



AI in Cryptocurrency

Alexander I. Iliev^{1,2}(✉), Malvika Panwar¹

¹ SRH University Berlin, Charlottenburg, Germany,
2. Institute of Mathematics and Informatics
Bulgarian Academy of Sciences, Sofia, Bulgaria
8 Acad. Georgi Bonchev Str., Bulgaria, 1113, Sofia
ailiev@berkeley.edu,
malvika.panwar@stud.srh-campus-berlin.de

Abstract.

This study investigates the predictability of six significant cryptocurrencies for the upcoming two days using AI techniques i.e., machine learning algorithms like random forest and gradient for predicting the price of these six cryptocurrencies. The study presents to us that machine learning can be seen as a medium to predict the prices of cryptocurrencies. A machine learning system learns from past data, constructs prediction models, and predicts the output for new data whenever it gets it. Predicted output's accuracy is influenced by the quantity of data since the more data there is, the better the model can predict the output. The results shown that with the accuracy score performance metric, which we employed for this study, we were able to calculate the accuracy of the algorithms and find that both algorithms random forest and gradient boosting respectively performed well for the cryptocurrencies such as Solana (98.07%, 98.14%), Binance (96.56%, 96.85%), and Ethereum (96.61%, 96.60%), with the exception of Tether (0.38%, 12.35%) and USD coin (-0.59%, 1.48%), the results demonstrate that both algorithms work effectively with the majority of cryptocurrencies which can be further increased by using deep learning algorithms like ANN, RNN or LSTM.

Keywords: Cryptocurrency, Artificial Intelligence, Machine Learning, Deep learning, ANN, RNN, LSTM.

1 Introduction

Most developed countries throughout the world adopt cryptocurrencies on a wide scale. Some significant businesses have begun to accept cryptocurrency as a form of payment on a worldwide scale. Microsoft, Starbucks, Tesla, Amazon, and many others are just a few of them. The world today uses cryptocurrency, a form of virtual currency that is protected by cryptography and hence impossible to reproduce, to power its economy. Network-based blockchain technology is used to distribute cryptocurrencies. In its most basic form, the term "crypto" refers to the encryption

of different algorithms and cryptographic techniques that safeguard the entries, such as elliptical curve encryption, public-private key pairs, etc. Cryptocurrency exchanges are easy since they are not tied to any country. These allow users to buy and sell cryptocurrencies using a variety of currencies. Cryptocurrencies are stored in a sophisticated wallet, which is conceptually comparable to a virtual bank account. Block chain is a place where the timestamp information and record of many trades are stored. In a block chain, a square represents each record. Each square has a link to a previous informational square. On the blockchain, the information is encrypted. Only the wallet ID of the customer is made visible during exchanges; their name is not [14]. Crypto currencies can be obtained in a few ways, including as through mining, or buying them on exchanges. Retail transactions are not used by many cryptocurrencies. They serve as a means of exchange and are typically used as assets rather than for regular purchases like groceries and utilities. As intriguing as this new class of investments may sound, cryptocurrencies come with a significant amount of risk, therefore thorough study is required. This paper's main goal is to inform readers about the benefits and drawbacks of investing in cryptocurrencies. New investments are always accompanied by a great deal of uncertainty, so it is important to analyse every aspect of cryptocurrencies to provide users and investors with better information and to provide information that will encourage ethical usage of cryptocurrencies without having to consider the risks and potential failure of the investment. Cryptocurrencies are one of the riskiest but most prominent investments available in the modern technological and economic environment. Therefore, prediction of cryptocurrencies will play a greater role for investors to decide if it is best to invest on cryptocurrency or not. *In this study, we suggest the 6 cryptocurrencies, and we use two ensemble machine learning algorithms—random forest and gradient boosting—to predict these cryptocurrencies for the next two days.*

2 Related work

Early studies on bitcoin argued over whether it was a currency or purely a speculative asset, with most authors favouring the latter conclusion in light of the cryptocurrency's high volatility, extraordinary short-term gains, and bubble-like price behaviour [17; 7; 4; 3]. The idea that cryptocurrencies are just speculative assets with no underlying value prompted research into potential correlations between macro-economic and financial variables as well as other pricing factors related to investor behaviour [10]. Even for more conventional markets, it has been demonstrated that these factors are of utmost significance. For instance, [16] point out that Chinese companies with more attention from regular investors typically have a lower probability of stock price crashes.

The market for cryptocurrencies has exploded during the past ten years [19]. For a brief period, Bitcoin's market value exceeded \$3 trillion, and it became widely accepted as a kind of legal tender [1]. A major turning point in the general use of

blockchain technology was marked by these occurrences. Investors in cryptocurrencies currently have a wide range of choices, from Bitcoin and Ethereum to Dogecoin and Tether. Returns can be calculated in several different ways, and it can be challenging to predict returns for a volatile asset with wide price swings like a cryptocurrency [2].

There has been an increase in interest in cryptocurrency forecasting and profiteering using ML approaches over the past three years. Since the publication of [12], which, to the best of our knowledge, is one of the first studies to address this subject, Table 1 compiles numerous of those papers, presented in chronological order. The goal of this article is to contextualise and emphasise the key contributions of our study rather than to present a comprehensive list of all papers that have been published in this area of the literature. See, for instance, [6] for a thorough analysis of cryptocurrency trading and numerous further references on ML trading.

In the case of other cryptocurrencies, it was investigated whether bitcoin values are primarily influenced by public recognition [11] refer to it—measured by social media news, Google searches, Wikipedia views, Tweets, or comments in Facebook or specialist forums. To forecast changes in the daily values and transactions of bitcoin, Ethereum, and ripple, for instance, [9] investigated user comments and answers in online cryptocurrency groups. Their findings were encouraging, especially for bitcoin [15]. Hidden Markov models based on online social media indicators are used by [13] to create profitable trading methods for a variety of cryptocurrencies. Bitcoin, ripple, and litecoin are found to be unconnected to several economic and financial variables in both the time and frequency domains by [5].

In summary, these papers show that ML models outperform rival models like autoregressive integrated moving averages and exponential moving averages in terms of accuracy and improve the predictability of prices and returns of cryptocurrencies. This is true regardless of the period under analysis, data frequency, investment horizon, input set, type (classification or regression), and method. The performance of trading strategies developed using these ML models and the passive buy-and-hold (B&H) strategy is also contrasted in around half of the research surveyed (with and without trading costs). There is no clear winner in the race between various machine learning models, although it is generally agreed that ML-based strategies outperform passive ones in terms of overall cumulative return, volatility, and Sharpe ratio.

3 Artificial Intelligence and Machine Learning

3.1 Artificial Intelligence

Artificial intelligence (AI) eliminates repetitive activities, freeing up a worker's time for higher-level, more valuable work. Artificial intelligence (AI) is a cutting-

edge technology that can deal with data that is too complicated for a person to manage. Artificial Intelligence may be used to automate tedious and repetitive marketing processes, enabling sales reps to concentrate on relationship development, lead nurturing, etc.

Self-awareness – this is the greatest & most advanced standard of Artificial intelligence. The VP can design a successful strategy using AI data and recommendation systems. To sum it up, Artificial Intelligence looks prepared to function as upcoming for the globe.

Artificial Intelligence would effortlessly manage on patients without real human direction. Artificial Intelligence have programs in several various other fields. Artificial intelligence (AI) is already being employed in almost every industry, offering any company that adopts AI a competitive advantage. Deep learning is a kind of device learning that runs inputs through the biologically inspired neural network architecture. Comprehending the improvement between Artificial intelligence, device reading, and deep training tends to be confusing. With Artificial Intelligence, devices execute features as for example discovering, preparation, reasoning and problem-solving.

3.2 Machine Learning

A machine learning system builds prediction models based on historical data and predicts the results for fresh data whenever it receives it. The amount of data has an impact on the accuracy of the predicted output since the more data the model has, the better it can predict the outcome. The necessity for machine learning in the sector is constantly growing.

Computers are increasingly using machine learning, which enables them to draw lessons from their prior experiences. A variety of approaches are used by machine learning to build mathematical models and make predictions. Currently, the technique is employed for a wide range of purposes, including picture identification, speech recognition, email filtering, Facebook auto-tagging, recommender systems, and many more.

3.2.1 Machine Learning Algorithms

a. Supervised machine learning

Data that encompasses a predetermined label, such as spam/not-spam or a stock price at a specific time, it is known as training data. It must go through a training phase where it is asked to make predictions and is corrected when those predictions are untrue to produce a model. Until the model achieves the level of accuracy desired on the training data set, the training phase is repeated. There are other options, including Back Propagation Neural Networks and Logistic Regression.

b. Unsupervised machine learning

Unsupervised learning techniques are used when the only input variables exist and there are no corresponding output variables. To build models of the underlying data structure, they employ unlabeled training data. Data can be grouped using the clustering technique so that objects in one cluster are more like one another than they are to those in another.

c. Reinforcement machine learning

Reinforcement learning is a type of machine learning that enables agents to choose the best course of action based on their current state by training them to engage in actions that maximize rewards:

Reinforcement learning systems frequently discover the best behaviors through trial and error. Imagine you're playing a video game where you need to go to certain locations at certain times to earn points. A reinforcement algorithm playing that game would first move the character randomly, but over time and with plenty of trial and error, it would eventually figure out where and when to move the character to maximize its point total.

4 Methodology

In this section, we will talk about our approach for prediction of cryptocurrency using proposed machine learning algorithms [18]. For this study, proposed workflow will look like it is shown in figure 1:

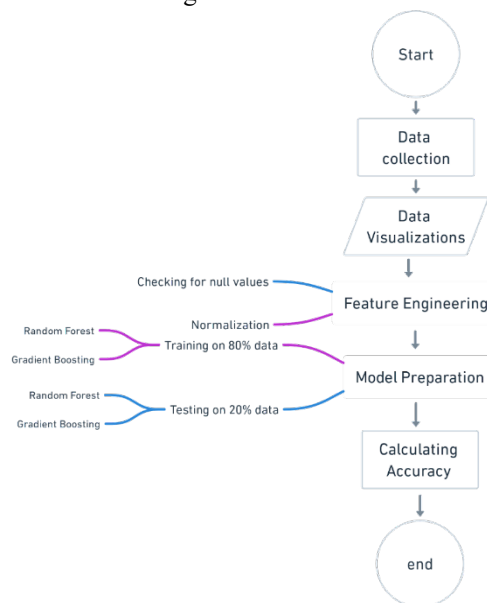


Fig. 1. Proposed Flowchart

4.1 Data Collection

Secondary data is chosen for this study and data is collected from in.investing.com which contains various stock price historical data for free. We have collected top 10 cryptocurrency data for this platform. Below tables show the name of selected cryptocurrency and date of the data taken from.

Table 1. Cryptocurrencies.

S.No.	Cryptocurrency Name	Data collected from date	Data collected till date	Number of observations
1	Binance	1/1/21	10/1/21	375
2	Cardano	1/1/21	10/1/21	375
3	Ethereum	1/1/21	10/1/21	375
4	Solana	1/1/21	10/1/21	372
5	Tether	1/1/21	10/1/21	375
6	USD coin	1/1/21	10/1/21	375

Collected dataset contains following features like Date, Open, High, Low, Volume. Date: Date of the observation, Open: Opening price on the given day, High: Highest price on the given day, Low: Lowest price on the given day, Price-Closing price on the given day, Volume: Volume of transactions on the given day.

4.2 Preliminary data analysis

We will examine the distribution of our data before creating our final machine learning model. Since price is our target variable, we are plotting prices for each cryptocurrency, figuring out average prices after 30 and 2 days, and determining whether our data can accurately predict prices 30 or 2 days in the future. Visualization of each cryptocurrency price is displayed below:

4.2.1 Binance plot

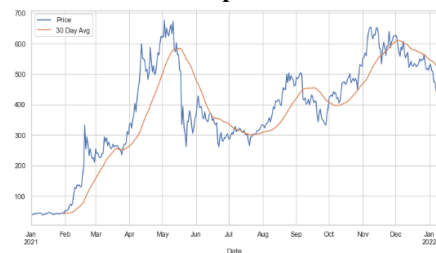


Fig. 2. Binance normal price vs monthly avg price

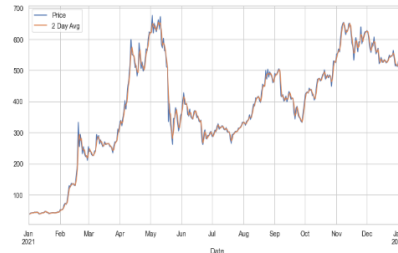


Fig. 3. Binance normal price vs 2 days avg price

Binance price randomness level is less as compared to the price of bitcoin **Fig2,3** and the price of Binance looks like increasing with time, but the current price of Binance is very less as compared to the price of bitcoin. If one wants to invest in cryptocurrency having less price value, then Binance can be a good choice.

Binance monthly avg price doesn't fit properly with normal price **Fig2.**, which means we cannot predict the 30 days future price of Binance, while if we roll a 2-day average against the price of Binance **Fig3.**, we get the clear fit for this problem. Therefore, we can predict the Binance price after 2 days in the future.

4.2.2 Cardano plot

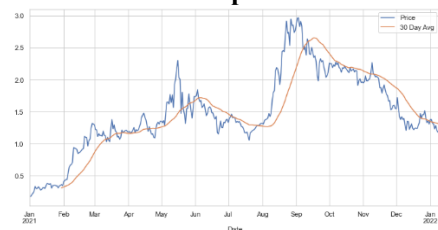


Fig. 4. Cardano normal price vs monthly avg price



Fig. 5. Cardano normal price vs 2 days avg price

Cardano price was at its peak on September 21, figures 4 and 5, but now Cardano price is decreasing also. So, it would be risky for investors to invest on Cardano. Prediction of Cardano is having the same issue as we saw in bitcoin and Binance. So, we can predict the price of Cardano after 2 days in the future as well.

4.2.3 Ethereum plot

From Fig6,7 we will predict the price of Ethereum just after two days in the future.



Fig. 6. Ethereum normal price vs monthly avg price

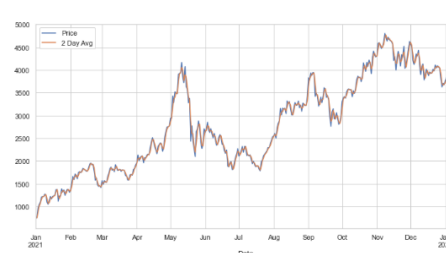


Fig. 7. Ethereum normal price vs 2 days avg price

Price of Ethereum shows an increasing trend but after December 21, the price is decreasing. Price value of Ethereum is quite good and investors can try Ethereum if they want to invest. But before the suggestion, we will check the predicted price of Ethereum on 12th of January to check if the price of Ethereum is showing any

variance i.e., either the price is increasing or decreasing from the price on 10th January.

4.2.4 Solana plot

Here we will predict the price after two days for Solana - figures 8 and 9.



Fig. 8. Solana normal price vs monthly avg price



Fig. 9. Solana normal price vs 2 days avg price

4.2.5 Tether plot

Here we will predict the price after two days for Tether - figures 10 and 11.

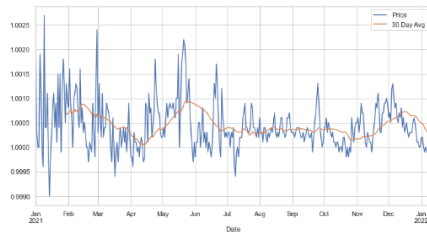


Fig. 10. Tether normal price vs monthly avg price

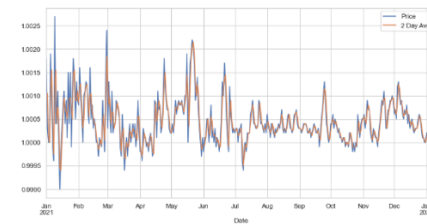


Fig. 11. Tether normal price vs 2 days avg price

4.2.6 USD plot

Here we will predict the price after two days for USD coin - figures 12 and 13.

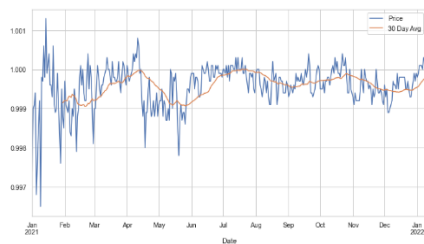


Fig. 12. USD coin normal price vs monthly avg price

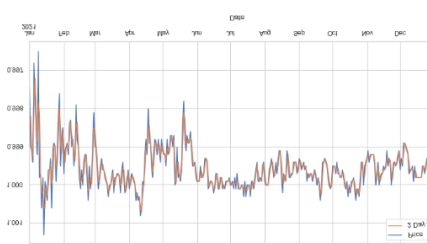


Fig. 13. USD Coin normal price vs 2 days avg price

Thus, from the above plots, the price of cryptocurrency is changing randomly i.e., there's no presence of trend in the price of each cryptocurrency. The monthly average price of cryptocurrencies doesn't fit properly with normal price which mean we cannot predict the 30 days future price of bitcoin, but if we roll a 2-day average against the price of bitcoin then we can predict the prices of each cryptocurrency after 2 days using machine learning.

4.3 Data preprocessing

Collected data is very raw in nature with which machine learning models cannot be applied therefore first data need to be processed. Data preprocessing is the process/methods of transforming data from unstructured data into structured data. Data preprocessing involves many techniques to transform the data. First, we will check if there's any null value present in the data which need to be fixed before applying any algorithms then we will normalize the data which will help to increase the accuracy of our machine learning models.

4.4 Model preparation

In this section, we will fit our machine learning models but before doing that first we will extract our dependent and independent variables because we are using supervised machine learning algorithm. Our dependent variable is "Price after two days" which can be obtained by shifting our price row by 2, and independent variables are Price, Open, High, Low and Ohlc.

$$ohlc\ average = \frac{Open + High + Low + Price}{4}$$

4.4.1 Algorithms

Algorithms that are taken into consideration for this study are: random forest and gradient boosting, both algorithms are very powerful machine learning ensemble techniques.

Random forest

The supervised learning technique known as random forest is typically used for classification and regression. Because it is composed of many decision tree algorithms and creates multiple decision trees depending on the classification samples, this approach is also known as a sort of ensemble learning. Regression trees are constructed using the samples' averages. Although we may predict outcomes from

both continuous (regression) and discrete (classification) data using these techniques, this algorithm performs better for classification issues than regression [8].

Understanding the ensemble technique is necessary to comprehend how random forest systems operate. In essence, ensemble combines two or more models. As a result, rather of using a single model, predictions are based on a group of models. Ensemble uses two different kinds of techniques:

1. **Bagging Ensemble Learning:** As was already said, several algorithms are considered in ensemble learning, but in the bagging method, the dataset is split into a few subsets before the same numerous algorithms are applied to each subset, and the final output data is decided by majority voting. Random forest employs this strategy for classification.
2. **Boosting Ensemble Learning:** Increasing ensemble learning: With this ensemble method, the final model produced provides the highest level of accuracy. This is accomplished by creating a sequential model after pairing weak and strong learners. A couple of these are ADA BOOST and XGBOOST.

Now that we are familiar with the ensemble approaches, let's examine the random forest method, which makes use of the first method, or bagging. The random first method considers the following steps, which are as follows:

Step1: The dataset is split into various subsets where each element is taken randomly.

Step 2: A single algorithm is used to train each subset; in the case of a random forest, a decision tree is employed.

Step 3: Each individual decision tree produces an output.

Step 4: The final output is created by adding the results of each predictor together, identifying the largest result, and then averaging the regression results to determine the final output.

Hyperparameter values taken into consideration: We selected 200 numbers of estimators at 42 random states for random forest algorithm.

Gradient boosting

Gradient boosting is a very powerful algorithm of machine learning. In gradient boosting, many weak algorithms are combined to form a strong algorithm. Gradient boosting can be optimized by changing its learning rate. In this algorithm, we will check our model accuracy with multiple numbers of different learning rates and among those learning rates, we will find the best model for our problem.

Hyperparameter values taken into consideration: Random state = 0, n_estimators = 100

5 Results

We achieved positive results on test data after training the data. The Solana cryptocurrency has the maximum accuracy, at 98.14 percent, while the USD coin and Tether cryptocurrency have the lowest accuracy, as their prices fluctuate erratically (see **Table 2**).

Table 2. Accuracy score of algorithms on each Cryptocurrency.

S.No.	Cryptocurrency Name	Random Forest Accuracy	Gradient Boosting Accuracy	Learning Rate
1	Binance	96.56%	96.85%	0.05
2	Cardano	94.88%	95.11%	0.05
3	Ethereum	96.61%	96.60%	0.05
4	Solana	98.07%	98.14%	0.1
5	Tether	38.00%	12.35%	0.05
6	USD coin	59.00%	14.80%	0.05

Since we last observed the price of each cryptocurrency on January 10th, we are predicting its price for January 12th and comparing it to the actual price of each coin (see **Table 3**).

Table 3. Predicted Results.

S.No.	Cryptocurrency Name	Prediction of price on Jan.12th using Random Forest	Prediction of price on Jan.12th using Gradients Boosting	Actual Price on Jan.12th
1	Binance	435.50	429.70	487.01
2	Cardano	1.20	1.20	1.31
3	Ethereum	3142.00	3131.40	3370.89
4	Solana	141.70	141.20	151.43
5	Tether	1.00	1.00	1.00
6	USD coin	1.00	1.00	0.99

6 Conclusions and Future work

In this study, we suggested the idea of cryptocurrencies and how it helps investors to gain money while taking loss risk into account. We proposed a prediction model employing artificial intelligence (AI) and machine learning, with which we are projecting the prices of the 6 cryptocurrencies for the next two days to lower the risk of loss. However, to discover the optimal algorithm for predicting the performance of these 6 cryptocurrencies, we applied two ensemble machine learning algorithms. The results show that both algorithms perform well with most cryptocurrencies,

except for Tether and USD coin (whose prices changed erratically). The performance metric we used for this study was accuracy score, with which we calculated the accuracy of algorithms and discovered that both algorithms perform well (about accuracy > 95 percent) for the cryptocurrencies Solana, Binance, and Ethereum.

In future, deep learning models like ANN, LSTM can be used to train the model to achieve better results which can be again optimized using optimization techniques like particle swarm optimization, generating algorithms etc. which are part of evolutionary algorithms.

References

1. AF, B. The inefficiency of bitcoin revisited: a dynamic approach. *Econ Lett* 161:1–4 (2017).
2. Anon. Top 10 Cryptocurrencies to Bet on for Good Growth in Feb 2022. Retrieved from <https://www.analyticsinsight.net/top-10-cryptocurrencies-to-bet-on-for-good-growth-in-feb-2022/> (2022).
3. Cheah ET, Fry J. Speculative bubbles in Bitcoin markets? An empirical investigation into the fundamental value of Bitcoin. *Econ Lett* 130:32–36 (2015).
4. Cheung A, Roca E. Crypto-currency bubbles: an application of the Phillips-Shi-Yu Yu. methodology on Mt.Gox Bitcoin prices. *Appl Econ* 47(23):2348–2358 (2015).
5. Corbet S, Meegan A. Exploring the dynamic relationships between cryptocurrencies and other financial assets. *Econ Lett* 165:28–34 (2018).
6. Fang F., V. Carmine. Cryptocurrency trading: a comprehensive survey. Preprint arXiv:2003.11352 (2020).
7. Dwyer G.P. The economics of Bitcoin and similar private digital currencies. *J Financ Stab* 17:81–91 (2015).
8. Jo, T. a. *Advances In Artificial Intelligence, Proceedings of the 13th Brazilian Symposium on Artificial Intelligence (Vol. 1159)*. (b. L. Kaestner, Ed.) Springer (1996).
9. Kim YB, Kim JH. Predicting fluctuations in cryptocurrency transactions based on user comments and replies. *PLoS ONE* 11(8):e0161197 (2016).
10. R. Böhme, N. Christin, B. Edelman, and T. Moore, “Bitcoin: Economics, technology, and governance,” *Journal of Economic Perspectives (JEP)*, vol. 29, no. 2, pp. 213–238 (2015).
11. Li X, Wang C.A. The technology and economic determinants of cryptocurrency exchange rates: the case of Bitcoin. *Decis Support Syst* 95:49–60 (2017).
12. Madan I, Saluja S. Automated bitcoin trading via machine learning algorithms. Retrieved from <http://cs229.stanford.edu/proj2014/Isaac%20Madan,20> (2019).
13. Phillips R C, Gorse. D. redicting cryptocurrency price bubbles using social media data and epidemic modelling. In: 2017 IEEE symposium series on computational intelligence (SSCI). Retrieved from <https://doi.org/10.1109/SSCI.2017.8280809> (2017).
14. A. Kaur, A. Nayyar, “Blockchain a path to the future” *Cryptocurrencies and blockchain technology applications*, John Wiley & Sons, Ltd. (2020)
15. Tiwari AK, Jana R.K. Informational efficiency of Bitcoin—an extension. *Econ Lett* 163:106–109 (2018).

16. Wen F, Xu L. Retail investor attention and stock price crash risk: Evidence from China. *Int Rev Financ Anal* 65:101376 (2019).
17. Yermack, D. Is bitcoin a real currency? An economic appraisal. Berlin: Springer (2015).
18. Catania L, Grassi. S. Predicting the volatility of cryptocurrency time-series. *Mathematical and statistical methods for actuarial sciences and finance*. Springer, Cham, pp 203–207 (2018).
19. Tran VL, Leirvik T. Efficiency in the markets of crypto-currencies. *Finance Res Lett* 35:101382 (2020).