https://doi.org/10.31107/2075-1990-2023-4-123-137

Bitcoin Price Short-term Forecast Using Twitter Sentiment Analysis

Alexey Mikhaylov[1], ***Vikas Khare***[2], ***Solomon Eghosa Uhunamure***[3], ***Tsangyao Chang***[4], ***Diana Stepanova***[5]

[1] *Financial University under the Government of the Russian Federation, Moscow, Russian Federation*
[2] *School of Technology, Management and Engineering NMIMS, Indore, India*
[3] *Cape Peninsula University of Technology, Cape Town, South Africa*
[4] *Department of Finance, Feng Chia University, Taichung, Taiwan*
[5] *Plekhanov Russian University of Economics, Moscow, Russian Federation*

[1] *alexeyfa@ya.ru, https://orcid.org/0000-0003-2478-0307*
[2] *vikaskharekhare@gmail.com, https://orcid.org/0000-0002-9915-5912*
[3] *uhunamures@cput.ac.za, https://orcid.org/0000-0002-8319-5143*
[4] *tychang@mail.fcu.edu.tw, https://orcid.org/0000-0003-1738-4621*
[5] *s_diana@mail.ru, https://orcid.org/0000-0001-5981-6889*

Abstract
The goal of the article is to develop an innovative forecasting approach based on the Random Forest and fuzzy logic models for predicting crypto-asset prices (IFSs, PFSs, q-ROFSs). The baseline forecast horizon is 90 days (additional horizons are 30, 60, 120 and 150 days), which allows to estimate the significance of the chosen features and the impact of time on the forecast accuracy. The paper proposes an optimal data selection approach for the Random Forest and fuzzy logic models to improve the prediction of the daily closing price of Bitcoin, using online social network activity, trading parameters, technical indicators, and data on other cryptocurrencies. This paper utilizes a tree-based machine learning prediction and a fuzzy logic model for Bitcoin. The article attempts to prove that automated Bitcoin forecasting using machine learning algorithms is very effective for the cryptocurrency market. Nevertheless, the latter is characterized by high volatility, significant rate hikes of the most liquid cryptocurrencies (mainly Bitcoin). Therefore, investments in cryptocurrencies, especially long-term ones, involve significant risks. This defines the paper's significance for investors and regulators. As shown by simulation studies of data selection approaches generalizing the accuracy performance of the Random Forest and fuzzy logic models to real preferences of forecasting, even under significant noise measurements, the proposed selection approach leads to fast convergence of estimates. The accuracy of the model's results exceed 85.21 on a 90-day time horizon.

Keywords: *cryptocurrency, investor behavior, Bitcoin, inflation, Twitter sentiment*

JEL: *D53, E31, E44, F21*

For citation: *Mikhaylov A.Yu. et al. (2023). Bitcoin Price Short-term Forecast Using Twitter Sentiment Analysis. Financial Journal, 15 (4), pp. 123–137. https://doi.org/10.31107/2075-1990-2023-4-123-137.*

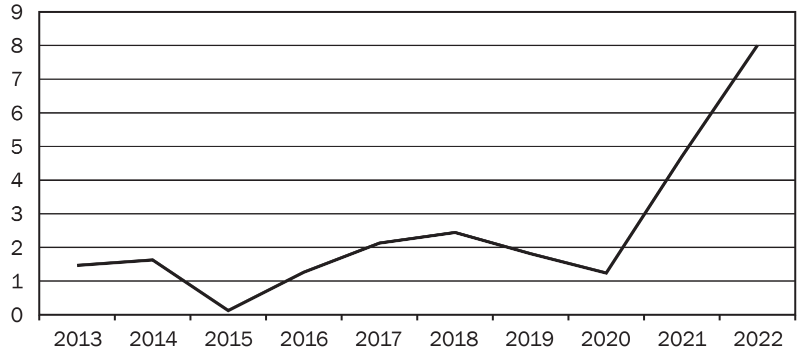
INTRODUCTION

The theoretical basis of the study is the Asset Price Theory (APT). The COVID-19 pandemic ended the longest period of U.S. market growth in history. It began in 2009 and lasted for 11 years. During this period, the S&P 500 index reached an all-time high of 3,386.15 points. But in February 2020, panic selling began — investors feared that the virus would lead to significant losses in global markets. By the end of March, the S&P 500 had collapsed by 33.6%, actually rolling back three years. After that, inflation rose to a historic level of 8–9% (Fig. 1).

The main hypothesis is that widespread government support for the U.S. population during the pandemic caused inflation to rise to these historic levels of 8–9%.

Figure 1

US Inflation in 2018–2022, %



Source: Federal Reserve Economic Data (FRED).

This study proposes an innovative forecasting approach based on the Random Forest and fuzzy logic models for crypto-asset prices forecasting (IFSs, PFSs, q-ROFSs) to predict prices of cryptocurrencies or stocks. The proposed approach is useful in research areas where time series data are used.

The purpose of this paper is to investigate methods for predicting the direction of Bitcoin price. It is also important to measure the influence of several data streams such as social media, other cryptocurrencies and Google Trends to evaluate whether they have any relationship with the BTC price and possibly affect its predicted trajectory [Sun et al., 2020; Sun et al., 2019; Borges and Neves, 2020; Derbentsev et al., 2020].

Since the paper is devoted to forecasting Bitcoin price changes, it is advisable to compare the forecasting results with similar works that focused on the same approaches for prediction of prices (direction of price movement). This paper proves higher accuracy and significance level of results compared to the previous work of [McNally et al., 2018].

These results contribute to the development of tree-based machine learning (ML) approaches for cryptocurrency prediction. This paper also proves that automated Bitcoin rate prediction using machine learning algorithms is very effective for the cryptocurrency market. In addition, this paper contributes to the literature on deep learning techniques in selection approaches [Chen et al., 2020a; Kumar and Rath, 2020; Chen et al., 2020b; Nayak, 2021; Ibrahim et al., 2021; Manahov, 2021; Cherati et al., 2021].

The paper consists of the following sections: Literature Review, Data and Methods, Results, and Conclusions.

LITERATURE REVIEW

From a crypto-investor's perspective, Bitcoin is a highly volatile asset. Since the scope of the study is very limited, so the paper adds novelty in data selection in 8 models. It utilizes the Random Forest and fuzzy logic models (IFSs, PFSs, q-ROFSs).

This paper fills a gap in the existing literature related to automated Bitcoin rate prediction using machine learning algorithms for the cryptocurrency market.

From an academic perspective, the price of a cryptocurrency, as of any other stock, is usually a time-series [Mikhaylov, 2020]. Its ability to calculate the variable importance provides an opportunity for advanced feature engineering and data optimization [Krauss et al., 2017]. Stock price prediction is generally considered as a very challenging task due to the nature of data. Many works incorporate deep learning techniques into selection approaches. They have helped to determine the scope and direction of this study. Perspectives on cryptocurrency price prediction have mainly focused on the use of LSTM, RNN and other tree-based ensembles [Lahmiri and Bekiros, 2019]. However, authors of those works focused on stocks of conventional companies rather than cryptocurrencies [Kang et al., 2018; Uematsu and Tanaka, 2017]. This study focuses on data streaming and feature selection, which have been frequently covered [Kenda et al., 2019; Shahrivari, 2014; Lisin, 2020], including the online methods [Fernandez-Basso et al., 2019].

This study aims to utilize the Random Forest and fuzzy logic algorithms for Bitcoin price short-term prediction in periods of high inflation using Twitter sentiment of the market participants. The short-term forecasts for any financial asset prices with q-ROF Multi-SWARA are investigated by many authors.

The general disadvantages of Random Forest are the problem of overfitting and the large number of hyperparameters to configure. At the same time, the complex branches of the tree are difficult to interpret. The approach used is less susceptible to overfitting. Only past values of the target variable are used as hyperparameters. When applying special methods, the problem of variance is solved.

DATA AND METHODS

In this study, seven Bitcoin constituents from Cryptocompare.com (GOLD, S&P 500, Oil WTI, ETH, Ripple and BNB) and [Tweet Sentiment Visualization, 2022] are used to construct the model. Moreover, to eliminate survivor bias and to optimize the fit by date, data are presented from January 2018 to May 2021.

Data

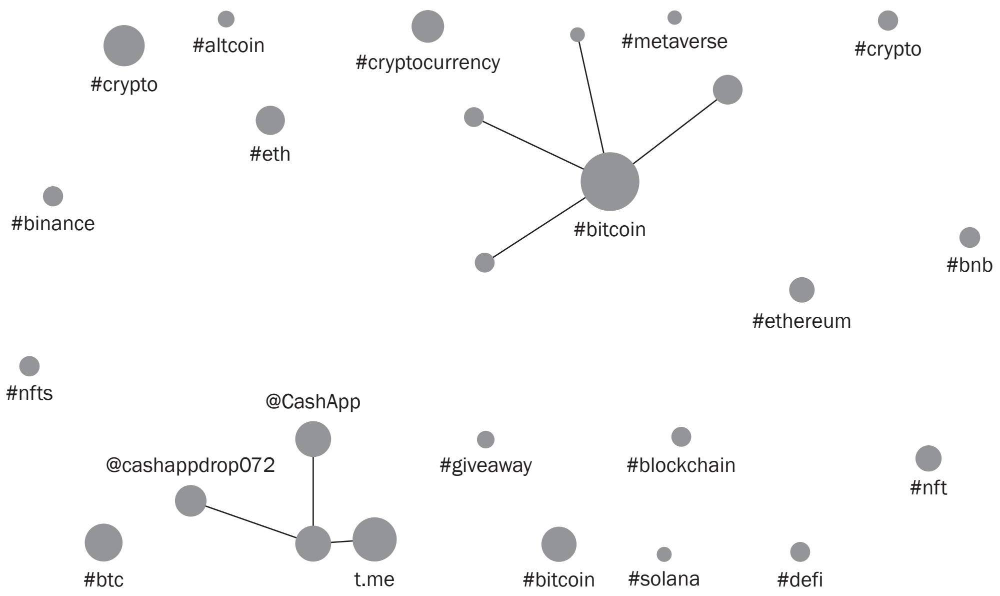
Mentions of Bitcoin from Twitter were collected from the same source. These then need to be consolidated into a single dataset for later use in the model and uploading to Mendeley [Mikhaylov, 2022a, 2022b].

The Random Forest model is relevant for these datasets because tree-based machine learning approaches to cryptocurrency forecasting have been used before (Fig. 2). This paper attempts to prove that automated Bitcoin rate prediction using machine learning algorithms is very effective for the cryptocurrency market. This model provides further development of the methodology and provides an opportunity to improve the base model's efficiency.

The regression model will be used to evaluate the importance of each group of variables in the dataset. Finally, the last model consists of all the features combined and its counterpart with filtered variables. Furthermore, the methods based on the deep learning techniques in selection approaches are used [Kumar and Rath, 2020]. Deep Learning (DL) feature selection approaches are very powerful in feature learning and selection. However, DL methods for automatic feature extraction are not as effective for very volatile daily time series as in the case of Bitcoin [Chen et al., 2020a; Krauss et al., 2017; Guyon et al., 2003] presented the data as a set of a scoring function, where S(i) is used to rank the variables and is computed from xk,l and yk.

Figure 2

The affinity of the word 'Bitcoin' in tweets (the larger the circle size — the bigger the affinity of the word)



Source: created by Authors.

Methodology

The Random Forest regression model

The paper proposes a model based on the Random Forest approach [Friedman et al., 2001; Louppe, 2015] with opportunity to improve its efficiency by adding new trees.

N = {(x1, y1), (x2, y2), ... (xn, yn)}. (1)

The tree is constructed on the basis of values of xn and yn in the training set N (daily changes of Bitcoin price (x) and the dependent parameter (y)). Each of these elements is an individual classifier, where:

K = {k1(x), k2(x), ... kj(x)}, (2)

where j is the number of trees (daily changes of Bitcoin's price).

Each tree also utilizes each variable d in the feature set U and decides whether the error is lower or higher:

U = {di1, di2, ... dim}, (3)

where m is the number of variables (daily changes of Bitcoin's price).

Each tree has a following formula:

Kj(x) = k(x/di). (4)

In research, the regression algorithm`s split criteria is Mean Squared Error:

$$\text{MSE} = \frac{\sum_{t=1}^{n}(g_t - f_t)^2}{n}, \tag{5}$$

where gt is the actual value, ft is the forecasted value, n is the number of data points:

$$\text{F`} = \frac{1}{B}\sum_{i=1}^{B} Fi(x). \tag{6}$$

$$\text{Ij} = \text{wjIvj} - \text{wleft(j)Ivleft(j)} - \text{wright(j)Ivright(j)}, \tag{7}$$

where Ij is the importance of node j, wj is the weighted number of samples, Ivj is the impurity value of node j, left(j) is the left child node, right(j) is the right child node.

The importance of each variable is calculated as:

$$\text{Fii} = \frac{\sum_i^j Ni_j}{\sum_k Ni_k}. \tag{8}$$

$$\text{normFii} = \frac{Fi_i}{\sum_j Fi_j}. \tag{9}$$

The final figure of importance is calculated according to the formula:

$$\text{TFii} = \frac{\sum_j normFi_{ij}}{T}, \tag{10}$$

where TFii is the importance of feature I from all the trees, T is the total number of trees.

Implementation of fuzzy sets

Intuitionistic fuzzy sets allow to get results using degrees.

$$\mathrm{I} = \{\langle \vartheta, \mu_{\mathrm{I}}(\vartheta), \mathrm{n}_{\mathrm{I}}(\vartheta)\rangle / \vartheta \epsilon \mathrm{U}\} \tag{11}$$

$$\mathrm{P} = \{\langle \vartheta, \mu_{\mathrm{P}}(\vartheta), \mathrm{n}_{\mathrm{P}}(\vartheta)\rangle / \vartheta \epsilon \mathrm{U}\} \tag{12}$$

$$0 \le (\mu_{\mathrm{P}}(\vartheta))^2 + (\mathrm{n}_{\mathrm{P}}(\vartheta))^2 \le 1 \tag{13}$$

Fuzzy sets with q-ROFSs are created.

$$\mathrm{Q} = \left\{\langle \vartheta, \mu_{\mathrm{Q}}(\vartheta), \mathrm{n}_{\mathrm{Q}}(\vartheta)\rangle / \vartheta \epsilon \mathrm{U}\right\} \tag{14}$$

$$0 \le \left(\mu_{\mathrm{Q}}(\vartheta)\right)^{\mathrm{q}} + \left(\mathrm{n}_{\mathrm{Q}}(\vartheta)\right)^{\mathrm{q}} \le 1\,,\ \ \mathrm{q} \ge 1 \tag{15}$$

The degree of indeterminacy can be implemented as following:

$$\pi_{\mathrm{Q}}(\vartheta) = \left(\left(\mu_{\mathrm{Q}}(\vartheta)\right)^{\mathrm{q}} + \left(\mathrm{n}_{\mathrm{Q}}(\vartheta)\right)^{\mathrm{q}} - \left(\mu_{\mathrm{Q}}(\vartheta)\right)^{\mathrm{q}}\left(\mathrm{n}_{\mathrm{Q}}(\vartheta)\right)^{\mathrm{q}}\right)^{1/\mathrm{q}} \tag{16}$$

$$\mathrm{Q}_1 = \left\{\langle \vartheta, \mathrm{Q}_1(\mu_{\mathrm{Q}_1}(\vartheta), \mathrm{n}_{\mathrm{Q}_1}(\vartheta))\rangle / \vartheta \epsilon \mathrm{U}\right\} \tag{17}$$

$$\mathrm{Q}_2 = \left\{\langle \vartheta, \mathrm{Q}_2(\mu_{\mathrm{Q}_2}(\vartheta), \mathrm{n}_{\mathrm{Q}_2}(\vartheta))\rangle / \vartheta \epsilon \mathrm{U}\right\} \tag{18}$$

$$\mathrm{Q}_1 \oplus \mathrm{Q}_2 = \left(\left(\mu_{\mathrm{Q}_1}^{\mathrm{q}} + \mu_{\mathrm{Q}_2}^{\mathrm{q}} - \mu_{\mathrm{Q}_1}^{\mathrm{q}}\mu_{\mathrm{Q}_2}^{\mathrm{q}}\right)^{1/\mathrm{q}}, \mathrm{n}_{\mathrm{Q}_1}\mathrm{n}_{\mathrm{Q}_2}\right) \tag{19}$$

$$\mathrm{Q}_1 \otimes \mathrm{Q}_2 = \left(\mu_{\mathrm{Q}_1}\mu_{\mathrm{Q}_2}, \left(\mathrm{n}_{\mathrm{Q}_1}^{\mathrm{q}} + \mathrm{n}_{\mathrm{Q}_2}^{\mathrm{q}} - \mathrm{n}_{\mathrm{Q}_1}^{\mathrm{q}}\mathrm{n}_{\mathrm{Q}_2}^{\mathrm{q}}\right)^{1/\mathrm{q}}\right) \tag{20}$$

$$\lambda Q = \left(\left(1-\left(1-\mu_Q^q\right)^\lambda\right)^{1/q}, \left(n_Q\right)^\lambda\right), \lambda > 0 \tag{21}$$

$$Q^\lambda = \left(\left(\mu_Q\right)^\lambda, \left(1-\left(1-n_Q^q\right)^\lambda\right)^{1/q}\right), \lambda > 0 \tag{22}$$

$$S(\vartheta) = \left(\mu_Q(\vartheta)\right)^q - \left(n_Q(\vartheta)\right)^q \tag{23}$$

M-SWARA and q-ROFSs implementation

Multi-SWARA can be implemented as following:

$$Q_k = \begin{bmatrix} 0 & Q_{12} & \cdots & \cdots & Q_{1n} \\ Q_{21} & 0 & \cdots & \cdots & Q_{2n} \\ \vdots & \vdots & \ddots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ Q_{n1} & Q_{n2} & \cdots & \cdots & 0 \end{bmatrix} \tag{24}$$

$$k_j = \begin{cases} 1 & j = 1 \\ s_j + 1 & j > 1 \end{cases} \tag{25}$$

$$q_j = \begin{cases} 1 & j = 1 \\ \frac{q_{j-1}}{k_j} & j > 1 \end{cases} \tag{26}$$

$$\text{If } s_{j-1} = s_j,\ q_{j-1} = q_j; \text{ If } s_j = 0,\ k_{j-1} = k_j \tag{27}$$

$$w_j = \frac{q_j}{\sum_{k=1}^{n} q_k} \tag{28}$$

w_j shows the coefficient for q-ROFNs feature:

$$X_{ij} = \begin{array}{c} \\ A_1 \\ A_2 \\ A_3 \\ \vdots \\ A_m \end{array} \begin{array}{c} \begin{array}{ccccc} C_1 & C_2 & C_3 & \dots & C_n \end{array} \\ \begin{bmatrix} x_{11} & x_{12} & x_{13} & \dots & x_{1n} \\ x_{21} & x_{22} & x_{23} & \dots & x_{2n} \\ x_{31} & x_{32} & x_{33} & \dots & x_{3n} \\ \vdots & \vdots & \ddots & \dots & \vdots \\ x_{m1} & x_{m2} & x_{m3} & \dots & x_{mn} \end{bmatrix} \end{array} \tag{29}$$

Then

$$\text{q-ROFWA}\ (X_1, X_2, \dots, X_n) = \left(\left(1 - \prod_{i=1}^{n}\left(1 - \mu_{x_i}{}^q\right)^{w_i}\right)^{1/q}, \prod_{i=1}^{n} n_{x_i}{}^{w_i}\right) \tag{30}$$

$$\text{q-ROFWG}\ (X_1, X_2, \dots, X_n) = \left(\prod_{i=1}^{n} \mu_{x_i}{}^{w_i}, \left(1 - \prod_{i=1}^{n}\left(1 - n_{x_i}{}^q\right)^{w_i}\right)^{1/q}\right) \tag{31}$$

RESULTS

Table 1

Performance metric based on S&P 500

RMSE	MAPE	MSE	MAE	PCC	Accuracy, %	Horizon
1697.81	16.5842	2856104	1345.32	−0.1212	78	30
1526.11	14.342	2303091	1315.02	−0.3333	75	60
1697.81	14.6147	2855322	1464.5	−0.0101	80	90
2431.07	21.5635	5853809	2214.93	0.404	60	120
2729.02	24.6945	7374130	3499.65	0.5959	37	150

Source: Authors' calculation.

The Closing Price is a major factor of impact (Table 1).

Table 2

Performance metric based on GOLD

RMSE	MAPE	MSE	MAE	PCC	Accuracy, %	Horizon
924.15	9.8677	845941.7	833.25	−0.404	81	30
1209.98	10.7363	1449919	1020.1	0.7878	82	60
1762.45	15.4833	3077473	1561.46	0.8484	83	90
3081.51	29.1587	9403293	2906.78	0.6262	65	120
1586.71	12.6149	2495727	1319.06	0.6666	80	150

Source: Authors' calculation.

This selection approach has been substantially improved (Table 2).

Table 3

Performance metric based on Oil WTI

RMSE	MAPE	MSE	MAE	PCC	Accuracy, %	Horizon
1196.85	10.605	1419334	1064.54	0.404	75	30
1511.97	11.7766	2266392	1301.89	0.2424	78	60
2082.62	21.7251	4295530	1885.67	−0.3232	75	90
3383.5	56.2671	11340813	2858.3	−0.8484	42	120
2921.93	66.5489	8456872	2777.5	−0.8787	33	150

Source: Authors' calculation.

The result of the selection approach № 3 is similar to № 2. Therefore, it is viable to say that the selected media coverage data alone is insufficient for predicting the price of cryptocurrencies.

As can be seen on the graph, despite the slight positive changes in the metrics, one source does not actually have a substantial impact on the regression model (Table 4). However, three of them might have a slight influence, which will be additionally tested in the following selection approaches.

Table 4

The performance metric based on ETH

RMSE	MAPE	MSE	MAE	PCC	Accuracy, %	Horizon
948.39	9.1102	891331.1	866.58	0.3131	81	30
1238.26	9.8071	1520011	1048.38	0.1717	82	60
1810.93	16.9478	3250598	1493.79	−0.3636	81	90
3696.6	63.9936	13535844	3296.64	0.4444	33	120
2006.87	43.9249	3988619	1870.52	0.202	51	150

Source: Authors' calculation.

Although the metrics are the same as the selection approach № 2, the influence of features is higher than that of user activity (Table 5). This may indicate that Random Forest tries to fit all the variables that compensate each other in the output and do not give an improvement in the model performance. The choice of cryptocurrencies may also influence the results [Moiseev et al., 2023b; Mikhaylov et al., 2023a, 2023b].

Table 5

The performance metric based on Ripple

RMSE	MAPE	MSE	MAE	PCC	Accuracy, %	Horizon
750.43	6.3529	557583.6	592.87	0.2727	83	30
1311.99	9.0092	1706854	1008.99	−0.2121	82	60
1556.41	15.0389	2398418	1244.32	0.7474	82	90
3218.87	55.6712	10258753	2744.17	−0.4141	40	120
1752.35	36.158	3042708	1569.54	−0.2929	61	150

Source: Authors' calculation.

The model specification changes when all features are combined (Table 6).

Table 6

The performance metric based on BNB

RMSE	MAPE	MSE	MAE	PCC	Accuracy, %	Horizon
580.75	4.4238	334366.6	386.83	0.0202	85	30
1642.26	13.4835	2670476	1317.04	0.5151	80	60
2242.2	18.7961	4979899	1924.05	0.8282	78	90
2549.24	21.5332	6431918	2245.23	0.4444	74	120
2548.23	21.0686	6430485	2198.77	0.2525	76	150

Source: Authors' calculation.

The next model uses technical indicators, other cryptocurrencies` trading parameters and user activities (Table 7).

Table 7

The performance metric based on Twitter sentiment analysis of market participants

RMSE	MAPE	MSE	MAE	PCC	Accuracy, %	Horizon
1668.52	19.6445	2755088	1609.94	0.4949	78	30
1258.46	12.7967	1567699	1083.73	0.8484	82	60
1197.86	11.2211	1419655	1038.28	0.8989	85	90
2709.83	23.7552	7272720	2455.31	0.7474	71	120
4273.31	40.5616	18086217	4074.34	0.2828	61	150

Source: Authors' calculation.

According to the above specification, selection based on Twitter sentiment of market participants is slightly better than the one with technical indicators (Table 7). In Table 8, some initially highly influential features were eliminated from the model:

Table 8 compares the obtained accuracy with other research [McNally et al., 2018].

Table 8

Forecasting performance of modeling using Twitter sentiment

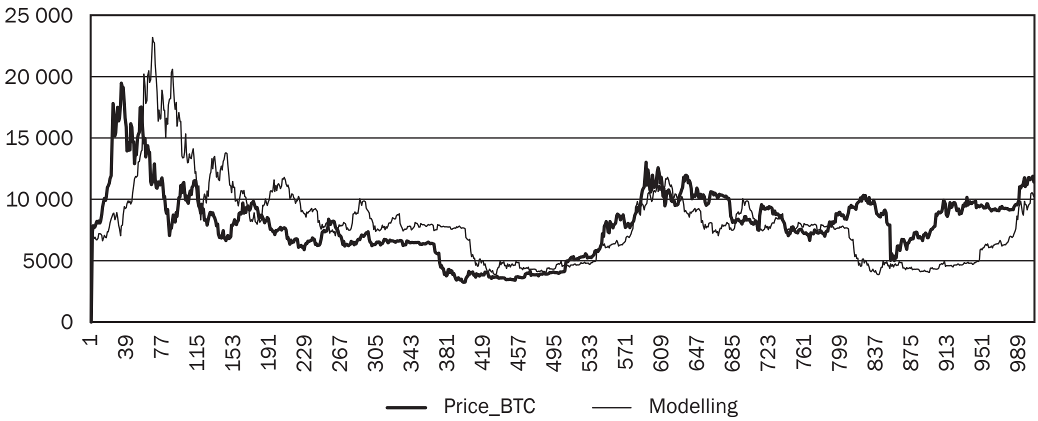
Models	LSTM (McNally)	RNN (McNally)	ARIMA (McNally)	Twitter sentiment
Accuracy	0.528	0.502	0.500	0.852

Source: Authors calculation.

The models would have a very high explanatory power but would not accurately represent reality and would have no practical/economic application [Krauss et al., 2017]. However, another way would be to specifically test the economic significance of forecasts made with the help of the models (Fig. 3).

Figure 3

Historical BTC price and forecast



Sources: Thomson Reuters, authors' calculation.

If the Twitter sentiment model predicts a move of more than 10% in the next 90 days, it is a strong signal for crypto investors. Google Trends` search parameters and multiple technical financial instruments are among the ones which had a very low impact.

The next part of the analysis is Bitcoin price forecasting with q-ROF Multi-SWARA (Table 9).

Table 9

Factors for BTC price prediction

Spillover effect	References
S&P500 (Model 1)	High influence (h)
Oil (Model 2)	Medium influence (m)
Gold (Model 3)	Medium influence (m)
ETH (Model 4)	Medium influence (m)
Ripple (Model 5)	No influence (n)
BNB (Model 6)	No influence (n)
Twitter sentiment (model 7)	Very high influence (vh)

Source: Authors' calculation.

The experts used scale as in Table 10.

Table 10

Membership and non-membership impact for Bitcoin price short-term forecast

Criteria	Membership Impact	Non-membership Impact
No influence (n)	0.15	0.95
Somewhat influence (s)	0.45	0.75
Medium influence (m)	0.50	0.50
High influence (h)	0.75	0.45
Very high influence (vh)	0.95	0.15

Source: Authors' calculation.

The analysis is presented below (Table 11).

Table 11

Data analysis for Bitcoin price short-term forecast

Participant 1				
	P1	P2	P3	P4
P1	M	H		M
P2	M	H	M	
P3	M	H		M
Participant 2				
	P1	P2	P3	P4
P1	M	H		VH
P2	M	H	M	
P3	M	H		VH
P4	M	H	M	
Participant 3				
	P1	P2	P3	P4
P1	M	H		VH
P2	M	H	M	
P3	M	H		VH
P4	M	H	M	

Source: Authors' calculation.

The average values for Bitcoin price short-term forecast (Table 12).

Table 12

The average values for Bitcoin price short-term forecast

	P1		P2		P3		P4	
	μ	v	μ	v	μ	v	μ	v
P1			0.66	0.45	0.89	0.33	0.44	0.14
P2	0.63	0.45			0.62	0.35	0.89	0.27
P3	0.80	0.37	0.85	0.23			0.69	0.46
P4	0.78	0.54	0.97	0.23	0.64	0.47		

Source: Authors' calculation.

The score function for Bitcoin price short-term forecast is calculated as shown in Table 13.

Table 13

Score function for Bitcoin short-term forecast

	P1	P2	P3	P4
P1	0.000	0.169	0.637	0.365
P2	0.247	0.000	0.209	0.514
P3	0.243	0.543	0.000	0.175
P4	0.142	0.576	0.178	0.000

Source: Authors' calculation.

The values of parameters are presented in Table 14.

Table 14

Parameters values for Bitcoin short-term forecast

P1	Sj	kj	qj	wj	P2	Sj	kj	qj	wj
P3	0.649	1.152	0.868	0.317	P4	1.152	0.868	0.317	0.386
P4	1.152	0.868	0.317	0.308	P1	1.152	0.868	0.317	0.307
P2	0.152	1.152	0.631	0.268	P3	0.259	1.152	0.868	0.317

P3	Sj	kj	qj	wj	P4	Sj	kj	qj	wj
P2	0.504	1.152	0.868	0.317	P2	0.514	1.010	1.010	0.371
P1	1.152	0.868	0.317	0.320	P1	0.152	0.172	1.152	0.868
P4	0.152	1.152	0.868	0.317	P3	0.152	1.152	0.868	0.317

Source: Authors' calculation.

The relation matrix for Bitcoin short-term forecast is featured below (Table 15).

Table 15

Relation Matrix for Bitcoin short-term forecast

	P1	P2	P3	P4
P1		0.271	0.415	0.326
P2	0.319		0.307	0.383
P3	0.335	0.424		0.265
P4	0.328	0.389	0.327	

Source: Authors' calculation.

Table 16 shows the stable matrix.

Table 16

Stable Matrix for Bitcoin short-term forecast

	P1	P2	P3	P4
P1	0.27246	0.27246	0.27246	0.27246
P2	0.29298	0.29298	0.29298	0.29298
P3	0.29526	0.29526	0.29526	0.29526
P4	0.2793	0.2793	0.2793	0.2793

Source: Authors' calculation.

The priorities and functions are presented in Tables 17–18.

Table 17

Priorities for Bitcoin short-term forecast

	IFSs	PFSs	q-ROFSs
P1	3	3	2
P2	1	2	3
P3	2	2	1
P4	4	4	4

Source: Authors' calculation.

Table 18

The function values for Bitcoin short-term forecast

	IFWA	IFWG	PFWA	PFWG	q-ROFWA	q-ROFWG
A1	0.57552	0.55481	0.5777	0.55154	0.47742	0.45017
A2	0.43491	0.41747	0.43709	0.4142	0.34662	0.32373
A3	0.50576	0.48505	0.50794	0.48287	0.41093	0.38477
A4	0.7194	0.70305	0.72049	0.70087	0.62239	0.60059

Source: Authors' calculation.

DISCUSSION

Since the paper uses RF as the base classifier, care must be taken in selecting and optimally finalizing the hyperparameters. About 30 hyperparameters were tested in this work, but the most important ones are given in the Table 19, such as:

number of variables;
sample size;
number of trees;
splitting rule;
replacement;
node size.

Table 19

Typical hyperparameters of the random forest and optimal values

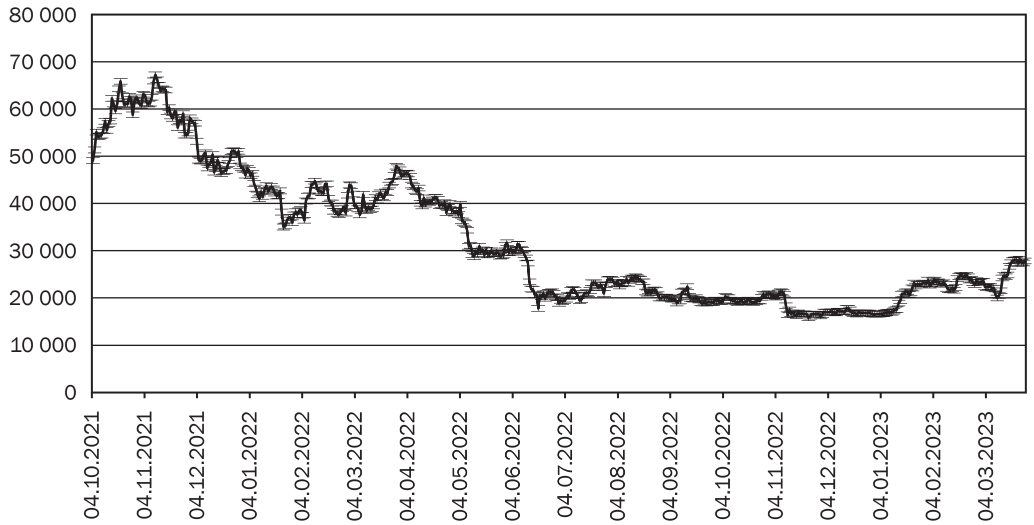
Hyper-parameter	Description	Typical default values	Optimal final tuning hyperparameters in the Twitter sentiment model
Sample size	Number of observations that are drawn for each tree	500	1100
Number of trees	Number of trees	From 500 to 1,000	125
Replacement	With or without replacement	With replacement	With replacement

Source: Authors' calculation.

The results show that investors obtain additional performance if they use the Twitter sentiment prediction model to trade Bitcoin. Investors will receive economic returns (positive and statistically significant returns that exceed the corresponding benchmark strategy, after adjusting for trading costs and risk).

Figure 4

Economic significance of BTC price predictions with positive forecasting border (POS) and negative forecasting border (NEG)



Source: Thomson Reuters, authors' calculation.

Nevertheless, the model does not detect price shocks. It is unable to predict market crashes (COVID-19 related price declines). The investors should follow a buy-and-hold trading strategy to get higher returns with lower risks. If the Twitter sentiment model predicts a price change of more than 10% in the next 90 days, it is a strong signal for crypto investors. The historical price of Bitcoin (BTC) is close to the positive forecasting border (POS) of the model. The negative forecasting border (NEG) of the model is far away from the historical price of Bitcoin (BTC) (Fig. 4).

In other words, the model can be applied in practice. In general, the ultimate goal of any research that involves the development of new forecasting models is to show that they have an advantage over existing alternatives (outperform them) in specific practical applications.

CONCLUSIONS

The research provided evidence on the value of the Random Forest model and fuzzy logic models (IFSs, PFSs, q-ROFSs) for all participants: academic, investors and households. This approach requires further research to compare it with other decision trees, random forest and fuzzy logic models (IFSs, PFSs, q-ROFSs). Utilizing tree-based machine learning forecasting approaches for Bitcoin, the article proved that automated Bitcoin forecasting using machine learning algorithms is very effective for the cryptocurrency market. The second limitation is the source data (Cryptocompare.com) which may be not so useful for future research.

The study confirmed the main hypothesis that approaches to Bitcoin price forecasting using the Random Forest model and fuzzy logic models have high accuracy. However, the cryptocurrency market is characterized by high volatility, significant hikes in the rate of the most popular cryptocurrencies (mainly Bitcoin). Therefore, investments in cryptocurrencies, especially long-term ones, are associated with significant risks. The article has practical applications for investors and policy makers: the results obtained contribute to the development of fine tree-based machine learning approaches to Bitcoin price forecasting. This paper also proved that automated Bitcoin price prediction using machine learning algorithms is highly effective for the cryptocurrency market emerging from globalization. The potential beneficiaries who can use the findings of this paper are investment funds and commercial banks around the world. The article highlighted Bitcoin price prediction methods in accordance with the contribution to the body of knowledge.

References

1. Borges T.A., Neves R.N. (2020). Ensemble of Machine Learning Algorithms for Cryptocurrency Investment with Different Data Resampling Methods. *Applied Soft Computing Journal*, 90, 106–187.
2. Chen W., Xu H. et al. (2020a). Machine learning model for Bitcoin exchange rate prediction using economic and technology determinants. *International Journal of Forecasting*, 37 (1).
3. Chen Z. et al. (2020b). Bitcoin Price Prediction Using Machine Learning: An Approach to Sample Dimension Engineering. *Journal of Computational and Applied Mathematics*, 365, 112395.
4. Cherati M.R. et al. (2021). Cryptocurrency direction forecasting using deep learning algorithms. *Journal of Statistical Computation and Simulation*, 91 (12). https://doi.org/10.1080/00949655.2021.1899179.
5. Derbentsev V. et al. (2020). Forecasting of Cryptocurrency Prices Using Machine Learning. In: Advanced Studies of Financial Technologies and Cryptocurrency Markets. Springer, 211–231. https://doi.org/10.1007/978-981-15-4498-9.
6. Fernandez-Basso C. et al. (2019). Finding tendencies in streaming data using Big Data frequent itemset mining. *Knowledge-Based Systems*, 163, 666–674.
7. Friedman J. et al. (2001). The elements of statistical learning. New York: Springer series in statistics, vol. 1.
8. Guyon I. et al. (ed.) (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3 (7–8), 1157–1182.
9. Ibrahim A., Kashef R. and Corrigan L. (2021). Predicting market movement direction for bitcoin: A comparison of time series modeling methods. *Computers & Electrical Engineering*, 89, 106905. https://doi.org/10.1016/j.compeleceng.2020.106905.
10. Kang Q., Zhou H., Kang Y. (2018). An Asynchronous Advantage Actor-Critic Reinforcement Learning Method for Stock Selection and Portfolio Management, 141–145. https://doi.org/10.1145/3291801.3291831.
11. Kenda K., Kažič B., Novak E. and Mladenić D. (2019). Streaming Data Fusion for the Internet of Things. *Sensors*, 19 (8), 1955.

12. Krauss C., Anh Do X. and Huck N. (2017). Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. *European Journal of Operational Research*, 259. https://doi.org/10.1016/j.ejor.2016.10.031.
13. Kumar D., Rath S.K. (2020). Predicting the Trends of Price for Ethereum Using Deep Learning Techniques, in: Artif. Intell. Evol. Comput. Eng. Syst., Springer, 103–114.
14. Lahmiri S., Bekiros S. (2019). Cryptocurrency forecasting with deep learning chaotic neural networks. *Chaos, Solitons & Fractals*, 118, 35–40. https://doi.org/10.1016/j.chaos.2018.11.014.
15. Lisin A. (2020). Prospects and Challenges of Energy Cooperation between Russia and South Korea. *International Journal of Energy Economics and Policy*, 10 (3). https://doi.org/10.32479/ijeep.9070.
16. Louppe G. (2015). Understanding Random Forest from theory to practice, pp. 25–53.
17. Manahov V. (2021). Cryptocurrency liquidity during extreme price movements: is there a problem with virtual money? *Quantitative Finance*, 21:2, 341–360. https://doi.org/10.1080/14697688.2020.1788718.
18. McNally S., Roche J., Caton S. (2018). Predicting the price of bitcoin using machine learning, in: 2018 26th Euromicro International Conference on Parallel, Distributed and Network-based Processing, PDP, IEEE.
19. Mikhaylov A. (2020). Cryptocurrency Market Development: Hurst Method. Finance: *Theory and Practice*, 24 (3), 81–91. https://doi.org/10.26794/2587-5671-2020-24-3-81-91.
20. Mikhaylov A. (2022a). “Bitcoin”, Mendeley Data, VI. https://doi.org/10.17632/5r89zptx96.1.
21. Mikhaylov A. (2022b). Python code for Bitcoin, Mendeley Data, V3. https://doi.org/10.17632/8jt9sch7yj.3.
22. Mikhaylov A., Dinçer H., Yüksel S. et al. (2023b). Bitcoin mempool growth and trading volumes: Integrated approach based on QROF Multi-SWARA and aggregation operators. *Journal of Innovation & Knowledge*, 8, 3, 100378. https://doi.org/10.1016/j.jik.2023.100378.
23. Mikhaylov A., Dinçer H., Yüksel S. (2023a). Analysis of financial development and open innovation oriented fintech potential for emerging economies using an integrated decision-making approach of MF-X-DMA and golden cut bipolar q-ROFSs. *Financial Innovation*, 9 (1), 4. https://doi.org/10.1186/s40854-022-00399-6.
24. Moiseev N., Mikhaylov A., Dinçer H. et al. (2023b). Market capitalization shock effects on open innovation models in e-commerce: golden cut q-rung orthopair fuzzy multicriteria decision-making analysis. *Financial Innovation*, 9 (1), 55. https://doi.org/10.1186/s40854-023-00461-x.
25. Nayak S.C. (2021). Bitcoin closing price movement prediction with optimal functional link neural networks. Evol. Intel. https://doi.org/10.1007/s12065-021-00592-z.
26. Shahrivari S. (2014). Beyond Batch Processing: Towards Real-Time and Streaming Big Data. Computers, [online] 3(4), pp. 117–129. Available at: https://arxiv.org/pdf/1403.3375 [Accessed 21 Sep. 2019].
27. Sun X., Liu M., Sima Z. (2020). A Novel Cryptocurrency Price Trend Forecasting Model Based on LightGBM, *Finance Research Letters*, 32, 101084.
28. Sun J., Zhou Y., Lin J. (2019). Using machine learning for cryptocurrency trading, in: 2019 IEEE Int. Conf. Ind. Cyber Phys. Syst., IEEE, 647–652.
29. Tweet Sentiment Visualization (2022). Available at: https://www.csc2.ncsu.edu/faculty/healey/tweet_viz/tweet_app/.
30. Uematsu Y., Tanaka S. (2017). High-dimensional Macroeconomic Forecasting and Variable Selection via Penalized Regression. *The Econometrics Journal*, 22 (1), 34–56. https://doi.org/10.1111/ectj.12117.

Information about the authors

***Alexey Mikhaylov**, Candidate of Economic Sciences, Head of laboratory, Financial University under the Government of the Russian Federation, Moscow, Russian Federation*
***Vikas Khare**, PhD (Economics), Associate Professor, Department of Electrical, NMIMS Indore, Indore, India*
***Solomon Eghosa Uhunamure**, PhD (Economics), Associate Professor, Cape Peninsula University of Technology, Cape Town, South Africa*
***Tsangyao Chang**, Doctor in Economics, Professor, Department of Finance, Feng Chia University, Taichung, Taiwan*
***Diana Stepanova**, Candidate of Economic Sciences, Associate Professor, Plekhanov Russian University of Economics, Moscow, Russian Federation*

Article submitted January 16, 2023
Approved after reviewing July 8, 2023
Accepted for publication August 10, 2023

https://doi.org/10.31107/2075-1990-2023-4-123-137

Краткосрочный прогноз цены биткоина с использованием индикатора анализа настроений Twitter

Алексей Юрьевич Михайлов, *кандидат экономических наук, заведующий лабораторией, Финансовый университет при Правительстве Российской Федерации, г. Москва, Российская Федерация*
E-mail: alexeyfa@ya.ru, ORCID: 0000-0003-2478-0307
Викас Кхаре, *доктор экономических наук, профессор, Школа технологии, менеджмента и инжиниринга NMIMS, г. Индор, Индия*
E-mail: vikaskharekhare@gmail.com, ORCID: 0000-0002-9915-5912
Соломон Эгоса Ухунамуре, *доктор экономических наук, профессор, Технологический университет полуострова Кейп, г. Кейптаун, ЮАР*
E-mail: uhunamures@cput.ac.za, ORCID: 0000-0002-8319-5143
Цангяо Чан, *доктор экономических наук, профессор, Финансовый факультет, Университет Фэн Чиа, г. Тайчжун, Тайвань*
E-mail: tychang@mail.fcu.edu.tw, ORCID: 0000-0003-1738-4621
Диана Игоревна Степанова, *кандидат экономических наук, доцент, Российский экономический университет им. Г. В. Плеханова, г. Москва, Российская Федерация*
E-mail: s_diana@mail.ru, ORCID: 0000-0001-5981-6889

Аннотация
Целью этой статьи является автоматическое прогнозирование цены биткоина с помощью алгоритмов машинного обучения. Для этого разработан и протестирован инструмент, основанный на моделях случайного леса и нечеткой логики для прогнозирования цен на биткоин (IFSs, PFSs, q-ROFSs). Базовый горизонт для прогнозирования цены биткоина составляет 90 дней (дополнительные горизонты составляют 30, 60, 120 и 150 дней), чтобы оценить значимость горизонта прогнозирования и оценки сентимента в социальных сетях на точность прогнозирования.
В статье предлагается оптимальный подход к выбору данных из временных рядов для алгоритма случайного леса и моделей нечеткой логики в целях улучшения прогноза дневной цены закрытия биткоина с использованием активности инвесторов в социальных сетях в интернете, торговых параметров, технических индикаторов, а также данных других криптовалют. Однако рынок криптовалют характеризуется высокой волатильностью, значительными скачками курса наиболее популярных криптовалют, поэтому инвестиции в криптовалюты, особенно долгосрочные, связаны со значительными рисками. Вот почему эта статья интересна для инвесторов и регуляторов рынка.
Как показали имитационные исследования подходов к выбору данных и показатели точности моделей случайного леса и нечеткой логики, оптимальный авторский подход приводит к быстрой сходимости оценок. Точность результатов модели превышает 85,21% на 90-дневном временном горизонте.

Ключевые слова: *криптовалюта, поведение инвесторов, биткоин, инфляция, настроения в «Твиттере»*

JEL: *D53, E31, E44, F21*

Финансирование: *Исследование выполнено за счет гранта РНФ № 23-41-10001, https://rscf.ru/project/23-41-10001/.*

Для цитирования: *Mikhaylov A.Yu. et al. (2023). Bitcoin Short-Term Forecast Using Twitter Sentiment // Financial Journal, vol. 15, no. 4, pp. 123–137.*
https://doi.org/10.31107/2075-1990-2023-4-123-137.

Статья поступила в редакцию 16.01.2023
Одобрена после рецензирования 08.07.2023
Принята к публикации 10.08.2023