, M. Rajcaniova et al., "Virtual relationships: Short-and long-run evidence from bitcoin and altcoin markets," Journal of International Financial Markets, Institutions and Money, vol. 52, pp. 173–195, 2018.

[13] Y. L. Ooi, "Analysis of cryptocurrency public sentiment shifts," Ph.D. dissertation, Dublin, National College of Ireland, 2020.

[14] S. Ji, J. Kim, and H. Im, "A comparative study of bitcoin price prediction using deep learning," Mathematics, vol. 7, no. 10, p. 898, 2019.

[15] S. Basak, S. Kar, S. Saha, L. Khaidem, and S. R. Dey, "Predicting the direction of stock market prices using tree-based classifiers," The North American Journal of Economics and Finance, vol. 47, pp. 552–567, 2019.

[16] I. Sibel Kervanci and A. Fatih, "Review on bitcoin price prediction using machine learning and statistical methods," Sakarya University Journal of Computer and Information Sciences, vol. 3, no. 3, pp. 272–282, 2020.

[17] W. Lu, J. Li, Y. Li, A. Sun, and J. Wang, "A cnn-lstm-based model to forecast stock prices," Complexity, vol. 2020, 2020.

[18] Y. Li and W. Dai, "Bitcoin price forecasting method based on cnn-lstm hybrid neural network model," The journal of engineering, vol. 2020, no. 13, pp. 344–347, 2020.

[19] C.-C. Hung and Y.-J. Chen, "Dpp: Deep predictor for price movement from candlestick charts," Plos one, vol. 16, no. 6, p. e0252404, 2021.

[20] E. Hoseinzade and S. Haratizadeh, "Cnnpred: Cnn-based stock market prediction using several data sources," arXiv preprint arXiv:1810.08923, 2018.

[21] A. Dutta, S. Kumar, and M. Basu, "A gated recurrent unit approach to bitcoin price prediction," Journal of Risk and Financial Management, vol. 13, no. 2, p. 23, 2020.

[22] P. Ciaian, M. Rajcaniova, and d. Kancs, "The digital agenda of virtual currencies: Can bitcoin become a global currency?" Information

Systems and e-Business Management, vol. 14, no. 4, pp. 883–919, 2016.

[23] Sˇ. Lyo´csa, P. Molna´r, T. Pl´ıhal, and M. Sˇiranˇova´, “Impact of macroeco- nomic news, regulation and hacking exchange markets on the volatility of bitcoin,” Journal of Economic Dynamics and Control, vol. 119, p. 103980, 2020.

[24] R. Auer and S. Claessens, “Regulating cryptocurrencies: assessing market reactions,” BIS Quarterly Review September, 2018.

[25] N. Gandal and H. Halaburda, “Competition in the cryptocurrency market,” ECONStar: Bank of Canada - working paper, no. 33, 2014.

[26] Y. LeCun, Y. Bengio et al., “Convolutional networks for images, speech, and time series,” The handbook of brain theory and neural networks, vol. 3361, no. 10, p. 1995, 1995.

[27] J. Schmidhuber, “Learning complex, extended sequences using the principle of history compression,” Neural Computation, vol. 4, no. 2, pp. 234–242, 1992.

[28] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[29] K. Cho, B. Van Merrie¨nboer, D. Bahdanau, and Y. Bengio, “On the properties of neural machine translation: Encoder-decoder approaches,” arXiv preprint arXiv:1409.1259, 2014.

[30] M. Sajjad, Z. A. Khan, A. Ullah, T. Hussain, W. Ullah, M. Y. Lee, and S. W. Baik, “A novel cnn-gru-based hybrid approach for short-term residential load forecasting,” Ieee Access, vol. 8, pp. 143 759–143 768, 2020.

[31] “Reproducible research repository,” (Date last accessed 25-July-2022). [Online]. Available: https://drive.google.com/drive/folders/10WSoPFOfmkuYBgOQB-LB4gcQwoZlAzEs