Volume XXVII 2024
ISSUE no.2
MBNA Publishing House Constanta 2024

Scientific Bulletin of Naval Academy

SBNA PAPER • OPEN ACCESS

Bitcoin Price Forecasting using Blockchain Metrics, Ensemble Machine Learning Algorithms and Facebook Prophet

To cite this article: Simona-Vasilica Oprea and Adela Bâra, Scientific Bulletin of Naval Academy, Vol. XXVII 2024, pg. 43-52.

Submitted: 23.05.2024
Revised: 27.09.2024
Accepted: 07.10.2024

Available online at www.anmb.ro

ISSN: 2392-8956; ISSN-L: 1454-864X

doi: 10.21279/1454-864X-24-I2-004

Bitcoin Price Forecasting using Blockchain Metrics, Ensemble Machine Learning Algorithms and Facebook Prophet

Simona-Vasilica Oprea, Adela Bâra

Bucharest University of Economic Studies, Department of Economic Informatics and Cybernetics, no. 6 Piaţa Romană, Bucharest, Romania
Simona-Vasilica Oprea and simona.oprea@csie.ase.ro

Abstract. As one of the most well-known cryptocurrencies, Bitcoin has gradually been embraced, showcasing the potential to emerge as a common method of payment. The year 2021 witnessed considerable fluctuations in Bitcoin's value, sparking a vast amount of research published in the Web of Science since the first studies on Bitcoin and cryptocurrencies appeared in 2012. The volatility of 2021, marked by successive bear and bull markets, highlighted the limitations of existing prediction tools. The main concern with Bitcoin lies not in its volatility, but rather in its unpredictability. Effective forecasting is crucial as it enhances retailers' strategies towards Bitcoin, encouraging its broader adoption. This study aims to conduct mid-term price predictions with a five-day outlook, utilizing Bitcoin previous prices, blockchain metrics and Twitter sentiments as input data. The prediction technique employs a combination of ensemble machine learning methods, including bagging, boosting, voting, and stacking, to achieve its forecasts. The analysis covers daily data collected from January 1, 2019, to May 31, 2022. For longer-term forecasts, the Facebook Prophet tool is used, suitable for univariate data sets only, demonstrating commendable accuracy for long-term projections. The Mean Absolute Error (MAE) metric is used to evaluate the precision of these predictions.

1. Introduction

Over the years, trading has evolved significantly, transitioning from barter systems to the use of fiat currency, advancements in banking, and the advent of blockchain-based cryptocurrencies. Bitcoin, since its inception, has captured significant attention, gradually establishing itself as a more reliable long-term investment option [1].

Both mid- and long-term predictions of Bitcoin price are challenging. The goal of such predictions is to make the Bitcoin market more predictable and increase the trust in cryptocurrencies. Most of the scientific works focused on the short-term predictions, but the very volatile Bitcoin price in 2021 discouraged such approaches due to the very high errors that were encountered by prediction tools [2], [3]. Some investors are scared by bear markets and usually sell when the prices start to go down. Some investors wait for the prices to go down and then buy the assets. Nevertheless, they are interested in knowing the evolution of Bitcoin in advance. The existing data regarding the blockchain transactions can offer valuable insights into the mid-term evolution of the prices. Several on-chain metrics are used as input to predict the prices [4] for the next five days.

Moreover, knowing the Bitcoin price evolution for some months ahead brings advantages in trading with Bitcoin. Therefore, a strong prediction tool – Facebook Prophet is set up to predict the prices for the following 4 months. The input data is open data downloaded from Glassnode platform, Twitter, Binance and Kaggle.

2. Literature review

More than 9,000 papers were written and published about Bitcoin and cryptocurrencies according to Web of Science platform [5]. The interest in Bitcoin increased gradually and more and more scientific papers are focused on various topics such as blockchain platform [6], dark markets, transitions [7], price forecast [8], [9], [10], security of the trading, transparency, relationship with macroeconomics [11], and facilities of payment for various users [12].

Due to their complexity and efficiency, ensemble models are one of the state-of-the-art models used in practice to provide predictions in the uncertain environment [13], [14]. They were applied in various fields such as energy, air pollution, etc. Usually, ensemble models rely on the forecasting ability of several performant models.

Moreover, Facebook Prophet is popular for predictions and sometimes competes with the well-known models such ARIMA [15]. It is used in forecasting weather conditions, including hydrology [16], PV systems output, solar irradiation [17], economic variables such as sales and stock prices [18]. It was employed for predicting the Bitcoin price as well and provided good results [19].

3. Methodology

For prediction, we create Python pipelines in order to obtain a mid or long-term forecast of the Bitcoin prices. The pipeline of algorithms for the mid-term prediction consists of: Random Forest (RF) that is a bagging model, three boosting models: eXtreme Gradient Boosting (XGB), Histogram Gradient Boosting (HGB) and Light Gradient Boosting (LGB), a Voting Regressor (VR) that embeds the results of RF, XGB, HGB and LGB and a stacking model that includes the voting regressor results and is a meta-model from the ensemble methods. The Voting Regressor combines the predictions from all the individual models. It can combine different models including linear regression, Support Vector Machine (SVM), Multi-Layer Perceptron (MLP) and other Artificial Neural Network (ANN) models.

There are different ways to combine these predictions: averaging, weighted averaging, median where the median of the predictions is considered as the final prediction mitigating the effect of outliers. Stacking is an ensemble machine learning technique that combines the predictions of multiple base models (also known as level-0 models) to make a final prediction using a meta-model (level-1 model). Stacking is a form of model stacking or blending that aims to improve the overall predictive performance of a machine learning system by leveraging the strengths of different models – as in figure 1. A meta-model is a simple model like linear regression or another machine learning model (decision tree, gradient boosting), trained on the meta-features (predictions) generated by the base models (level-0 models). The goal of the meta-model is to learn how to combine the predictions of the base models to make a final prediction that optimizes predictive performance.

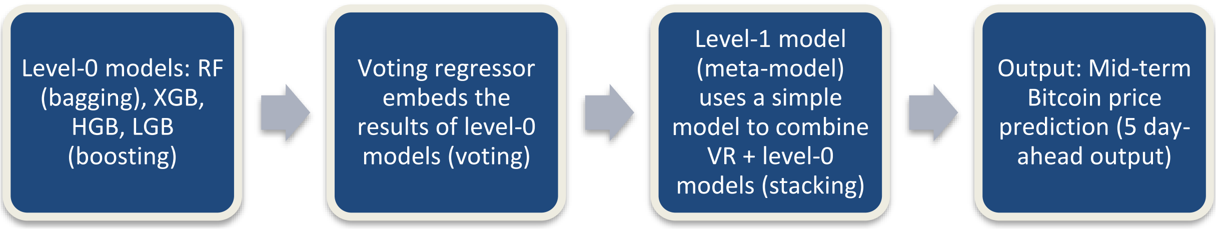


Figure 1. Proposed ensemble model

The long-term prediction is based on the Bitcoin price time series and Facebook Prophet algorithm. Thus, we take a univariate dataset and define the intervals for training and testing the results. Both an in-sample and out-of-sample predictions are performed to test the results.

4. Results

The input daily Bitcoin price for mid-term prediction is displayed in figure 2.

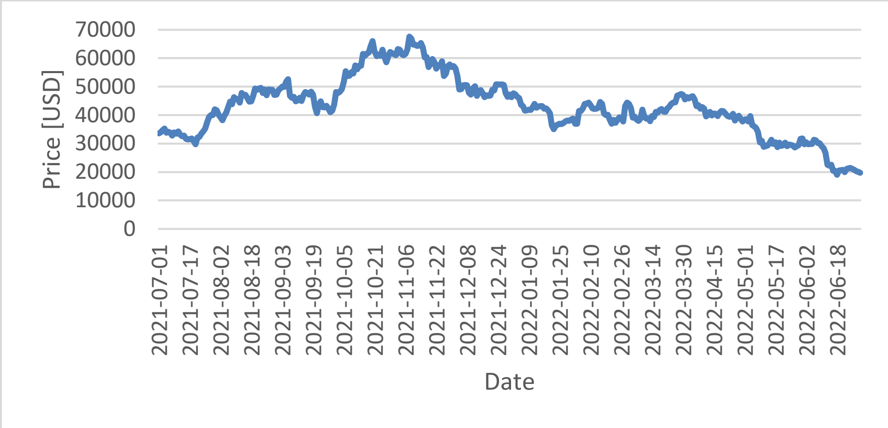


Figure 2. Bitcoin price evolution from 1st of July 2021 until end of June 2022

In Table 1, the correlation coefficients between on-chain metrics, Twitter sentiment and Bitcoin price is showcased. The correlation illustrates different intensities, and the prediction relies on the relationship between some of the metrics (such as: market cap, NUPL, SOPR) and Bitcoin prices. However, the correlation with Twitter data is weak.

Table 1. On-chain metrics, Twitter sentiment and correlations with Bitcoin price

No.	***Feature***	**Correlation**	***No.***	***Feature***	**Correlation**
1	*Active Addresses*	0.2386	*29*	*3iq Holdings*	0.5014
2	*Sending Addresses*	0.1825	*30*	*Coin Days Destroyed*	-0.0396
3	*Receiving Addresses*	0.2711	*31*	*ASOL*	-0.0209
4	*Addresses with Non-zero Balance*	-0.5994	*32*	*MSOL*	-0.0981
5	*Over 0.01*	-0.6244	*33*	*Suply Last Active 1+ Years Ago*	-0.6551
6	*Over 0.1*	-0.6012	*34*	*Suply Last Active 2+ Years Ago*	0.0977
7	*Over 1*	-0.6027	*35*	*Suply Last Active 3+ Years Ago*	-0.5328
8	*Over 10*	0.2104	*36*	*Suply Last Active 5+ Years Ago*	-0.1816
9	*Over 100*	0.5080	*37*	*MVRV*	0.9457
10	*Over 1000*	-0.3413	*38*	*NVT*	0.1313
11	*Over 10000*	-0.5159	*39*	*SOPR*	0.5137
12	*New Addresses*	0.3830	*40*	*NUPL*	0.8991
13	*Exchange Balance*	0.3545	*41*	*Block Height*	-0.4768
14	*Exchange Net Position Change*	0.2805	*42*	*Blocks Mined*	0.2005
15	*Exchange Inflow Volume*	-0.2184	*43*	*Block Interval Mean*	-0.2109
16	*Exchange Outflow Volume*	-0.2545	*44*	*Block Size Mean*	-0.0423
17	*Exchange Withdrawals*	0.1083	*45*	*Issuance*	0.2005
18	*In-House Exchange Volume*	-0.3086	*46*	*Inflation Rate*	0.2242
19	*Inter-Exchange Transfers*	0.6548	*47*	*Difficulty*	-0.3911
20	*Inter-Exchange Volume*	-0.1193	*48*	*Hash Rate*	-0.3036
21	*Fees Total*	-0.1792	*49*	*Transaction Count*	0.2938
22	*Fees Mean*	-0.2376	*50*	*Transaction Size*	0.1049
23	*Free Ratio Multiple*	0.1396	*51*	*Transfer Volume Total*	-0.0421
24	*Miner Revenue Fees*	-0.1937	*52*	*Transfer Volume Mean*	-0.0989
25	*Thermocap*	-0.4431	*53*	*UTXO Total*	-0.5736
26	*Market Cap to Thermocap Ratio*	0.9286	*54*	*Market Cap*	0.9998
27	*Purpouse Bitcoin ETF Holdings*	-0.4277	*55*	*Tweets Volume*	-0.0646
28	*Bitcoin Fund Holdings*	0.5702	*56*	*Tweets Sentiment*	-0.0190

For long-term prediction using Facebook Prophet algorithm in Python, the Bitcoin price time series is graphically represented in figure 3. Hourly prices were taken into account. In this longer interval, several bull and bear markets are evident. The first important increase took place in January and February 2021 followed by the first downward slope in May 2021, the second and third upward slopes were recorded in August and October followed by a second downward slope in November 2021.

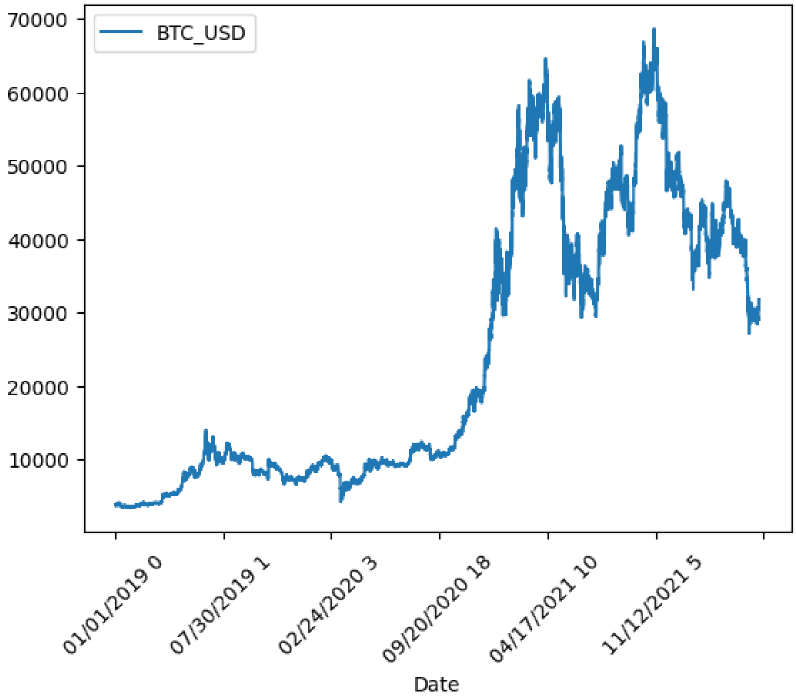
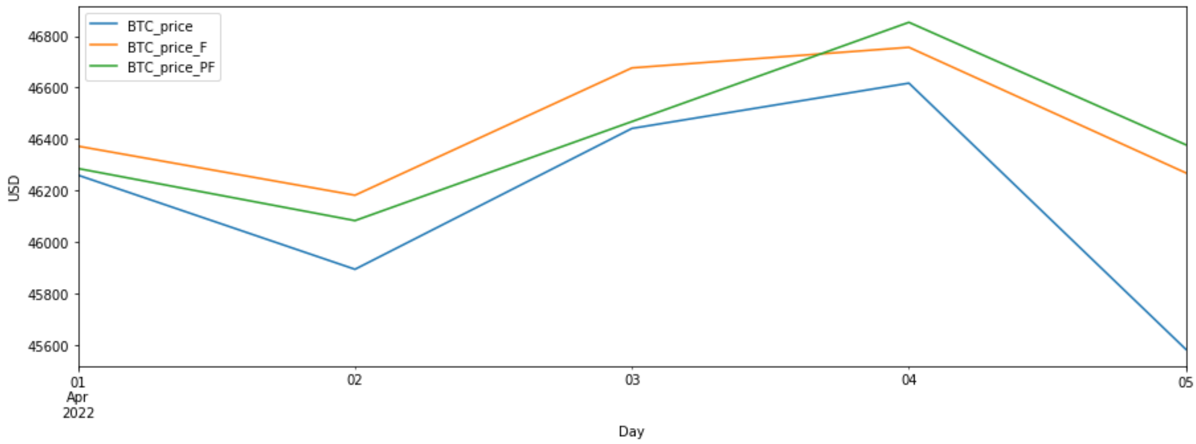
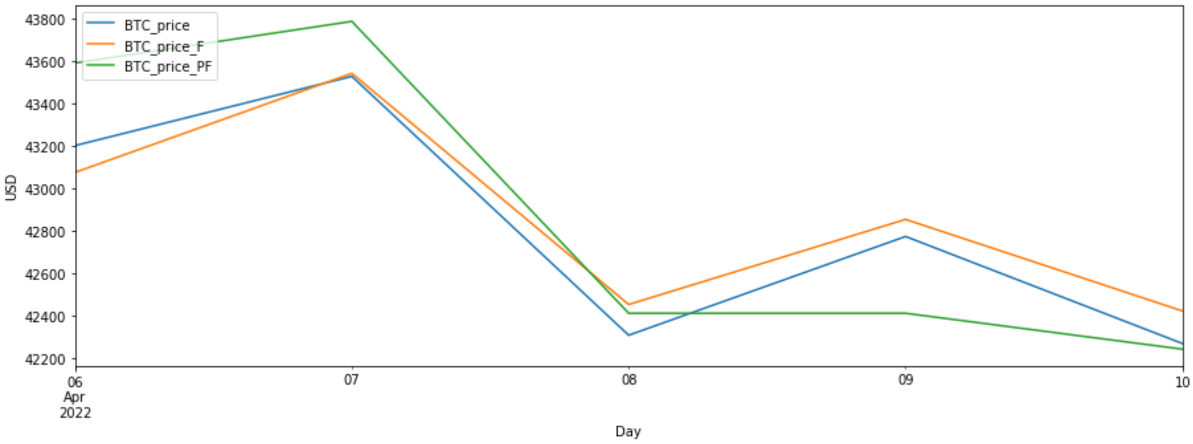


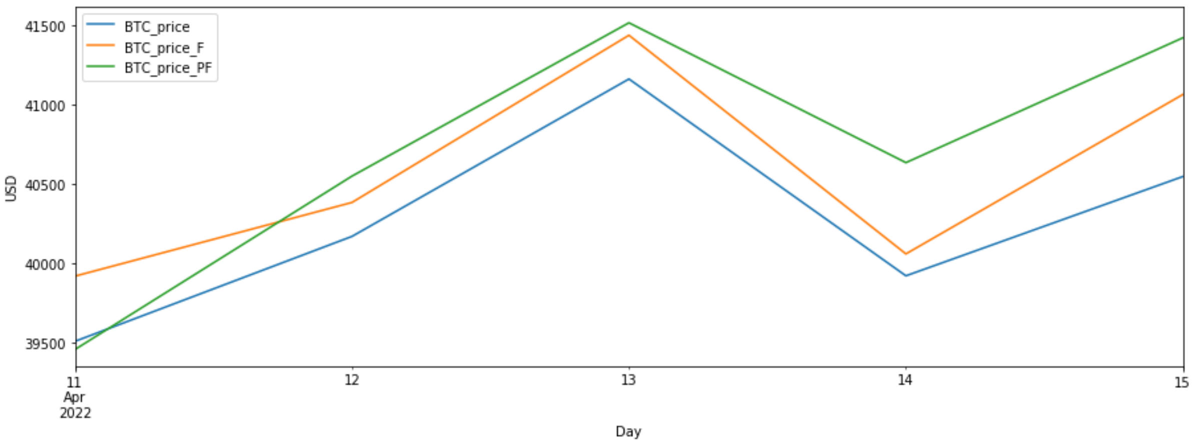
Figure 3. Bitcoin price evolution in time (from January 2019 until end of May 2022)

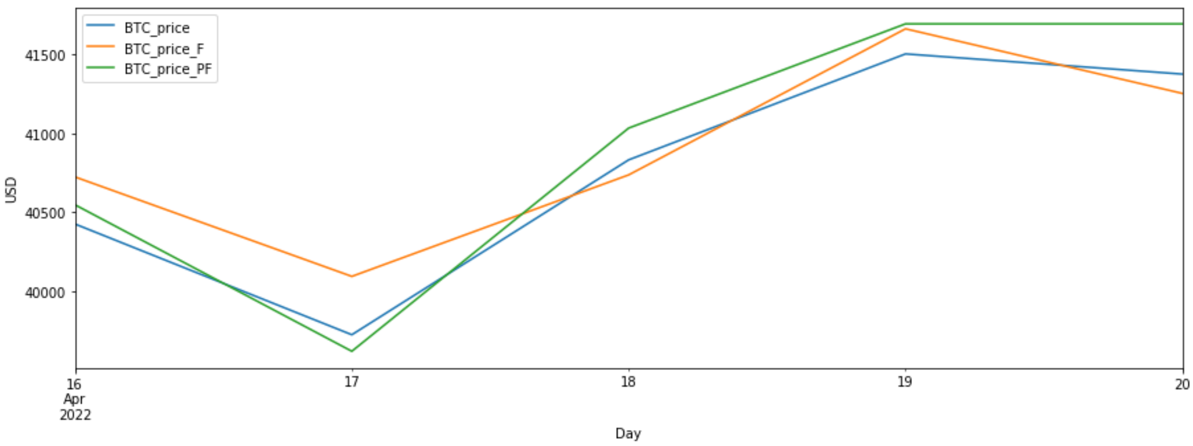
4.1. Mid-term prediction

We trained the models for interval 1st of July 2021 to 31st of March and obtained the following results for the first 5 days in April. Then we trained the models for 1st of July 2021 to 5th of April and obtained the following results for the 6th – 10th of April and so on we continued to train the models and obtain the forecast for the entire month – April 2022, as in figure 4.




BTC_price
BTC_price_F
BTC_price_PF
USD
43800
43600
43400
43200
43000
42800
42600
42400
42200
06
Apr
2022
07
08
09
10
Day


BTC_price
BTC_price_F
BTC_price_PF
USD
41500
41000
40500
40000
39500
11
Apr
2022
12
13
14
15
Day


BTC_price
BTC_price_F
BTC_price_PF
USD
41500
41000
40500
40000
16
Apr
2022
17
18
19
20
Day

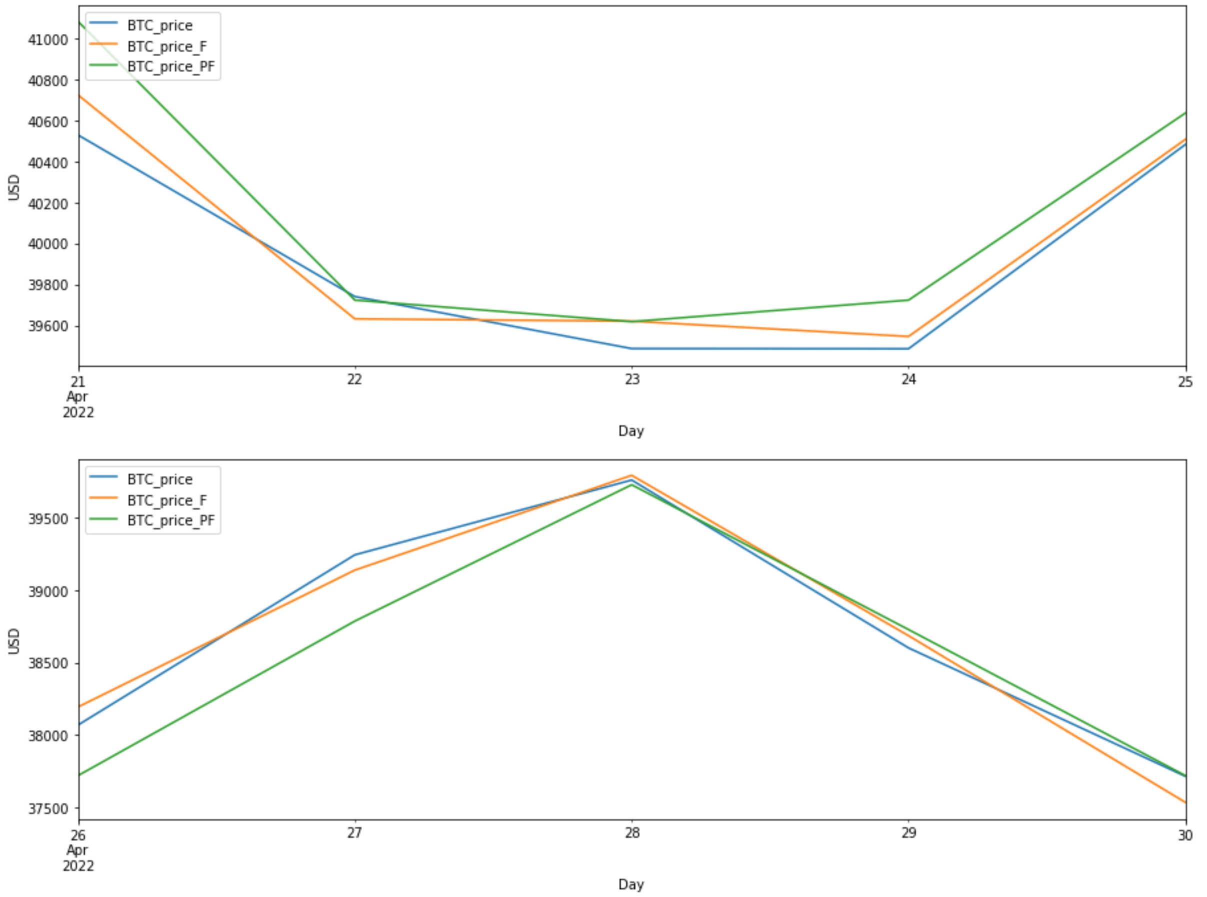


Figure 4. Bitcoin price prediction for April 2022

Figure 4 depicts the daily price movements of Bitcoin (BTC) over several days in April 2022, visualized through three different line graphs labelled "BTC_price", "BTC_price_F", and "BTC_price_PF". Here is a breakdown of what each graph represents, and the general trends observed: BTC_price - this line represents the actual historical daily closing prices of Bitcoin; BTC_price_F - this line represents a forecasted price based on a specific model or algorithm. Its closeness to "BTC_price" suggests it may be predictions made very close to or during the actual time; BTC_price_PF - this is another set of forecasted or predicted, possibly from a slightly different model or method than "BTC_price_F". The variations between "BTC_price_F" and "BTC_price_PF" indicate different modelling approaches or assumptions.

Across the charts, there are general ups and downs in the Bitcoin prices indicating typical volatility. In some periods, the forecasted lines ("BTC_price_F" and "BTC_price_PF") closely follow the actual prices, suggesting that the forecasting models are somewhat accurate during these times. There are also periods where the forecasted prices diverge significantly from the actual prices, highlighting potential limitations or errors in the forecasting models under certain market conditions. Notably, during sharp rises or drops in the actual prices, the forecasts seem to lag slightly or fail to predict the extent of the change, which is common in predictive modelling of volatile markets like cryptocurrencies.

The forecasts, especially when closely aligned with the actual prices, can be useful tools for traders and investors trying to anticipate market movements. The discrepancies between actual and forecasted prices are critical points for refining the forecasting algorithms or models to better handle sudden market shifts.

Overall, figure 4 provides useful insights into the performance and potential reliability of different Bitcoin price forecasting models. It also highlights the inherent challenges in predicting cryptocurrency prices due to their volatility.

4.2. Long-term prediction

First, we train the models and test in-sample for interval Jan. 2019 – Dec 2021. The results are showed in figure 5.

Figure 5. First in-sample long-term prediction

The out-of-sample prediction for the first 4 months in 2022 (Jan-Apr) is displayed in figure 6.

Figure 6. First out-of-sample long-term prediction

Second, we train the model and test using dataset that spans from Jan. 2019 until May 2022 as in figure 7. Out-of-sample prediction for the next 4 months (until September 2022) as in figure 8.

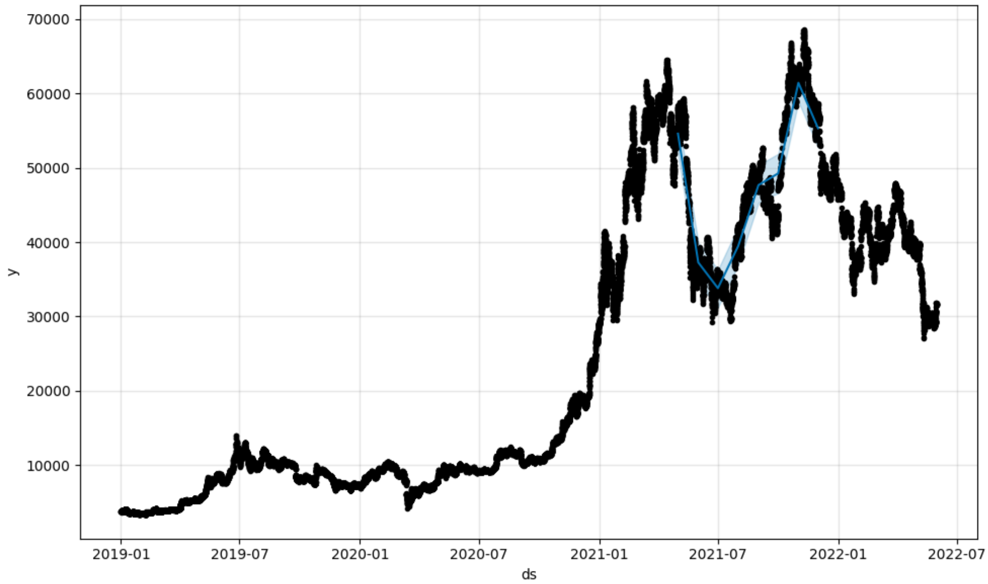


Figure 7. Second in-sample long-term prediction

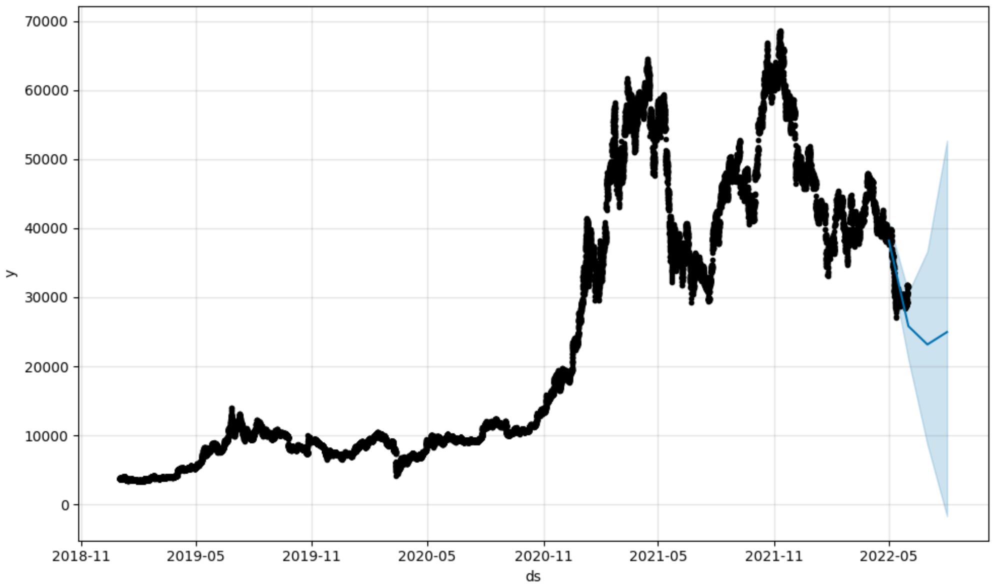


Figure 8. Second out-of-sample long-term prediction

These figures illustrate the historical price data of Bitcoin along with predictive models, showing both in-sample fits and out-of-sample forecasts. The out-of-sample forecast (Figure 8) - The black line represents historical Bitcoin prices up to a point, with a predictive model overlaid in blue. The blue shaded area indicates the confidence interval or prediction interval, suggesting where the model predicts prices could realistically fall. The model forecasts a potential continuation or slight uptick in prices, but with significant uncertainty as indicated by the wide confidence interval. The In-Sample Fit (Figure 7) - The historical data is again shown in black, with the blue line indicating the model's fit to historical data. This chart is useful for assessing how well the model captures past trends and fluctuations in Bitcoin's price. The In-Sample Fit with Trend Lines (Figure 5) - Similar to the second, but the blue line more actively follows the peaks and troughs of the actual price, suggesting a model that closely fits the historical data. The Out-of-Sample Forecast with Trend Line (Figure 6) - This displays another predictive model's out-of-sample forecast, where the blue line suggests an expected future trend based on historical data. The blue shaded area shows the uncertainty in the forecast, which appears to increase over time, indicating less certainty in the model's predictions as it moves further from the last known data point.

The models that fit the historical data closely provide insights into the model's accuracy and reliability. If the model captures the historical trends and fluctuations well, it may be more trusted for short-term forecasting. The confidence intervals in the forecasts (shaded areas) highlight the uncertainty and risk involved in using these models for predicting future prices. A wider interval suggests higher uncertainty. These models are particularly useful for investors and analysts trying to anticipate market movements or for strategizing entries and exits in trading. However, the inherent volatility of Bitcoin means that predictions can rapidly become outdated. These visualizations underscore the challenges in predicting volatile markets like cryptocurrencies.

5. Conclusion

In this paper, we proposed two approaches for mid- and long-term Bitcoin price prediction. For mid-term, 54 on-chain metrics and Twitter data for one year are used to train ensemble models and obtain a 5-day prediction, whereas for long-term prediction, Facebook Prophet algorithm is applied on the Bitcoin price time series in order to obtain the prediction for a couple of months ahead.

Based on the results, both mid and long-term predictions provide good estimations of the future prices of Bitcoin. However, it is important to note that for the two cases, different approaches are considered as the mid-term estimation rely on the on-chain metrics whereas for the long-term prediction, the Facebook Prophet provided better results.

Acknowledgements

This work was supported by a grant of the Ministry of Research, Innovation and Digitization, CNCS-UEFISCDI, project number PN-III-P4-PCE-2021-0334, within PNCDI III.

References

[1] S. Foley, B. Frijns, A. Garel, and T. Y. Roh, "Who buys Bitcoin? The cultural determinants of Bitcoin activity," *Int. Rev. Financ. Anal.*, 2022, doi: 10.1016/j.irfa.2022.102385.

[2] A. Băra, I. A. Georgescu, S. V. Oprea, and M. P. Cristescu, "Exploring the Dynamics of Brent Crude Oil, S&P500 and Bitcoin Prices Amid Economic Instability," *IEEE Access*, 2024, doi: 10.1109/ACCESS.2024.3370029.

[3] A. Bâra, S.-V. Oprea, and M. Panait, "Insights into Bitcoin and energy nexus. A Bitcoin price prediction in bull and bear markets using a complex meta model and SQL analytical functions," *Appl. Intell.*, 2024, doi: 10.1007/s10489-024-05474-2.

[4] N. Jagannath *et al.*, "A Self-Adaptive Deep Learning-Based Algorithm for Predictive Analysis of Bitcoin Price," *IEEE Access*, 2021, doi: 10.1109/ACCESS.2021.3061002.

[5] A. Bâra and S.-V. Oprea, "The Impact of Academic Publications over the Last Decade on Historical Bitcoin Prices Using Generative Models," *J. Theor. Appl. Electron. Commer. Res.*, vol. 19, no. 1, pp. 538–560, 2024, doi: 10.3390/jtaer19010029.

[6] F. Qin, Y. Wu, F. Tao, L. Liu, L. Shi, and A. J. Miller, "Multi-input address incremental clustering for the Bitcoin blockchain based on Petri net model analysis," *Digit. Commun. Networks*, 2022, doi: 10.1016/j.dcan.2022.09.003.

[7] S. Kasahara, "Performance modeling of bitcoin blockchain: Mining mechanism and transaction-confirmation process," *IEICE Trans. Commun.*, 2021, doi: 10.1587/transcom.2021ITI0003.

[8] S. Aras, "On improving GARCH volatility forecasts for Bitcoin via a meta-learning approach," *Knowledge-Based Syst.*, 2021, doi: 10.1016/j.knosys.2021.107393.

[9] H. Wang and X. Zhou, "Less is More: Bitcoin Volatility Forecast Using Feature Selection and Deep Learning Models," 2022. doi: 10.1109/INDIN51773.2022.9976100.

[10] T. Papadimitriou, P. Gogas, and A. F. Athanasiou, "Forecasting Bitcoin Spikes: A GARCH-

SVM Approach," *Forecasting*, 2022, doi: 10.3390/forecast4040041.
[11] S. A. Basher and P. Sadorsky, "Forecasting Bitcoin price direction with random forests: How important are interest rates, inflation, and market volatility?," *Mach. Learn. with Appl.*, 2022, doi: 10.1016/j.mlwa.2022.100355.
[12] S.-V. Oprea, I. A. Georgescu, and A. Bâra, "Is Bitcoin ready to be a widespread payment method? Using price volatility and setting strategies for merchants," *Electron. Commer. Res.*, 2024, doi: 10.1007/s10660-024-09812-x.
[13] O. A. Ejohwomu *et al.*, "Modelling and Forecasting Temporal PM2.5 Concentration Using Ensemble Machine Learning Methods," *Buildings*, 2022, doi: 10.3390/buildings12010046.
[14] S. K. Gunturi and D. Sarkar, "Ensemble machine learning models for the detection of energy theft," *Electr. Power Syst. Res.*, 2021, doi: 10.1016/j.epsr.2020.106904.
[15] B. Jin, S. Gao, and Z. Tao, "ARIMA and Facebook Prophet Model in Google Stock Price Prediction," *Proc. Bus. Econ. Stud.*, 2022, doi: 10.26689/pbes.v5i5.4386.
[16] M. M. Hossain, A. H. M. Faisal Anwar, N. Garg, M. Prakash, and M. Bari, "Monthly Rainfall Prediction at Catchment Level with the Facebook Prophet Model Using Observed and CMIP5 Decadal Data," *Hydrology*, 2022, doi: 10.3390/hydrology9060111.
[17] R. Gupta, A. K. Yadav, S. K. Jha, and P. K. Pathak, "Time Series Forecasting of Solar Power Generation Using Facebook Prophet and XG Boost," 2022. doi: 10.1109/DELCON54057.2022.9752916.
[18] S. Kaninde, M. Mahajan, A. Janghale, and B. Joshi, "Stock Price Prediction using Facebook Prophet," *ITM Web Conf.*, 2022, doi: 10.1051/itmconf/20224403060.
[19] R. K. Jana, I. Ghosh, and M. W. Wallin, "Taming energy and electronic waste generation in bitcoin mining: Insights from Facebook prophet and deep neural network," *Technol. Forecast. Soc. Change*, 2022, doi: 10.1016/j.techfore.2022.121584.