Forecasting Cryptocurrency Prices Time Series Using Machine Learning

Vasily Derbentsev[0000-0002-8988-2526], Natalia Datsenko,
Olga Stepanenko[0000-0003-2319-1383] and Vitaly Bezkorovainyi[0000-0002-4998-8385]

Kyiv National Economic University named after Vadym Hetman, 54/1, Peremohy Ave., Kyiv, 03057, Ukraine
derbvd@gmail.com, d_tashakneu@ukr.net,
olga_stepanenko@kneu.edu.ua, retal.vs@gmail.com

Abstract. This paper describes the construction of the short-term forecasting model of cryptocurrencies' prices using machine learning approach. The modified model of Binary Auto Regressive Tree (BART) is adapted from the standard models of regression trees and the data of the time series. BART combines the classic algorithm classification and regression trees (C&RT) and autoregressive models ARIMA. Using the BART model, we made a short-term forecast (from 5 to 30 days) for the 3 most capitalized cryptocurrencies: Bitcoin, Ethereum and Ripple. We found that the proposed approach was more accurate than the ARIMA-ARFIMA models in forecasting cryptocurrencies time series both in the periods of slow rising (falling) and in the periods of transition dynamics (change of trend).

Keywords: cryptocurrency market, short-term forecasting model, machine learning approach.

1 Introduction

The rapid development of digital currencies during the last decade is one of the most controversial and ambiguous innovations in the modern global economy.

Significant fluctuations in the exchange rate of cryptocurrencies and their high volatility, as well as the lack of legal regulation of their transactions in most countries resulted in significant risks associated with investment into crypto assets. This has led to heated discussions about their place and role in the modern economy (see, for example [1-5]).

Therefore, the issue of developing appropriate methods and models for predicting prices for cryptographic products is relevant both for the scientific community and for financial analysts, investors and traders.

Methodological approaches to forecasting prices for financial assets depend on an analyst's understanding of the causal relationships in the pricing process.

For example, the forecasting model can be specified as a price formation model:

- Based on the interaction of market players (demand-supply models) that make economic decisions based on some indicators or regularities, taking into account objective economic laws or laws of behavioral finance (econometric and balance models);
- Given the past dynamics (time series models and autoregressive models),
- Taking into account production-technological possibilities of creating the corresponding asset (in particular, for commodity markets, fundamental valuation of shares, technological opportunities for mining cryptocurrency, etc.);
- Based on the consideration of random factors and events, for example, external shocks, which complicate the formal description of cause and effect relationships (stochastic models).

It should be noted that forecasting cryptocurrencies' prices is fundamentally different from forecasting other financial assets, in particular, ordinary (fiat) currencies, which have a large number of theoretical and empirical studies focused on studying their dynamics model.

There are two fundamentally different approaches to forecasting the exchange rate dynamics of currencies. The first approach is to build a cause and effect casual model that describes the relationship between exchange rates and other macroeconomic variables (in particular, the rates of economic growth, trade and balance of payments, purchasing power parity, public debt, inflation rates, etc.) within a certain theoretical economic concept.

The other approach is to study only the time series and make a prediction based on the processing and analysis of past observations. The most common models are the Box-Jenkins ARIMA time series models and their modifications, GARCH models, or artificial neural networks.

It should be noted that there is no consensus on the fundamental value of cryptocurrencies among scholars. The prevailing thesis is that the exchange rate of the majority of cryptocurrencies is determined only by the ratio of demand and supply [3, 4, 6-10].

Liu and Tsyvinski's [11] empirical analysis of the three most capitalized crypto currencies (Bitcoin, Ripple, and Ethereum) did not reveal a static relationship between the yield of cryptocurrencies and the complexity of their extraction.

At the same time, macroeconomic factors, which usually determine the dynamics of currency, stock and commodity markets have no significant effect on the dynamics of the cryptocurrencies market.

Conrad, et al. [12], also found that influence of the US stock market (SP500 index) and the global stock market index (Nikkei 225 index) on bitcoin's volatility was not significant.

In addition, the studies reported in [1, 8, 9] show that the price dynamics of cryptocurrencies is described by classical log-periodic models of price bubbles of Sornette [13] and their modifications.

A number of recent cryptocurrency market studies show that, unlike other financial assets, cryptocurrency prices are influenced by a number of specific factors that shape their demand, such as the number of Google trends searches, the number of posts in

social networks and other mass media [6, 14-16]. These studies substantiated the feasibility of using non-typical factors as predictors.

All of these factors complicate the development of casual econometric models of cryptocurrency price dynamics.

Recently, non-parametric methods based on Machine Learning and Deep Learning have gained popularity for the analysis and forecasting of financial and economic time series.

Models of Machine Learning are based on special artificial networks that allow to solve the problem of prediction and classification by utilizing learning sequences in the data. The effectiveness of such models depends on the training speed and the degree of universality of approximating functions.

These models combine an arsenal of powerful methods, such as Artificial Neural Network (ANN), Support Vector Machines (SVM), Decision and Classification Tree (DT, CT), Fuzzy Logic, Genetic Algorithms (GA), linear and nonlinear statistical models, etc.

Examples of their effective use in forecasting exchange rates and stock indices are given, in particular, by Peng et al. [17].

Several studies [18-20] reported the results of the Bitcoin exchange rate forecasting using classical ARIMA models and using different methods of machine learning, such as Random Forest (RF), Logistic Regression (LR) and Linear Discriminant Analysis (LDA), and Long Short-Term Memory (LSTM). The results from these analyses showed that the models that relied on training proved to be better suited for forecasting both the prices of cryptocurrencies and their volatility.

Rebane and Karlsson [21] conducted a comparative analysis of the ARIMA forecasting properties with recurrent neural networks (RNNs) for such cryptocurrencies as DASH, Ethereum (ETH), Litecoin (LTC), Siacoin (SC), Stellar (STR), NEM (XEM), Monero (XMR) and Ripple (XRP). The results showed that neural networks had better forecasting properties than ARIMA models.

Thus, in our view, the second approach, which is based on the application of the time series analysis using the CRISP-DM methodology [22], is more appropriate for predicting price trends in cryptocurrency.

The purpose of our work is to construct a short-term price forecasting model for the 3 cryptocurrencies with the highest market capitalization using binary autoregressive models and machine learning technology.

2 Methodology

2.1 CRISP-DM Approach

To solve the problem of forecasting the dynamics of cryptocurrencies, we used the CRISP-DM (Cross-industry standard process for data mining) methodology (Fig. 1-2).

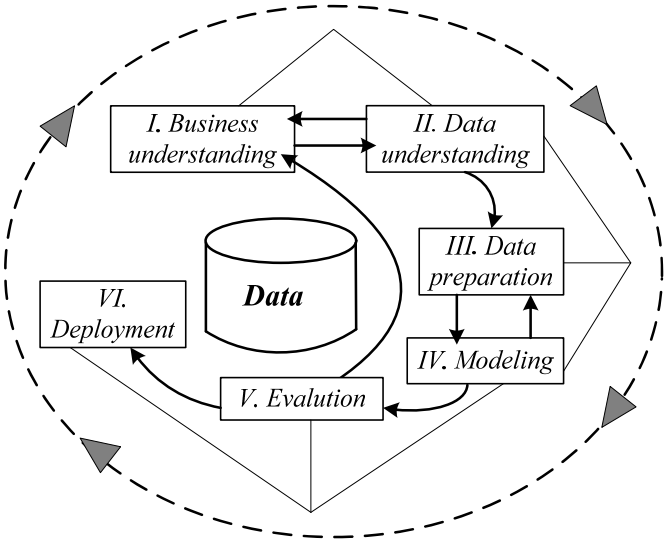


Fig. 1. The conceptual diagram of cryptocurrency forecasting based on the standard CRISP-DM.

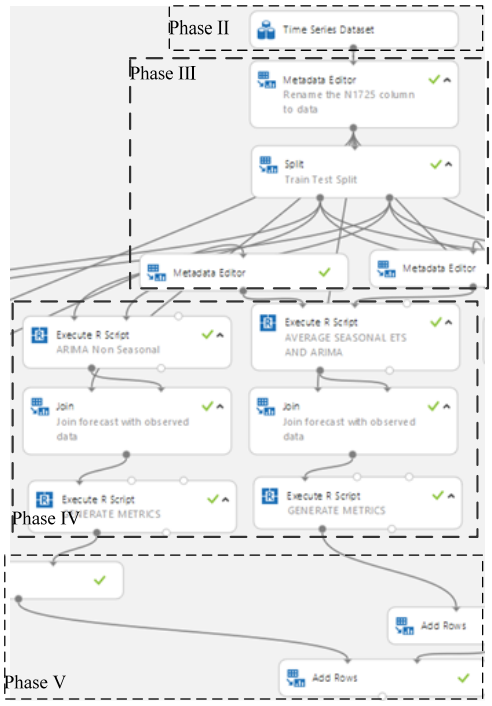


Fig. 2. An excerpt from the cryptocurrencies functional dynamics diagram in the Microsoft Azure Machine Learning Studio environment based on the CRISP-DM standard.

According to CRISP-DM, intelligent analysis is a continuous process with many cycles and feedback loops, and has six phases (I-VI).

The main advantage of the CRISP-DM is that it is platform- and application neutral and that it can be adapted to various applied problems.

Fig. 2 shows some of the CRISP-DM phases of the cryptocurrency forecasting functional dynamics diagram: Phase II: Data understanding, Phase III: Data preparation, Phase IV: Modeling, Phase V: Evaluation.

Methodology CRISP-DM is the most widespread publicly available standard process model that describes major phases and common data mining methods.

2.2 Regression Tree

The regression tree is a class of regression models that allows separating the input space of factor variables into segments. Subsequently, a separate piecewise regression model can be constructed for each of them representing a regression function in an intuitive and visual form [23-24].

In such a tree, internal nodes contain rules for splitting the space of explanatory variables; branches indicate the conditions and the transition between the nodes; and tree leaves are local regression models.

The essence of this method is in sequential division of the data set into non-intersecting classes, which, in turn, are also subject to a breakdown by a partition efficiency criterion.

The decision tree consists of the following elements: “nodes”, “leaves” and “branches”. “Branches” contain records of attributes which define the target function (result variable), the “leaves” are the values of the target function, and “nodes” are the remaining attributes under which the classification takes place.

There are two types of trees: (i) for classification, in this case, the result of the prediction is the data ownership class; and (ii) for regression, the result in this case is the predicted value of the target function.

2.3 BART Algorithm

Let us consider the proposed approach we call BART (Binary Auto Regressive Tree). It is a generalization of standard models of regression trees and is adapted to time series data. BART combines the classic classification and regression trees (C&RT) [24-25] algorithm and the standard autoregressive integrated moving average (ARIMA) models and their components (AR, MA). Models of ART (Auto Regressive Tree) are closely related to the models of the TAR (threshold autoregressive model) threshold auto-regression models of the class and their modifications SETAR and ASTAR [24]. The SETAR and ASTAR models are linear models that construct multiple adaptive regression splines (MARS) based on time series [26-27]. BART models differ from the SETAR and ASTAR models in two ways:

(1) Error estimates for models based on BART differ from one another;
(2) BART models allow for the gap between built-in auto regression models.

To convert a time series, the "window" data conversion method is used. The result variable Y_t in this algorithm corresponds to the previous value (Y_{t-1}) and the value with the lag p (Y_{t-p}). This separation of the input space into segments (Fig. 3) allows to construct a separate (local) model for each of them and to represent a piecewise function as an autoregressive tree (Fig. 4) in an intuitive visual form.

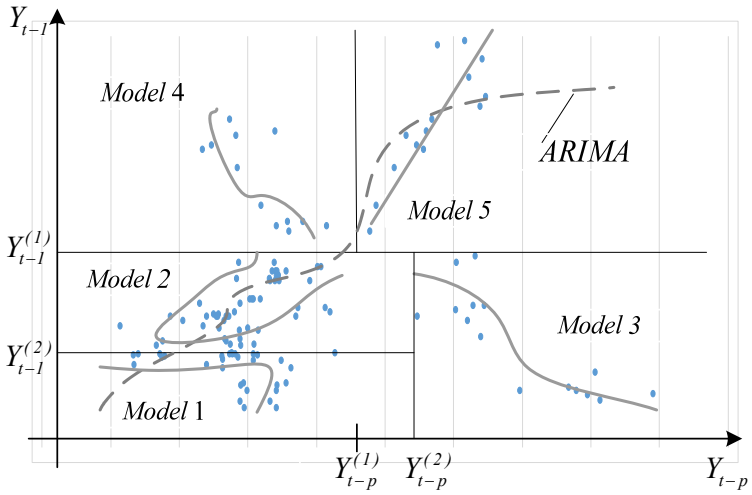


Fig. 3. Separation of the input space into segments.

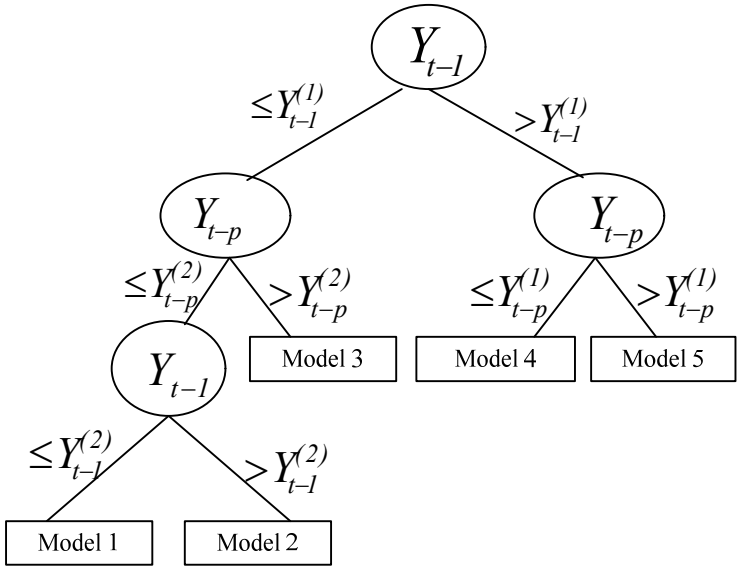


Fig. 4. Autoregressive tree building diagram.

Most such algorithms apply a recursive separation of training data. In BART, unlike other algorithms, *a step-by-step (staged and iterative) method of constructing a tree is used*:

Step 1. The construction of a regression tree begins from a single value (root node), which is defined as the Median (*Me*, second quartile $Q_{50\%}$) of the entire time series Y_t and is calculated the equation

$$ME=Q_{50\%}=0.5(Y_i^{min}+Y_i^{max}) \tag{1}$$

The median of the time series is defined as the median of the distribution of realization of a random variable at time *t*, that is, a real number with probability of exceeding an arbitrary dimension equal to 0.5. For a stationary series and a series with a symmetric distribution, this value does not depend on the time of observation $Me=\overline{Y}_t$ and coincides with the mean value of the series. Sometimes in the literature, the median is considered to be a prototype of a simple stable output.

Step 2. The best split is found for each unprocessed node, and it is selected according to a predefined rule.

These procedures are performed similarly to the C&RT algorithm. The difference lies in the accepted rules, criteria for evaluation and termination of splitting. We have used an alternative selection criterion (or informational criterion) for better splitting based on the entropy indicator, because it gives preference to options with less tree complexity. This algorithm will determine an entropy information gain.

In constructing BART, the number of branches (branching) is 2, that is, each node has two child nodes. The final tree is chosen from these nodes, and we have to evaluate informativeness of not only the predictor nodes that divide the time series into subsets, but also of those that separate a certain group of subsets from the set, that is, the subtree from the rest of the tree.

Entropy criterion. Initially, the probability is estimated as the frequency of assigning a particular observation to a certain subset (subtree) and the entropy $\hat{H}$ sampling Y^l is calculated using the following equation:

$$\hat{H}\left(P,N\right)=H\left(\frac{P}{P+N}+\frac{N}{P+N}\right). \tag{2}$$

After all the information in the node is obtained for a certain predecessor, entropy is calculated using the following equation:

$$\hat{H}_{\phi}\left(P,N,p,n\right)=\frac{p+n}{P+N}\hat{H}\left(p,n\right)+\frac{P+N-(p+n)}{P+N}\hat{H}\left(P-p,N-n\right), \tag{3}$$

where P – is the number of objects that correspond to a subset *C*, and p – is the number of objects that correspond to the membership conditions of a subset, $p\in P$, similarly n and N are such that $n\in N,\ N\in C$.

Then the entropy of the sample $\left\{ x \in Y^l \,\middle|\, \phi(x) = 1 \right\}$ will be $\hat{H}(p,n)$, and the probability of obtaining an element from this sample will be calculated as $\frac{p+n}{P+N}$. Similarly, for the sample $\left\{ x \in Y^l \,\middle|\, \phi(x) = 0 \right\}$ entropy $\hat{H}(P-p, N-n)$ can be calculated with probability $\frac{(P-p)+(N-n)}{P+N}$. Thus, the entropy of the whole sample after obtaining information φ is calculated using equation (3).

Then the decrease of entropy can be calculated as:

$$\mathrm{IGain}_{\mathrm{C}}\left(\phi, Y^l\right) = \hat{H}(P,N) - \hat{H}(P,N,p,n) \tag{4}$$

which is called *entropy information gain*, which is the amount of information about the current division of the tree into two classes *«c»* and *«not c»*.

In addition, in the BART algorithm for the early termination criterion Q, we used an extended Bayesian information criterion [28], which minimizes the statistic:

$$\mathrm{EBIC} = n \cdot \ln \frac{\mathrm{SSE}}{n} + J \cdot \left[\ln(n) + 2\ln(p) \right], \tag{5}$$

where SSE – is the sum of squares of the residuals of the model; J – is the number of model parameters; n – is the number of examples of training sample; p – is the quantity that characterizes the complexity of the model space (it is the product of the tree size and the number of explanatory variables).

In equation (5), the first term is the maximum value of the plausibility logarithmic function, and the second is a penalty for the model complexity.

Splitting of the nodes continues until the EBIC value is reduced. Note that the application of this criterion in the recursive approach of the algorithm of the regression tree is not possible. This is due to the fact that in the recursive method during tree construction only part of the model is considered at a time without considering the complete model as a whole.

For BART, the simplification procedure (i.e., early termination of the tree branching) is more important than, for example, for classification trees. This is due to the fact that regression trees tend to be more complex, because the variety of the investigated metric values (for example, the price of regression) is much more diverse than for qualitative data.

Step 3. If the selected split improves the model and it is valid with an entropy information gain, then this split is performed and step 2 is repeated. Otherwise, the final tree is selected and the BART algorithm execution procedure is considered complete.

The rejection of recursion in the BART algorithm and the transition to the iterative version allows for a complete control of the tree construction process, that is, it provides a “softer” control of the tree construction process at the expense of the following:

(1) Determining the arbitrary order of split nodes;

(2) Introducing early termination rules / algorithms that analyze both separate nodes and the whole regression tree as a whole;
(3) Termination of the construction of the regression tree at any time.

Because the ultimate goal of the proposed algorithm is forecasting, the standard regression model of the ARIMA class, which is a traditional tool for forecasting financial series, needs to be built on the nodes-leaves:

$$\Phi(L)(1-L)^d X_t = \mu + \Theta(L)\varepsilon_t,\ \varepsilon_t \sim N(0,\sigma^2) \tag{6}$$

where Y_t – is the time series, L – is the lag operator, $\Phi(L)$ – is the polynomial degree p from L, μ – is the average process value, $\Theta(L)$ – is the polynomial degree q from L, ε_t – is white noise, d – is the order of process integration Y_t. If d=0, then process X_t can be described by ARMA (p, q) or ARIMA (p, 0, q).

This process is stationary and has a short memory. If d=1, then the series has infinite memory, that is, each perturbation has an impact on the behavior of the process indefinitely.

Thus the result variable Y_t in this algorithm corresponds to the previous value (Y_{t-1}) and the lag p (Y_{t-p}). Also, the separation of the input space into segments allows to construct an own (local) model for each of them and to represent a piecewise function as an autoregressive tree in an intuitive visual form.

3 Empirical Results

For performing empirical analysis, we selected three cryptocurrencies which are the market capitalization leaders: Bitcoin (BTC), Ethereum (ETH) and Ripple (XRP). We have taken daily closing prices for the period from 01/01/2017 to 01/03/2019, according to Yahoo Finance [29] and calculated their time series in log-return.

To compare the predictive properties of the BART algorithm, we also made a forecast using the classical ARIMA (1, 0, 1) and ARFIMA (1, d, 1) models.

As a parameter d in ARFIMA we can use appropriate Hurst exponents (see, for example E. Peters [30]). So we selected as the difference parameter d for ARFIMA models for each currency such values [31]:

$H(BTC) = 0.75,\ \ H(ETC) = 0.83,\ \ H(XRP) = 0.66$.

The sample size for training for all sub-periods for the BART algorithm was 80% of the total sample size, and 20% was used as out-of-sample dataset.

To implement the models, we chose the Microsoft Azure Machine Learning Studio Cloud Application. A fragment of the implementation of machine experiments is shown in Fig. 5.

For each model the target variable is the log-return for the next time period. The forecast was carried out on five different time horizons: 5, 10, 14, 21, and 30 days using three models for each cryptocurrency.

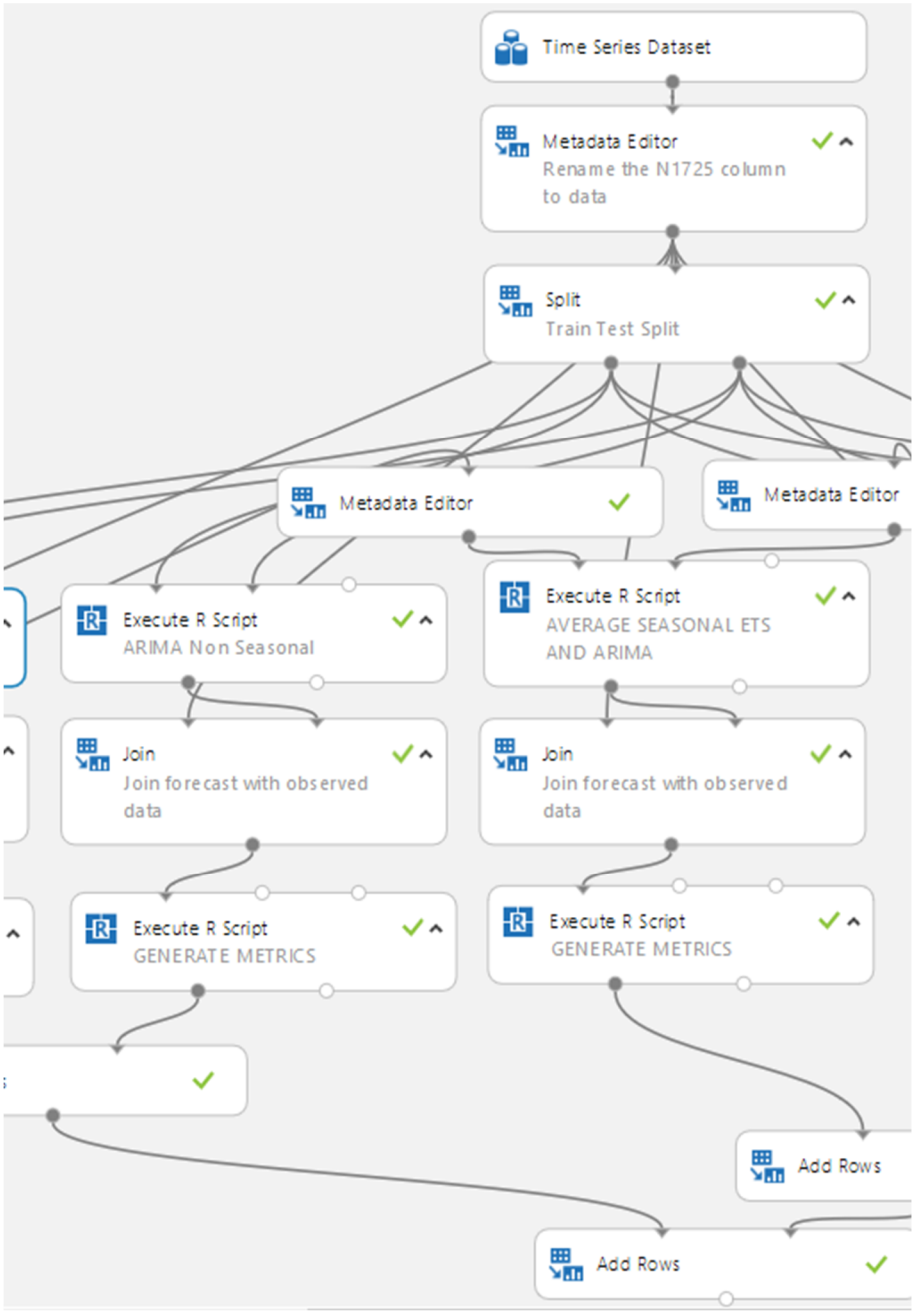
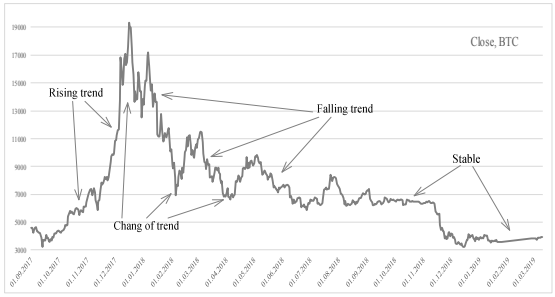


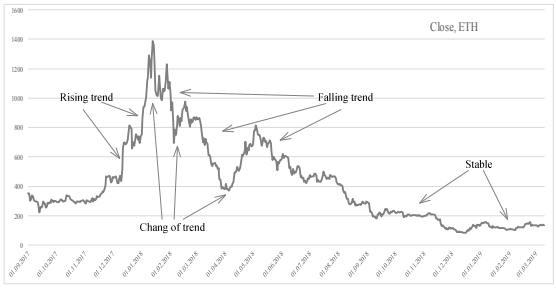
Fig. 5. A diagram of the forecasting cryptocurrencies times series experiments in MS Azure Machine Learning environment.

To check the effectiveness of the BART algorithm and that of the classical models, we conducted tests for periods with different types of dynamics of cryptocurrencies time series (two subperiods for each type), namely (Fig. 6):

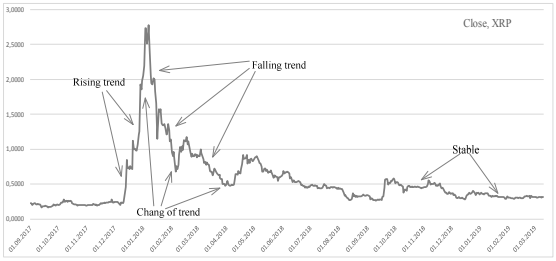
(1) Stable period;
(2) Falling trend;
(3) Transition dynamics (change of trend);
(4) Rising trend.



(a)



(b)



(c)

Fig. 6. Selected periods with different types of dynamics of cryptocurrencies time series: (a) BTC, (b) ETH, (c) XRP.

As we can see, BTC is a driver and other cryptocurrencies repeat its dynamics.

Fig. 7-8 illustrate the forecast accuracy for 3 models for ETH in the period of slow rising (falling) (Fig. 7) and rapid trend change period (Fig. 8). Forecasting accuracy for BTC and XRP have the same properties as ETH.

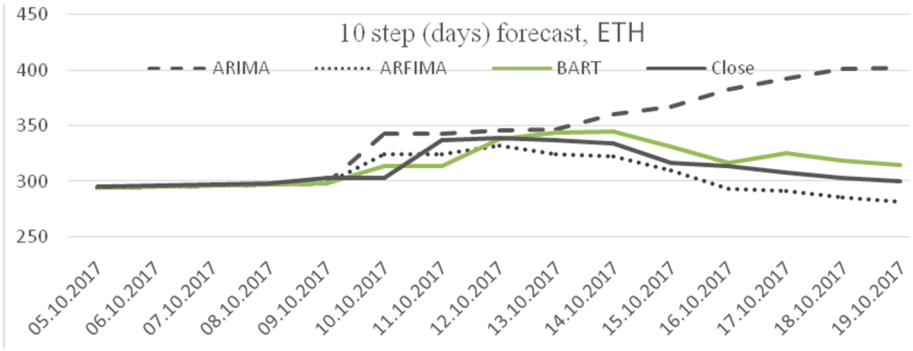


Fig. 7. 10-step (days) forecast performance for ETH in the period of slow rising (falling).

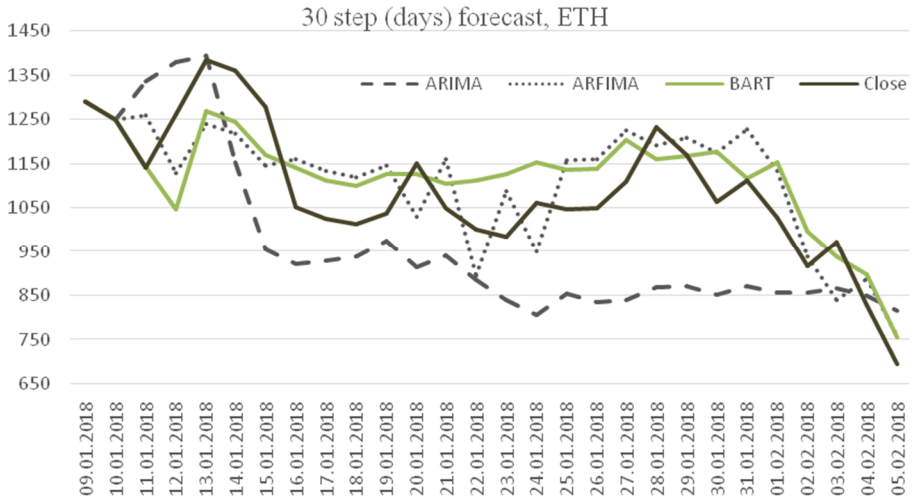


Fig. 8. 30-step (days) forecast performance for ETH in transition dymamics period.

To estimate the prognostic properties of the models we used the Root Mean Square Error metric (RMSE).

Results (averaged over three cryptocurrencies) of forecasting performance for all sub-periods are shown in Table 1.

The obtained results indicate that for the investigated time series of cryptocurrencies, the proposed approach gives RMSE over the range 4% for the 14 days forecast horizon without reference to the type of dynamic behavior, over the range 6% for the 21 days and 8% for the 30 days forecast horizon, respectively.

Table 1. Summary (average in three cryptocurrencies) of the forecast accuracy RMSE, %

Model	Step forecast, days				
	5	10	14	21	30
Stable period *					
ARIMA	4.27	6.54	6.80	12.20	15.08
ARFIMA	3.93	4.87	5.30	6.00	8.60
BART	2.76	3.11	3.45	3.73	4.83
Falling trend **					
ARIMA	7.13	10.12	11.10	12.47	16.40
ARFIMA	4.98	5.34	7.70	8.73	9.26
BART	2.97	3.64	3.30	4.15	5.60
Transition dynamics (change of trend) ***					
ARIMA	6.76	7.03	7.30	13.61	18.51
ARFIMA	3.34	3.67	4.00	6.44	8.76
BART	2.82	3.16	3.50	5.59	7.61
Rising trend ****					
ARIMA	6.82	11.99	13.15	14.07	15.34
ARFIMA	4.63	4.42	7.20	7.81	8.93
BART	2.25	2.98	3.70	3.34	5.64

* sub-periods: 11/09/2018-09/11/2018, 27/12/2018-25/02/2019
** sub-periods: 06/01/2018-09/02/2018, 12/03/2018-10/04/2018
*** sub-periods: 04/12/2017-02/01/2018, 22/04/2018-21/05/2018, 02/03/2018-04/05/2018
**** sub-periods: 09/10/2017-19/12/2017, 10/02/2018-11/03/2018

The results show that for selected time series for the short-term forecast, the error of BART algorithm is half the size of the error of ARIMA model, on average, and it is 15-20% lower than the error of ARFIMA model for slowly changing periods (both falling and rising).

Note that all of our models show worse forecast accuracy for the periods of complex dynamic modes (rapid trend change periods).

In addition, the proposed algorithm is more accurate in the periods of transition dynamics (change of trend) compared to ARIMA-ARFIMA models.

4 Concluding Remarks

The modified model of Binary Auto Regressive Tree (BART) is adapted from the standard models of regression trees to the data of time series. BART combines the classic algorithm C&RT and autoregressive models ARIMA.

One of the advantages of the proposed method is the use of the “window” data transformation method for the time series.

The obtained results proved that the BART algorithm is more accurate for all investigated time series of cryptocurrencies and subperiods. In particular, RMSE for this algorithm for the horizon of 14, 21, and 30 days was within the range of 4%, 6%, and 8%, respectively.

The proposed BART method for analyzing and forecasting cryptocurrecies time series demonstrated higher efficiency for building forecast estimates in comparison

with traditional time series technique, regardless of whether the target data is collected before, during or after a recession.

References

1. Baek, C., Elbeck, M.: Bitcoins as an investment or speculative vehicle? A first look. Appl. Econ. Lett. **22**(1), 30–34 (2015)
2. Bariviera, A., Basgall, M., Hasperue, W., Naiouf, M.: Some stylized facts of the Bitcoin market. Physica A. **484**(3), 82–90 (2017)
3. CNBC: Top Economists Stiglitz, Roubini and Rogoff Renew Bitcoin Doom Scenarios. https://www.cnbc.com/2018/07/09/nobel-prize-winning-economist-joseph-stiglitz-criticizes-bitcoin.html (2018). Accessed 15 Feb 2019
4. Selmi, R., Tiwari, A., Hammoudeh, S.: Efficiency or speculation? A dynamic analysis of the Bitcoin market. Econ. Bull. **38**(4), 2037-2046 (2018)
5. Stepanenko, O.: Problems and Prospects of Financial Provision Modeling of Ecological, Economics and Production Systems in the Digital Economy. In: Proceedings of the 7th International youth conference on Perspectives of science and education, pp. 421–430 (2019)
6. Bouoiyour, J., Selmi, R.: What does Bitcoin look like? Ann. Econ. Financ. **16**(2), 449–492 (2015)
7. Cheah, E., Fry, J.: Speculative bubbles in Bitcoin markets? An empirical investigation into the fundamental value of bitcoin. Econ. Lett. **130**, 32–36 (2015)
8. Ciaian, P., Rajcaniova, M., Kancs, D.: The economics of BitCoin price formation. Appl. Econ. **48**(19), 1799–1815 (2016)
9. Blau, B.: Price dynamics and speculative trading in bitcoin. Res. Int. Bus. Financ. **43**, 15–21 (2018)
10. Balcilar, M., Bouri, E., Gupta, R., Roubaud, D.: Can volume predict bitcoin returns and volatility? A quantiles-based approach. Econ. Model. **64**, 74–81 (2017)
11. Liu, Y., Tsyvinski, A.: Risks and Returns of Cryptocurrency. Technical report. National Bureau of Economic Research https://www.nber.org/papers/w24877 (2018). Accessed 15 Feb 2019
12. Conrad, C., Custovic, A., Ghysels, E.: Long- and Short-Term Cryptocurrency Volatility Components: A GARCH-MIDAS Analysis. J. Risk Financ. Manag. **11**(2), 23 (2018). doi:10.3390/jrfm11020023
13. Sornette, D.: Critical Phenomena in Natural Sciences, Chaos, Fractals, Self-organization and Disorder: Concepts and Tools. Springer, Heidelberg (2004)
14. Jain, R., Nguyen, R., Tang, L., Miller, T.: Bitcoin Price Forecasting using Web Search and Social Media Data. Oklahoma State University. Paper 3601-2018. https://www.sas.com/content/dam/SAS/support/en/sas-global-forum-proceedings/2018/3601-2018.pdf (2018). Accessed 15 Feb 2019
15. Kim, Y., Kim, J., Kim, W., Im, J.: Predicting Fluctuations in Cryptocurrency Transactions Based on User Comments and Replies. PLoS ONE. **17**, 1–17 (2016). doi: 10.1371/journal.pone.0161197
16. Galeshchuk, S., Vasylchyshyn, O., Krysovatyy, A.: Bitcoin Response to Twitter Sentiments. CEUR Workshop Proceedings. **2104**, 160–168. http://www.ceur-ws.org/Vol-2104/paper_199.pdf (2017). Accessed 15 Feb 2019

17. Peng, Y., Albuquerque, P.H.M., de Sá, J.M.C., Padula, A.J.A., Montenegro, M.R.: The best of two worlds: Forecasting high frequency volatility for cryptocurrencies and traditional currencies with Support Vector Regression. Expert Syst. Appl. **97**, 177–192 (2018)
18. Amjad, M., Shah, D.: Trading Bitcoin and Online Time Series Prediction. In: NIPS 2016 Time Series Workshop. http://proceedings.mlr.press/v55/amjad16.pdf (2016). Accessed 15 Feb 2019
19. Saxena, A., Sukumar, T.: Predicting bitcoin price using LSTM and compare its predictability with ARIMA model. Int. J. Pure Appl. Math. **119**(17), 2591–2600 (2018)
20. McNally, S.: Predicting the price of Bitcoin using Machine Learning. Doctoral dissertation, National College of Ireland (2016)
21. Rebane, J., Karlsson, I.: Seq2Seq RNNs and ARIMA models for Cryptocurrency Prediction: A Comparative Study. In: SIGKDD Fintech'18, August 2018, London, UK. https://fintech.kdd2018.a.intuit.com/papers/DSF2018_paper_papapetrou.pdf (2018). Accessed 15 Feb 2019
22. Kurgan, L., Musilek, P.: A survey of Knowledge Discovery and Data Mining process models. Knowl. Eng. Rev. **21**(1), 1–24 (2006)
23. Breiman, L., Friedman, J., Stone, C., Olshen, R.: Classification and Regression Trees. Chapman and Hall/CRC, Boca Ration (1984)
24. Breiman, L.: Random forests. Mach. Learn. **45**(1), 5–32 (2001)
25. Tong, H.: Threshold models in Nonlinear Time Series Analysis. Springer-Verlag, NewYork (1983)
26. Friedman, J.H.: Greedy function approximation: A gradient boosting machine. Ann. Stat. **29**, 1189–1232 (2001)
27. Chipman, H., George, E., McCulloch, R.: Bayesian CART model search. J. Am. Stat. Assoc. **93**, 935–948 (1998)
28. Chen, J., Chen, Z.: Extended Bayesian information criteria for model selection with large model spaces. Biometrika. **95**(3), 759–771 (2008)
29. Yahoo Finance. https://finance.yahoo.com (2019). Accessed 2 Mar 2019
30. Peters, E.E.: Fractal Market Analysis: Applying Chaos Theory to Investment and Economics. John Wiley and Sons, Hoboken (1994)
31. Derbentsev, V., Kibalnyk, L., Radzihovska, Yu.: Modeling Multifractal Properties of Cryptocurrency Market using Hurst Exponent and Detrended Fluctuation Analysis. J. Environ. Econ. Manag. Special Issue (2019, in press)
32. Derbentsev, V., Datsenko, N., Stepanenko, O., Bezkorovainyi, V.: Forecasting cryptocurrency prices time series using machine learning approach. SHS Web of Conferences. **65**, 02001 (2019). doi:10.1051/shsconf/20196502001