Applied Soft Computing Journal 75 (2019) 596–606

Contents lists available at ScienceDirect

Applied Soft Computing Journal

journal homepage: www.elsevier.com/locate/asoc

Predicting the direction, maximum, minimum and closing prices of daily Bitcoin exchange rate using machine learning techniques

Dennys C.A. Mallqui [a], Ricardo A.S. Fernandes [b,*]

[a] Graduate Program in Computer Science, Federal University of São Carlos, Brazil
[b] Department of Electrical Engineering, Federal University of São Carlos, Rod. Washington Luís (SP-310) km 235, São Carlos 13565-905 SP, Brazil

HIGHLIGHTS

- We predict the Bitcoin price direction and forecast the Bitcoin exchange rates (maximum, minimum and closing prices), considering daily data.
- The proposed methodology is based on the analysis of attribute selection methods combined with machine learning algorithms to predict the Bitcoin exchange rates.
- Machine Learning Ensemble algorithms are proposed, where the combination of Recurrent Neural Networks and a Tree classifier obtained the best results to predict the Bitcoin price direction.
- The SVM algorithm obtained the best results to forecast the Bitcoin exchange rates.
- The proposed methodology provides very good results when compared to the state-of-the-art studies.

ARTICLE INFO

Article history:
Received 24 January 2018
Received in revised form 29 August 2018
Accepted 20 November 2018
Available online 1 December 2018

Keywords:
Bitcoin
Prediction
Price direction
Maximum price
Minimum price
Closing price
Regression
Attribute selection
Machine learning

ABSTRACT

Bitcoin is the most accepted cryptocurrency in the world, which makes it attractive for investors and traders. However, the challenge in predicting the Bitcoin exchange rate is its high volatility. Therefore, the prediction of its behavior is of great importance for financial markets. In this way, recent studies have been carried out on what internal and/or external Bitcoin information is relevant to its prediction. The increased use of machine learning techniques to predict time series and the acceptance of cryptocurrencies as financial instruments motivated the present study to seek more accurate predictions for the Bitcoin exchange rate. In this way, in a first stage of the proposed methodology, different feature selection techniques were evaluated in order to obtain the most relevant attributes for the predictions. In the sequence, it was analyzed the behavior of Artificial Neural Networks (ANN), Support Vector Machines (SVM) and Ensemble algorithms (based on Recurrent Neural Networks and the *k*-Means clustering method) for price direction predictions. Likewise, the ANN and SVM were employed for regression of the maximum, minimum and closing prices of the Bitcoin. Moreover, the regression results were also used as inputs to try to improve the price direction predictions. The results showed that the selected attributes and the best machine learning model achieved an improvement of more than 10%, in accuracy, for the price direction predictions, with respect to the state-of-the-art papers, using the same period of information. In relation to the maximum, minimum and closing Bitcoin prices regressions, it was possible to obtain Mean Absolute Percentage Errors between 1% and 2%. Based on these results, it was possible to demonstrate the efficacy of the proposed methodology when compared to other studies.

1. Introduction

Virtual Currencies are becoming popular and used for financial transactions worldwide. The Cryptocurrencies are the most representative [1,2] because have received much attention by the media and investors, which can be attributed to their innovative characteristics, transparency, simplicity and increasing acceptance [3]. Currently, Bitcoin [4] is probably the most successful cryptocurrency. According to the website https://coinmarketcap.com, accessed on March 3rd, 2018, the cryptocurrency market capitalization value represents approximately USD 441 billion of dollars, where the Bitcoin represents more than 42%. In addition, the Bitcoin does not need a central bank or institution to emit and control it. Its transactions are evaluated and stored in the network itself, in a peer-to-peer format [5], using cryptography models to validate transactions [6]. Moreover, it is very attractive

* Corresponding author.
E-mail addresses: dennys.mallqui@xertica.com (D.C.A. Mallqui), ricardo.asf@ufscar.br (R.A.S. Fernandes).

https://doi.org/10.1016/j.asoc.2018.11.038

for payments because the cryptocurrencies are anonymous, faster and simpler than using traditional credit cards [7].

Due to its innovative characteristics, the Bitcoin has increased its acceptance among the public. Thus, in accordance with [8], in February 2015, more than 100,000 companies accept Bitcoins. The list includes famous companies like Amazon, CVS, Dell, Expedia, Home Depot, Pay Pal, Subway, Target, Victoria Secret, Gap, among others. Furthermore, the list continues to grow, such as presented by [9]. Currently, the authors of [10] show another list with retail companies as Overstock, eGifter, Newegg, Microsoft (funds for purchase movies, games and apps) and Shopify stores.

Although there are criticisms regarding security aspects of anonymity for Bitcoin transactions, recent studies such as those made by [11] and [12] suggest that there are opportunities to improve these aspects. However, it would be needed the adaptation of the current architecture of the Bitcoin to support the evolution of its demand and advances in the cryptographic and data security research fields.

For traders or general users, the greatest challenge is the Bitcoin exchange rate volatility. Therefore, idealize a model that can explain the Bitcoin price behavior for this unsettled market is meaningful [13]. As mentioned by [14], the author claims that the high volatility of the Bitcoin cannot be a factor that makes it a currency, but it is a motivation for traders and the general public to seek solutions to reduce their risk. Thus, in the financial world, the possibility to predict the price direction of assets is a practical matter that strongly influences a trader decision to buy or sell an instrument of investment. As mentioned by [15], a trade driven by an accurate price direction prediction can generate better returns for investors.

The number of studies about the time series of Bitcoin exchange rate is increasing, but is relatively recent. Many of them try to identify factors or attributes that show more correlation with the Bitcoin price variation [16–23]. Other previous studies try to make predictions for the Bitcoin exchange rate behavior [24–26].

Based on the above context, this research seeks to contribute to the decision support literature, identifying relevant attributes and machine learning algorithms to make predictions of the Bitcoin exchange rate (Bitcoin against US dollar), in order to obtain greater accuracy than recent studies and, consequently:

1. identify techniques of feature selection that can obtain the most relevant attributes;
2. analyze the best attributes that explain the behavior of the daily variation of exchange rate;
3. get the configurations of machine learning algorithms that obtain the best results.

The present study is focused on improving the accuracy of forecasts of the daily exchange rate behavior of the Bitcoin, considering the direction, maximum, minimum, and closing prices. Thus, the goal of this paper is to propose a methodology that can improve the decision making process for Bitcoin traders. For this purpose, it was employed feature selection algorithms to determine the most relevant attributes and analyzed the performances of machine learning algorithms to predict the Bitcoin exchange rate.

This paper is organized as follows: Section 2 presents the foundations of the Bitcoin and a background review; Section 3 introduces the proposed methodology; Section 4 presents the numerical results; in Section 5 are analyzed and discussed the results; and, finally, Section 6 presents the conclusions, highlighting the main contributions of the paper.

2. Foundations of Bitcoin and background review

Bitcoin works using open source peer-to-peer system that was created by an organization or person under the pseudonym of "Satoshi Nakamoto" in 2008 [7]. This cryptocurrency or electronic coin is defined as a chain of digital signatures, where each owner transfers the coins to the other using a digital signature, which contain a hash of the previous transactions and the public key of the next owner, adding this information to the end of the coin. This signature can be verified by the holder to prove the chain of ownership [6]. Bitcoin network is composed by a high number of computers connected through the Internet. In order to avoid the need of a trusted party to validate transactions, it was implemented a proof-of-work mechanism where the nodes perform complex mathematical procedures to verify the correctness and truthfulness of the transactions [7]. When the proof-of-work is finished, one block is added to the chain (Blockchain). The block cannot be changed without redoing the work. As an incentive, the first transaction in a block is a special transaction that creates new coins owned by the responsible for generating the block. This represents a similar scheme to the gold miners [6]. In addition, the system provides a limited total amount of money in circulation, equal to 21 million of Bitcoins. Consequently, this action avoids the risk of increasing the number of coins and generating inflation [7].

Due to the innovative characteristics of Bitcoin, many companies have exploited it as a way of making payments for interchange of services and products. Thus, in [5] the author highlights the technological advantages of Bitcoin, converting it into an alternative to credit cards and traditional bank transfers. Unlike the previous study, which was focused on the advantages of Bitcoin as a currency, in [27] the authors present solid indications that new users are not primarily interested in its transaction advantages, but to participate in a new investment vehicle. Similar to the previous study, in [28] and [20], the behavior of Bitcoin is analyzed based on their capabilities as an instrument of investment.

Relating to the identification of relevant attributes for Bitcoin price and trend forecasting, in [16] the author claims that the searches on Google Trends and Wikipedia contain relevant correlation with Bitcoin price. The same author, in another paper [17], analyzes long-term and short-term correlations with different sorts of factors, finding a significant relation to fundamental economic factors and periods where the price of Bitcoin rises (possible cause of price bubbles). In [20], the authors suggested that the determinant factors of Bitcoin exchange rate are classified in technical (hash rate and public recognition) and economics (economic fundamentals and trading volume). Specifically, in long-term models, the exchange rate shows a significant reaction to economic fundamentals (including money supply, gross domestic product, inflation, and interest rate) and, in the short-term, it responds promptly to changes in hash rate and public recognition (Google searches and Twitter mentions) [20]. In other research, [18], is analyzed whether social media activity or information extracted from web search media could be helpful to predict the behavior of Bitcoin price. As a result, Google Trends could be seen as a sort of predictors, because of its high cross-correlation. In a similar way, the authors of [19] conclude that Wikipedia views have a statistically significant impact on Bitcoin price. However, in contrast to other studies, such as [23], the authors suggest that macro financial indicators (e.g., Dow Jones Index or crude oil price) do not significantly affect Bitcoin price.

In [21], the authors show that Bitcoin price has a strong cross-correlation with the number of transactions and transaction fees. In addition, good cross-correlation with gold and crude oil price and a moderated cross-correlation with contemporary stock market indices (such as NASDAQ, DAX and S&P500). Finally, in an economical study presented by [22] is analyzed the causal relation between trading volume, Bitcoin returns and volatility. Thus, it is presented that volume can predict returns, but not volatility.

In terms of time series prediction (that is the context of this paper), in [29], the authors also select relevant attributes. However, the prediction was done by using a Bayesian Neural Network

(BNN). Thus, the BNN was compared to a Support Vector Regression (SVR) and linear models. The time series covers the daily Bitcoin data from Sep 11, 2011, to Aug 22, 2017. Based on this data set, the BNN was parameterized, trained and tested to predict the log price and the log volatility of Bitcoin price. The results obtained present MAPEs equals to 0.0198 and 0.6302 for log price and log volatility, respectively.

The paper of [2] combines a GARCH model with the SVR. Thus, it was evaluated its performance for cryptocurrencies (Bitcoin, Ethereum and Dash market price) and traditional currencies (Euro, British pound and Japanese yen). All of them were considered in US dollars. Moreover, it was used low (daily) and high (hourly) frequency data to predict the volatilities. The authors show that the errors, RMSE and MAE, obtained from high frequency data are much lower than for low frequency data.

Other research has been conducted on predicting fluctuations in the price and number of transactions of three relevant cryptocurrencies (Bitcoin, Ethereum and Ripple) [26]. Thus, it was used comments (from people) in online cryptocurrency communities (bitcointalk.org, forum.ethereum.org and www.xrpchat.com). In this study, a total of 793 instances was divided into 88% to train and 12% to test the model. This data were tagged in positive or negative, using VADER engine. However, it was determined 5 categories: very positive, positive, neutral, negative and very negative. Fluctuations in the price of Bitcoin demonstrated to be significantly correlated with the number of positive/very positive comments and with positive replies. Granger causality test was used to get the maximum accuracy of 79.57%, an f1-score of 0.796 and Matthews correlation coefficient of 0.606.

On works related to predicting direction of the Bitcoin exchange rate, in an empirical study realized by [25], the authors used Open, High, Low, and Close (OHLC) data from CoinDesk and the hash rate taken from the Blockchain (1066 instances — using 80% as training data and the remaining 20% as test data). This data are standardized (mean equal to 0 and standard deviation equal to 1) and used to obtain the Simple Moving Average (SMA), which can improve the capacity of the model to recognize trends by smoothing the data. In addition, all extracted attributes were used as inputs of deep learning models: Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network. It was observed that LSTM achieves the highest classification accuracy, about of 52%.

Therefore, the present study argues that the aforementioned literature offers possibilities to explore a new set of attributes and different configurations of machine learning techniques to improve both the direction prediction and the maximum, minimum and closing price prediction for Bitcoin.

3. Proposed methodology

The methodology proposed in this paper can be visualized on Fig. 1, which summarizes the: (1) sources where the data were collected; (2) transformations used for data pre-processing; (3) the data partitioning for training and testing purposes; (4) sort of attribute selection methods applied; (5) application of machine learning techniques to predict the price direction using single and ensemble approaches, including classification by regression to predict the maximum, minimum and closing prices; and (6) performance evaluation metrics used.

3.1. Collected data

The sources of information can be categorized into internal (the behavior of different parameters of Bitcoin) and external (the economic factors, external demand or information obtained from social networks or specialized forums, also named as public).

In this sense, as an internal data source, the Blockchain information is considered in a similar way as suggested by [22,25]. This information includes opening, maximum, minimum, and closing Bitcoin exchange rate (OHLC), the volume of trades, total transaction fees, number of transactions, cost per transaction and average hash rate. In Fig. 2, it is presented the behavior of the Bitcoin exchange rate (OHLC). Thus, it can be highlighted the high volatility, especially for the minimum price (Low price) that presents a high fall in Jun 23th, 2016. This represents that Bitcoin market, in general, is immature yet.

However, it is possible to observe, in Fig. 3, that the Blockchain data shows a similar volatility, which can be used to predict the exchange rate.

As a contribution of the present paper to the identification of relevant attributes for the prediction of the Bitcoin price trend, external information was considered and obtained from international economic indicators. These indicators were used due to the high correlation identified by [21] and the good results obtained by [30]. Thus, the following indicators were used: crude oil future prices, gold future prices, S&P500 future, NASDAQ future, and DAX index. Fig. 4 shows the historical data for these economic indicators.

The OLHC exchange rates (Bitcoin against US dollar) were collected from the website http://bitcoincharts.com and the remaining internal data were obtained from the website http://quandl.com. The external information was collected from http://investing.com.

In order to compare the proposed methodology with the state-of-the-art, specifically with the models proposed by [25], an identical data interval (named as *interval 1*) was considered, ranging from August 19th, 2013 to July 19th, 2016. However, a second interval was considered, ranging from April 1st, 2013 to April 1st, 2017 (named as *interval 2*).

3.2. Data pre-processing

In this paper, the data pre-processing stage suggested by [30] was used, i.e., the lag period concept and the smoothing of the data. Thus, in the pre-processing stage, the value "1" was assigned to the class if the closing exchange rate of Bitcoin at a Day (D) is greater than or equal to the previous day ($D-1$). Otherwise, it was assigned the value "0". Unlike the case presented by [30], the Bitcoin cryptocurrency is traded every hour and every day. For this reason, it was considered a lag period of 7 days. This way, for each class ("0" or "1"), at time D, it was considered historical data from the previous 7 days as input attributes. Similarly, it was considered a period of data smoothing of 30 days and, consequently, the Weighted Moving Average (WMA) was calculated for 30 days to all input attributes. The WMA calculation is used to identify possible trends in the exchange rate, which can be expressed as:

$$WMA_M = \frac{\sum_{n=1}^{M} np_n}{\sum_{n=1}^{M} n}, \tag{1}$$

where $M = 30$ due to the number of days considered in the WMA. So, p_n corresponds to the value $M-n$ days before the current day.

After the pre-processing stage, it was obtained a data set composed of the attributes shown in Table 1.

3.3. Data partitioning

In order to compare the obtained results with the methodology proposed by [25], an identical data interval (named as *interval 1*) was considered, which considers the same data partitioning (80% of the data for training and the remaining 20% most recent data for validation/test). In addition, a larger data interval was also considered and used to generate a baseline for future researches.

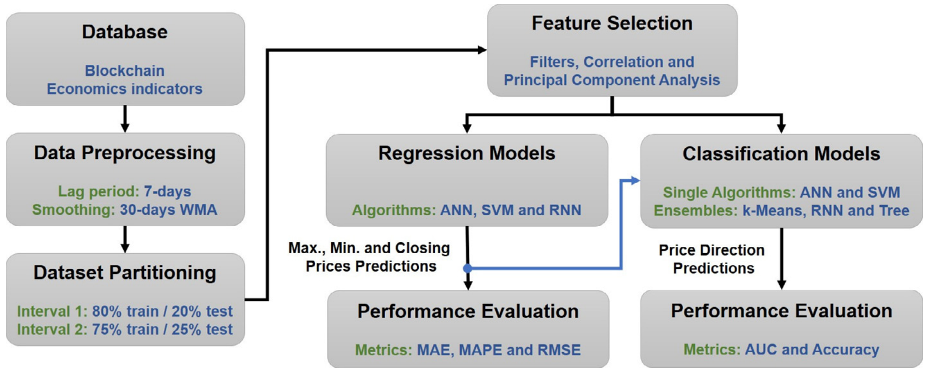

Fig. 1. Overview of the proposed methodology.

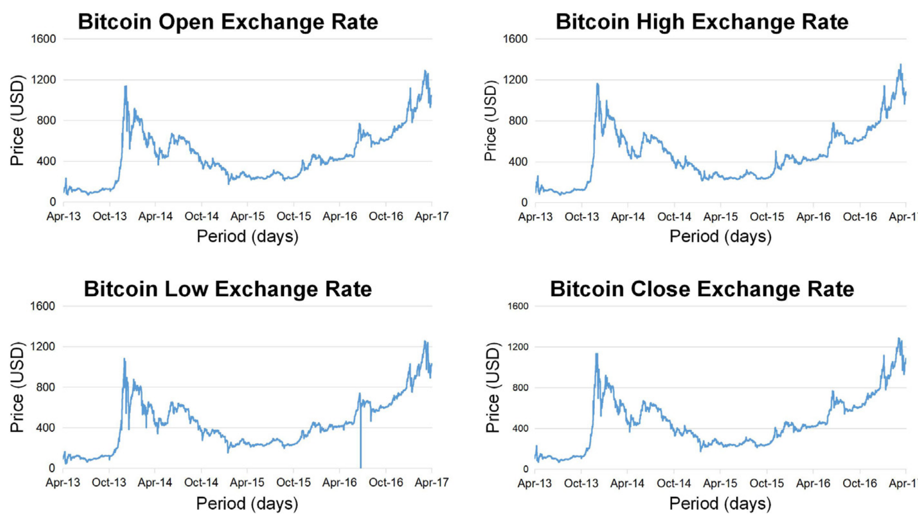

Fig. 2. Bitcoin exchange rate (OHLC).

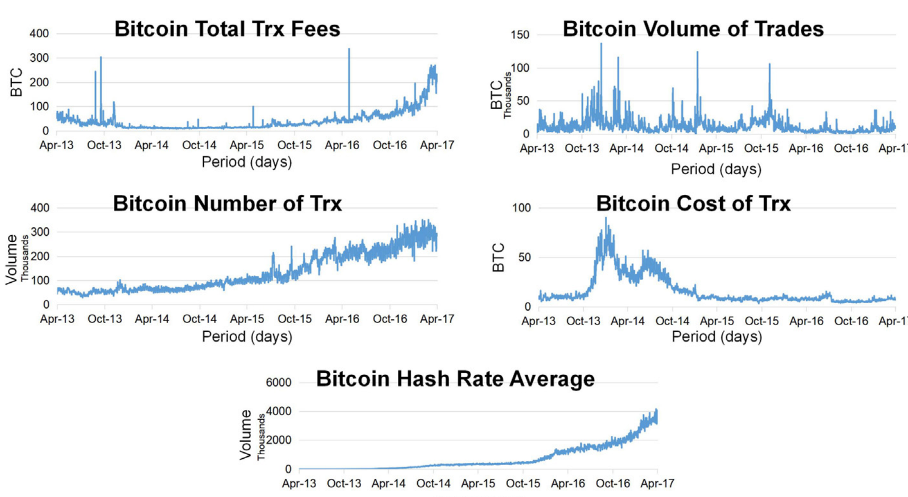

Fig. 3. Additional Blockchain information.

This larger interval was prepared (named as *interval 2*), which considers 75% of the data for training and the remaining 25% most recent data for validation/test.

These data sets, *interval 1* and *interval 2*, were used in the training and validation/testing process of all machine learning algorithms that will be presented in Section 3.5.

3.4. *Attribute selection*

As indicated in Section 3.2, it was considered up to 86 possible input attributes. Therefore, in order to reduce this high dimensionality, it was necessary to use methods to select the most relevant attributes. From the review of the literature, it was verified that

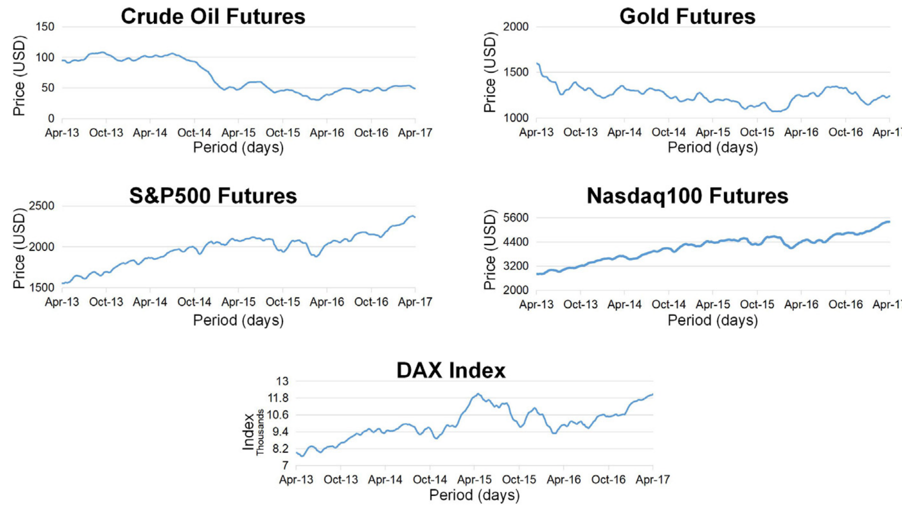

Fig. 4. Economic indicators.

Table 1
List of possible input attributes.

Day D	Day $(D-i)$	30-day WMA
Opening price	Price direction	Opening price
Timestamp	Opening price	Maximum price
	Maximum price	Minimum price
	Minimum price	Closing price
	Closing price	Volume of trades
	Volume of trades	Number of txn
	Number of txn	Transaction fees
	Transaction fees	Cost per txn
	Cost per txn	Hash rate avg
	Hash rate avg	Closing crude price
		Closing gold price
		Closing S&P500 price
		Closing Nasdaq price
		Closing DAX price

the proposal of [30] uses the degree of correlation to identify the most relevant attributes for the stock market. However, because the nature of Bitcoin is different from a stock market, as cited by [7] and [20], it was preferred to explore different selection and transformation techniques to reduce the dimensionality of the data set. Thus, five attribute selection techniques were considered, which are summarized as follows:

1. Correlation analysis (*Corr*) — this function evaluates the value of an attribute by measuring the cross-correlation (Pearson's coefficient) between it and the class;
2. Relief technique (*Relief*) — this method estimates the worth of an attribute by repeatedly sampling an instance and considering the value of the given attribute of the nearest instance of the same and different class [31];
3. Information Gain method (*InfoGain*) — this technique assesses the value of an attribute by measuring the gain of information relative to the class, using the concept of entropy [32];
4. Principal Component Analysis (*PCA*) — reduces dimensionality by choosing sufficient eigenvectors to explain a percentage of the variance in the original data (95%) [33,34];
5. Correlation-based Feature Subset selection (*CFS*) — evaluates the value of a subset of attributes, considering the individual predictive capacity of each attribute along with the degree of redundancy between them [35].

It is important to mention that for all of the five selectors, the 20 best attributes were selected. These algorithms were executed by means of Waikato Environment for Knowledge Analysis version 3.8.1 (WEKA).

3.5. Soft computing algorithms applied to the predictions

As can be seen in the previous sections, some machine learning techniques such as Artificial Neural Networks (ANN) and Support Vector Machines (SVM) are widely used in stock market predictions. Thus, in this paper, it was proposed a comparison between such techniques in relation to ensembles that combine regression models with classification and clustering algorithms.

3.5.1. Artificial neural network

The ANNs have been widely applied in the forecasting and prediction of direction for stock values. Thus, the Multilayer Perceptron (MLP) architecture was used due to its flexibility and good results in other studies [30,36,37]. In this paper, the hidden layers use the hyperbolic tangent transfer function and, for the output layer, it was used the logistic transfer function, such as presented in Fig. 5.

The scaled conjugate gradient was employed as learning method and *crossentropy* as performance metric (recommended for classification purposes). Thus, several configurations with one and two hidden layers were tested, with combinations of 5, 10, 15, 20, 25, 30 and 35 neurons with a number of epochs ranging from 20 to 500. This algorithm was implemented and parameterized in Matlab® platform.

3.5.2. Support vector machine

The SVM algorithm is based on the principle of minimization of structural risk. Moreover, it estimates a function that reduces the generalization error, demonstrating a resistance to the problem of overfitting. It is important to mention that the SVM is not a stochastic technique. Therefore, if the dataset is not changed, the same result will be always obtained [15].

The basic idea is to create a hyperplane that can separate the classes of the problem [37]. Since each sample in each side of the hyperplane have a distance to it, the smallest distance is called the separation margin. The hyperplane is optimal, if the margin is maximized. Therefore, the training process of the SVM consists of finding the optimal hyperplane, that is the one with the maximum distance from the nearest training samples [38]. In order to avoid the excessive computational cost for calculating the optimal hyperplane, the concept of "soft margin" is used, which establishes a

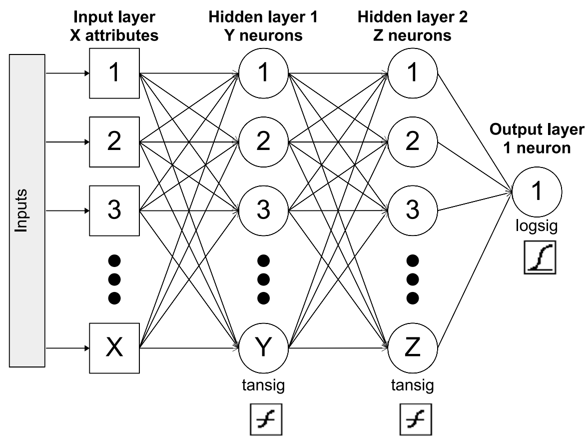

Fig. 5. ANN architecture employed.

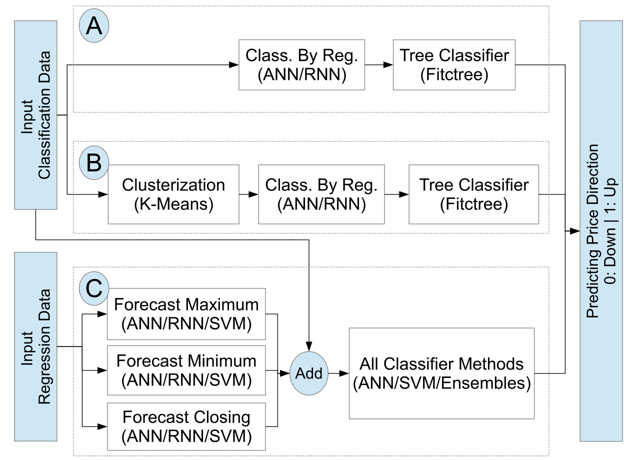

Fig. 6. Ensemble of machine learning techniques.

tolerance level (C) to accept samples that are not within the limit established by the hyperplane [39].

In cases where the data are not linearly separable, Cover's theorem is used, which suggests raising the dimensionality to achieve a linear separation. In this way, the SVM makes use of *kernels* that allow to raise the dimensionality of the data and, thus, achieve to separate them linearly [39]. The *kernels* used in this paper are described in Eqs. (2) and (3):

$$Polynomial = (x^T x')^d, \tag{2}$$

$$Gaussian = \exp(-\frac{\|x - x'\|^2}{2\sigma^2}), \tag{3}$$

where d is the degree and σ is the gamma parameters.

Thus, when using a polynomial kernel, it will be necessary to define the parameter d, that represents the degree of the polynomial expressed in Eq. (2). On the other hand, if the model uses a radial kernel, then the standard deviation (σ) must be defined in Eq. (3).

It is important to mention that the classical SVM algorithm requires to solve a quadratic optimization problem. In order to avoid the amount of memory needed, it was used the Sequential Minimal Optimization algorithm, described in [40].

3.5.3. Ensembles

Three machine learning ensemble algorithms are proposed to compare their prediction performance with the classifiers (ANN and SVM) mentioned above. Fig. 6 presents an overview of ensembles A, B and C proposed in this paper.

Ensemble A

First, with the input data prepared in Section 3.2, it is executed a regression model based on RNN (Jordan architecture) to classify the Bitcoin price direction. After that, the result is used as input of a tree classifier model that predicts the Bitcoin price direction.

For classification by regression, the RNNs use hyperbolic tangent transfer functions in the hidden layers and, for the output layer, it was used the linear transfer function, such as presented in Fig. 7.

As a learning method, the gradient descent with momentum and an adaptive learning rate was employed. The Mean Square Error (MSE) metric was used to evaluate the performance during the training stage. Thus, several configurations with one and two hidden layers were tested, with combinations of 5, 10, 15, 20, 25, 30 and 35 neurons with a number of epochs ranging from 20 to 500. This method was also implemented in Matlab® platform.

The tree classifier model (Fig. 8) was configured with only one decision rule that is $X > n$, where X is the input value and n is a

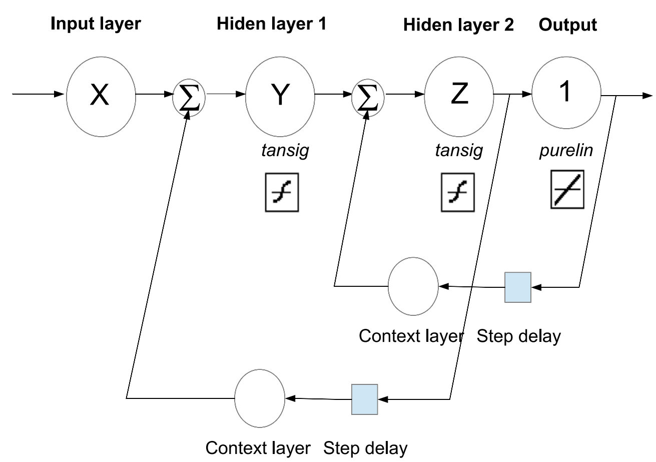

Fig. 7. RNN architecture.

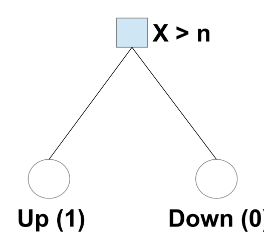

Fig. 8. Decision tree classifier.

threshold adjusted during the training process. Thus, it is worth to mention that, after train all the models generated, n ranged from 0.5 to 0.65.

Ensemble B

In this ensemble model, a clustering method is executed previously to apply the RNN (classification by regression) and the tree classifier used in the Ensemble A. The clustering method is based on the k-Means algorithm, where k represents the number of clusters that will be generated.

According to [41], the k-Means works by stages: (a) two-dimensional input data with three clusters; (b) three seed points selected to generate k cluster centroids and initial classification of the data points to these clusters; (c) intermediate iterations x updating cluster labels and their centroids; (d) final clustering obtained after the convergence. For implementation of this technique was considered two clusters and used the *city block* distance metric.

Ensemble C

The last ensemble is responsible for executing two tasks: (1) the forecasting of Maximum, Minimum and Closing Bitcoin exchange rates, using ANN, RNN and SVM in their regression versions; and (2) use the outputs of the forecasting process as inputs to classifiers (ANN, SVM, and Ensembles A and B) in order to predict the Bitcoin price direction.

Thus, as inputs for each forecasting method, it was used the most relevant attributes selected by each attribute selection technique mentioned in Section 3.4.

The ANNs used a similar architecture presented in Fig. 5, changing only the transfer function of the output layer by a linear function. In addition, it was used the Levenberg–Marquardt learning method and the MSE as performance metric. The same combinations of the neurons, layers and epochs presented in Section 3.5.1 were used.

Each model generated was evaluated by regression performance metrics presented in the following Eqs. (4), (5) and (6):

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|Y_i - \overline{Y_i}\right|, \tag{4}$$

$$MAPE = \frac{100}{N}\sum_{i=1}^{N}\frac{\left|Y_i - \overline{Y_i}\right|}{Y_i}, \tag{5}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \overline{Y_i}\right)^2}, \tag{6}$$

where N is the number of data to be evaluated, Y_i is the ith predicted value, and $\overline{Y_i}$ is the ith desired value.

Finally, the maximum, minimum and closing Bitcoin exchange rates predicted values were added to the input of each classifier (ANN, SVM, and Ensembles A and B), which are responsible for predicting the Bitcoin price direction.

3.6. Performance metrics used to evaluate the price direction prediction

To compare the performance of each individual classifier and ensemble used for the purpose of predicting the Bitcoin price direction, it was used the area under the ROC curve (AUC). Thus, it was calculated the sensitivity and the specificity, such as described in the Eqs. (7) and (8), respectively:

$$Sensitivity = \frac{TP}{TP + FN}, \tag{7}$$

$$Specificity = \frac{TN}{TN + FP}, \tag{8}$$

where TP = True Positive; TN = True Negative; FP = False Positive; and FN = False Negative.

The accuracy (Eq. (9)) is considered to compare the models proposed in this paper (with the best performances) with the state-of-the-art results:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \times 100\%. \tag{9}$$

It can be highlighted that each generated model was evaluated (trained and validated/tested) 50 times in order to obtain values statistically significant for each considered performance metric.

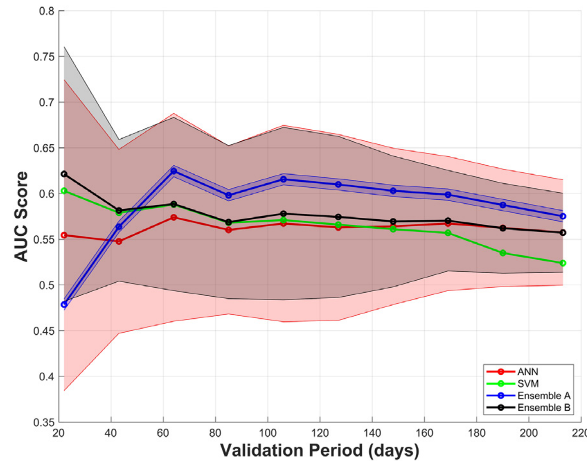

Fig. 9. Performance validation for the first data set (interval 1), considering 95% of confidence.

4. Results

4.1. Prediction of price direction

The classification strategies described above were evaluated and the best results of each of them are presented, in Tables 2 and 3, for the first and second intervals described in Section 3.3, respectively.

In the column "Algorithm:Arch. (fs)" is mentioned the machine learning technique used, its architecture and, in parenthesis, the attribute selection method applied to the data set (where, *all* represents that the better results were obtained using all attributes). In the case of ANN and Ensembles A and B, the architecture used is described as: "h1"-"h2"-"e", where "h1" is the number of neurons used in the first hidden layer, "h2" is the number of neurons considered in the second hidden layer and "e" represents the number of epochs used in the training stage. For SVM, the architecture is described as "c"-"d", where "c" represents the cost parameter and "d" is the degree of the kernel polynomial function.

The column "Individually" means that the classifiers where individually employed, i.e., the Ensemble C was not considered. While the results presented in the column "Ensemble C" means that the Ensemble C was taken into consideration, i.e., the forecasting of Maximum, Minimum and Closing Bitcoin exchange rates was used as inputs for each classifier.

Analyzing the results presented in Tables 2 and 3, it can be observed that the Ensemble C did not demonstrate good performance for both data sets (intervals 1 and 2). Therefore, in the sequence, Figs. 9 and 10 show the AUC score for different periods of validation (in days). In these graphs, the interval areas are highlighted with a statistical confidence of 95% (±2 times the standard deviation).

In the first interval (Table 2), the best result was obtained by Ensemble A that has the greatest value of AUC (0.58) and an accuracy of 62.91%. It was used the correlation analysis technique as attribute selection method, without including the predicted values of maximum, minimum and closing Bitcoin exchange rates. Table 4 shows the attributes selected by the *Corr* method, which were used to obtain this result. It is important to observe that only information from the Blockchain was used.

Analyzing the results obtained for the second interval (Table 3), the SVM was the algorithm with the best performance (with 0.58 of AUC and 59.45% of accuracy). The data set used was composed of all attributes described in Table 1, but without including the predicted values of maximum, minimum and closing Bitcoin exchange rates.

Table 2
Interval 1 - best performances obtained by each classifier.

Algorithm:Arch. (fs)	Individually		Algorithm:Arch. (fs)	Ensemble C	
	AUC	Acc. (%)		AUC	Acc. (%)
ANN:20-0-100 (Corr)	0.56 ± *0.03*	58.84 ± *7.25*	ANN:25-30-500 (Corr)	0.51 ± *0.02*	46.10 ± *4.62*
SVM:1-1 (CFS)	0.52 ± *0.00*	56.81 ± *0.00*	SVM:1-1 (CFS)	0.51 ± *0.00*	56.34 ± *0.00*
Ens. A:5-10-500 (Corr)	**0.58** ± *0.00*	**62.91** ± *0.00*	Ens. A:25-5-20 (Corr)	0.51 ± *0.00*	42.72 ± *0.00*
Ens. B:5-15-50 (all)	0.56 ± *0.03*	61.31 ± *3.89*	Ens. B:20-5-1000 (all)	0.54 ± *0.01*	60.83 ± *0.48*

Table 3
Interval 2 - best performances obtained by each classifier.

Algorithm:Arch. (fs)	Individually		Algorithm:Arch. (fs)	Ensemble C	
	AUC	Acc. (%)		AUC	Acc. (%)
ANN:25-0-100 (InfoGain)	0.54 ± *0.03*	53.40 ± *5.40*	ANN:15-30-20 (InfoGain)	0.51 ± *0.02*	46.11 ± *5.78*
SVM:1-1 (all)	**0.58** ± *0.00*	**59.45** ± *0.00*	SVM:1-1 (all)	0.55 ± *0.00*	56.44 ± *0.00*
Ens. A:25-0-500 (InfoGain)	0.54 ± *0.00*	48.85 ± *0.00*	Ens. A:20-15-20 (InfoGain)	0.50 ± *0.00*	60.50 ± *1.67*
Ens. B:25-20-500 (Corr)	0.55 ± *0.02*	58.19 ± *2.37*	Ens. B:5-25-1000 (Corr)	0.52 ± *0.00*	42.16 ± *0.75*

Table 4
Attributes selected by the *Corr* method for interval 1.

Day *D*	Day *(D-i)*	30-day WMA
Open. price	Open. price (i:1,5,6,7) Max. price (i:6,7) Min. price (i:1,2,4,6,7) Closing price (i: 1,6,7)	Opening price Maximum price Minimum price Closing price Transaction fees

Table 5
Comparison of accuracy with the models proposed by [25].

Model	Accuracy (%)
Ensemble A:5-10-500 (Corr)	**62.91**
LSTM [25]	52.78
RNN [25]	50.25
ARIMA [25]	50.05

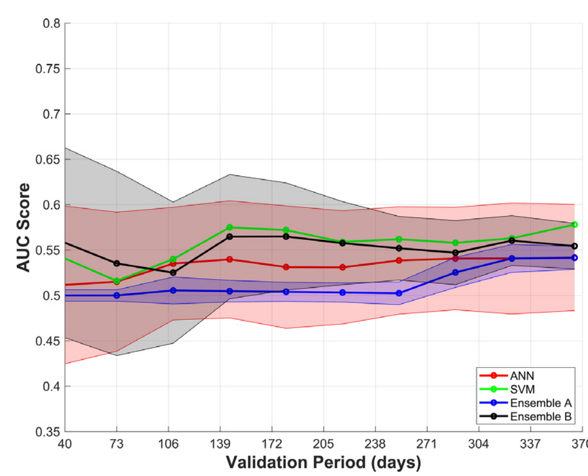

Fig. 10. Performance validation for the first data set (interval 2), considering 95% of confidence.

In addition, the best result obtained by the Ensemble A was compared with those presented by [25], considering the same range of data (interval 1). For this comparison, only the accuracy metric could be used (see Table 5).

Comparing the results obtained by the LSTM algorithm proposed by [25], it was possible to note that the performances, in terms of accuracy, of all the individual algorithms (shown in Table 2, column "Individually") proposed in this paper are better.

4.2. *Forecasting of maximum, minimum and closing Bitcoin exchange rates*

In Tables 6 and 7 are presented the best results for regression (forecasting) experiments considering the first and second intervals, respectively.

For both intervals, the best results were obtained by the SVM algorithm (in its regression version) using attributes selected by the Relief technique. Moreover, the SVM obtained the best results to forecast the maximum, minimum and closing Bitcoin exchange rates.

In the case of minimum exchange rate, the MAPE metric is very high because in Jun 23th, 2016 its value decreases from $588.03 to $1.5. However, if this date is not considered, the MAPE obtained using the SVM regression model decreases from 183.7% and 107.8% to 1.52% and 1.58% for the first and second intervals, respectively.

Tables 8–10 show the most effective attributes selected by the *Relief* method to forecast the maximum, minimum and closing Bitcoin exchange rates. It can be noted that only the information from Blockchain was considered as relevant.

Analyzing Table 8, the second interval presents some attributes equal to the first interval. The exceptions were the maximum price of the *D*-3 and the minimum price 30-day WMA. Moreover, the volume of trades *D*-5 and the cost of transaction *D*-1 were added.

Table 6
Interval 1 - best performances obtained by each regression/forecasting algorithm.

Value	Algorithm:Arch.(fs)	MAE	MAPE (%)	RMSE
Max.	ANN:5-0-100(Relief)	27.02 ± *37.58*	4.94 ± *6.72*	65.29 ± *110.7*
	RNN:10-10-500(InfoGain)	19.97 ± *0.00*	3.80 ± *0.00*	32.16 ± *0.00*
	SVM:0.8-1(Relief)	**6.70** ± *0.00*	**1.28** ± *0.00*	**12.12** ± *0.00*
Min.	ANN:5-5-20(CFS)	14.58 ± *3.52*	183.9 ± *1.37*	45.90 ± *4.39*
	RNN:30-10-500(CFS)	13.51 ± *0.00*	183.2 ± *0.00*	42.48 ± *0.00*
	SVM:0.8-1(Relief)	**10.08** ± *0.00*	**183.7** ± *0.00*	**42.66** ± *0.00*
Close	ANN:5-0-20(CFS)	19.06 ± *10.06*	3.86 ± *2.02*	25.85 ± *13.83*
	RNN:30-25-500(CFS)	14.54 ± *0.00*	3.08 ± *0.00*	18.56 ± *0.00*
	SVM:1-1(Relief)	**9.63** ± *0.00*	**1.91** ± *0.00*	**15.92** ± *0.00*

Table 7
Interval 2 - best performances obtained by each regression/forecasting algorithm.

Value	Algorithm:Arch.(fs)	MAE	MAPE (%)	RMSE
Max.	ANN:5-5-20(CFS)	55.03 ± *73.48*	6.51 ± *9.40*	83.42 ± *95.86*
	RNN:10-35-500(CFS)	14.04 ± *0.00*	2.03 ± *0.00*	20.38 ± *0.00*
	SVM:0.9-1(Relief)	**9.23** ± *0.00*	**1.14** ± *0.00*	**17.17** ± *0.00*
Min.	ANN:5-0-20(CFS)	44.40 ± *31.20*	112.3 ± *4.22*	79.99 ± *52.96*
	RNN:10-35-500(CFS)	32.65 ± *0.00*	113.6 ± *0.00*	51.97 ± *0.00*
	SVM:1-1(Relief)	**13.26** ± *0.00*	**107.8** ± *0.00*	**41.08** ± *0.00*
Close	ANN:5-0-20(CFS)	26.14 ± *5.18*	3.06 ± *0.60*	41.62 ± *7.22*
	RNN:30-30-500(CFS)	27.85 ± *0.00*	3.36 ± *0.00*	42.34 ± *0.00*
	SVM:1.3-1(Relief)	**14.32** ± *0.00*	**1.81** ± *0.00*	**25.47** ± *0.00*

Table 8
Attributes selected by the *Relief* method to forecast the maximum Bitcoin price.

Interval *I*	Day *D*	Day $(D-i)$	30-day WMA
Interval 1	Open. price	Open. price (i:1,2) Max. price (i:1-3) Min. price (i:1,2) Closing price (i:1-3) Vol. of trades (i:1-4) Cost Trx. (i:2)	Min. price Vol. of trades Cost Trx. Trx. Fee
Interval 2	Open. price	Open. price (i:1,2) Max. price (i:1,2) Min. price (i:1,2) Closing price (i:1-3) Vol. of trades (i:1-5) Cost Trx. (i:1,2)	Vol. of trades Cost Trx. Trx. Fee

Table 9
Attributes selected by the *Relief* method to forecast the minimum Bitcoin price.

Interval *I*	Day *D*	Day $(D-i)$	30-day WMA
Interval 1	Open. price	Open. price (i:1,2) Max. price (i:1,2) Min. price (i:1,2) Closing price (i:1-3) Vol. of trades (i:1-3) Cost Trx. (i:1-4,6,7)	Vol. of trades
Interval 2	Open. price	Open. price (i:1) Max. price (i:1,2) Min. price (i:1,2) Closing price (i:1-3) Vol. of trades (i:1-3) Cost Trx. (i:1-7)	Vol. of trades

Comparing the attributes selected for the first and second intervals to forecast the minimum price (Table 9), only the opening price of the *D*-2 day was excluded. On the other hand, the cost of transaction of the *D*-5 day was added.

In addition, in Table 10, it can be observed that, for closing prices, only the cost transaction of the *D*-7 day was not considered for the second interval. However, the transaction fee 30-day WMA was considered as relevant for this forecasting.

Table 10
Attributes selected by the *Relief* method to forecast the closing Bitcoin price.

Interval *I*	Day *D*	Day $(D-i)$	30-day WMA
Interval 1	Open. price	Open. price (i:1) Max. price (i:1,2) Min. price (i:1,2) Closing price (i:1,2) Vol. of trades (i:1-3) Cost Trx. (i:1-7)	Vol. of trades Cost Trx.
Interval 2	Open. price	Open. price (i:1) Max. price (i:1,2) Min. price (i:1,2) Closing price (i:1,2) Vol. of trades (i:1-3) Cost Trx. (i:1-6)	Vol. of trades Cost Trx. Trx. Fee

5. Discussions

5.1. Prediction of Bitcoin price direction

Analyzing the data of Figs. 9 and 10, it is observed that the ANNs show a greater variation within the whole validation period. Also, it is possible to observe that the Ensemble A has less variation in all cases, because it is based on Recurrent Neural Network. However, in all cases, the Ensemble A do not achieve better results than traditional ANNs. In the case of *SVM* algorithm, this variability cannot be noticed because it is not a stochastic model.

Regarding the attribute selection methods (dimensionality reduction), the most successful were: correlation score (*Corr*), *InfoGain* and *CFS*. However, for the second data set (interval 2), it was necessary to use all the attributes (including international economic indicators). This means that the attributes in this interval show a similar importance.

Although the Ensemble B did not improve the classification results, for both intervals it shows better results for the first 50 days of prediction. In addition, it shows a lower variability than the ANN algorithm.

5.2. Forecasting of maximum, minimum and closing bitcoin exchange rates

Similar to the prediction of Bitcoin price direction, stochastic models based on RNN show less variability than those based on ANN. However, here it can be mentioned that the SVM models show better performances in all cases and for both intervals.

In terms of dimensionality reduction, the best methods were: *Relief*, *CFS* and *InfoGain*. In both intervals, it was possible to reduce the number of attributes with an improvement in the performance of each forecasting algorithm.

One of the objectives of this paper is to present the best results in forecasting of Bitcoin exchange rates to serve as a baseline for future works and to guide traders. Thus, it is observed that without considering the abrupt fall of June 23th, 2016, the best model obtained an MAE between 6.70 and 9.63; and an MAPE between 1.28% and 1.91%. Although these values can still be considered very high, this is due to the high volatility of Bitcoin and, therefore, is a challenge for future efforts.

6. Conclusions

As previously presented, the proposed methodology verified the relevance of different attributes both for the price direction prediction and also for the predictions of maximum, minimum and closing prices of the Bitcoin. In terms of price direction, the Correlation analysis (*Corr*) was the technique with the highest effectiveness rate when considering the same data interval used in the literature (named as *interval 1*). However, for a larger interval (named as *interval 2*), the best result was obtained through the use of all proposed attributes, i.e., without the selection of attributes. On the other hand, when considering the predictions of maximum, minimum and closing prices, it was observed that *Relief* technique presented the best results.

In addition to the definition of the most relevant attributes, another important objective was reached, that is the achievement of better performances than those presented in the state-of-the-art for the Bitcoin price direction predictions. Thus, using the data of the *interval 1*, the better accuracy performance was obtained by the Ensemble A (62.91%). Comparing this result with that obtained by [25], the proposed methodology shows about 10% of accuracy improvement. It was also analyzed the *interval 2*, where the SVM algorithm presented the best performance (59.45%).

Regarding the regression experiments, the SVM algorithm obtained the best results for all predictions (maximum, minimum and closing prices) and for both intervals. In terms of maximum price prediction, it was obtained low MAPE (1.28% and 1.14% for *intervals 1* and *2*, respectively). The same occurs to forecast the closing price, where the SVM presents 1.91% and 1.81% of MAPE for *intervals 1* and *2*, respectively. The worst results were obtained for the minimum price (183.7% and 107.8% of MAPE). However, these results were a consequence of an abrupt decrease of the Bitcoin in Jun 23th, 2016. Thus, by disregarding this date, the SVM obtains 1.52% and 1.58% of MAPE, respectively, demonstrating its potential to predict the Bitcoin exchange rates.

Based on the results obtained by other studies applied to financial market time series, it is observed that the performances of the predictions are around 0.7 of AUC. Thus, it can be assumed that is still possible to improve the results for the Bitcoin exchange rate. One of the possibilities for improvement is to use technical indicators commonly employed by traders (e.g., Larry Williams (%R), Relative Strength Index (RSI) Stochastic oscillators (%K and %D), Moving Average Convergence/Divergence (MACD) oscillator, among others), in addition to the economic indicators, OHLC and Blockchain information used in this paper.

Acknowledgments

This work has been supported by CAPES PAEC and CNPq, Brazil (Grant Number 420298/2016-9).

References

[1] S. Abboushi, Global virtual currency - brief overview, Competition Forum 14 (2) (2016) 230–236.
[2] Y. Peng, P.H.M. Albuquerque, J.M.C. de Sá, A.J.A. Padula, M.R. Montenegro, The best of two worlds: forecasting high frequency volatility for cryptocurrencies and traditional currencies with support vector regression, Expert Syst. Appl. 97 (1) (2018) 177–192.
[3] A. Urquhart, Price clustering in bitcoin, Econom. Lett. 159 (2017) 145–148.
[4] P. Katsiampa, Volatility estimation for bitcoin: a comparison of garch models, Econom. Lett. 158 (2017) 3–6.
[5] J. Iglesias de Ussel, Bitcoin: a new way to understand Payment Systems (Ph.D. thesis), Massachusetts Institute of Technology, 2015.
[6] S. Nakamoto, Bitcoin: a peer-to-peer electronic cash system, 2008, Retrieved April 25, 2016 from www.bitcoin.org.
[7] L. Cocco, G. Concas, M. Marchesi, Using an artificial financial market for studying a cryptocurrency market, J. Econ. Interact. Coord. 12 (2) (2017) 345–365.
[8] A. Cuthbertson, Bitcoin now accepted by 100,000 merchants worldwide. international business times, Retrieved April 25, 2016 from 2015, www.venturebeat.com.
[9] J. Chokun, Who accepts bitcoins as payment? list of companies, stores, shops., Retrieved June 11, 2016 from www.bitcoinvalues.net, 2016.
[10] E. Moreau, 13 Major Retailers and Services That Accept Bitcoin., Retrieved April, 2018 from www.lifewire.com, 2018.
[11] M.C.K. Khalilov, A. Levi, A survey on anonymity and privacy in bitcoin-like digital cash systems, IEEE Commun. Surv. Tutor. 20 (3) (2018) 2543–2585.
[12] M. Conti, S. Kumar, C. Lal, S. Ruj, A survey on security and privacy issues of bitcoin, IEEE Commun. Surv. Tutor. (Early Access) (2018) 1–39.
[13] K.H. McIntyre, K. Harjes, Order flow and the bitcoin spot rate, Appl. Econ. Finance 3 (3) (2016) 136–147.
[14] M. van Alstyne, Why Bitcoin has value, Commun. ACM 57 (5) (2014) 30–32.
[15] W. Huang, Y. Nakamori, S.Y. Wang, Forecasting stock market movement direction with support vector machine, Comput. Oper. Res. 32 (10) (2005) 2513–2522.
[16] L. Kristoufek, BitCoin meets Google Trends and Wikipedia: Quantifying the relationship between phenomena of the Internet era, Sci. Rep. 3 (2013) 1–7.
[17] L. Kristoufek, What are the main drivers of the bitcoin price? Evidence from wavelet coherence analysis, PLoS ONE 10 (4) (2015) 1–15.
[18] M. Matta, I. Lunesu, M. Marchesi, Bitcoin Spread Prediction Using Social And Web Search Media, in: UMAP Workshops, 2015, pp. 1–10.
[19] P. Ciaian, M. Rajcaniova, d'Artis Kancs, The economics of bitcoin price formation, Appl. Econ. 48 (19) (2016) 1799–1815.
[20] X. Li, C.A. Wang, The technology and economic determinants of cryptocurrency exchange rates: the case of bitcoin, Decis. Support Syst. 95 (2017) 49–60.
[21] S. Vassiliadis, P. Papadopoulos, M. Rangoussi, T. Konieczny, J. Gralewski, Bitcoin value analysis based on cross-correlations, J. Internet Bank. Commer. 22 (S7) (2017) 1.
[22] M. Balcilar, E. Bouri, R. Gupta, D. Roubaud, Can volume predict Bitcoin returns and volatility? A quantiles-based approach, Econ. Modell. 64 (2017) 74–81.
[23] Y. Zhu, D. Dickinson, J. Li, Analysis on the influence factors of Bitcoin's price based on VEC model, Financ. Innov. 3 (1) (2017) 1–13.
[24] A. Greaves, B. Au, Using the bitcoin transaction graph to predict the price of bitcoin, Retrieved April 24, 2016 from http://snap.stanford.edu/, 2015.
[25] S. McNally, Predicting the price of Bitcoin using Machine Learning (Ph.D. thesis), National College of Ireland, 2016.
[26] Y.B. Kim, J.G. Kim, W. Kim, J.H. Im, T.H. Kim, S.J. Kang, C.H. Kim, Predicting fluctuations in cryptocurrency transactions based on user comments and replies, PLoS ONE 11 (8) (2016) 1–18.
[27] F. Glaser, K. Zimmermann, M. Haferkorn, M.C. Weber, M. Siering, Bitcoin-asset or currency? revealing users' hidden intentions, in: 22th European Conference on Information Systems, 2014, pp. 1–14.
[28] C.Y. Wu, V.K. Pandey, The value of Bitcoin in enhancing the efficiency of an investor ' s portfolio, J. Financ Plann. 27 (9) (2014) 44–52.
[29] H. Jang, J. Lee, An empirical study on modeling and prediction of bitcoin prices with bayesian neural networks based on blockchain information, IEEE Access 6 (2018) 5427–5437.
[30] L.A. Laboissiere, R.A. Fernandes, G.G. Lage, Maximum and minimum stock price forecasting of Brazilian power distribution companies based on artificial neural networks, Appl. Soft Comput. 35 (2015) 66–74.
[31] L. Gao, T. Li, L. Yao, F. Wen, Research and application of data mining feature selection based on relief algorithm, J. Softw. 9 (2) (2014) 515–522.

[32] J. Dai, X. Qing, Attribute selection based on information gain ratio in fuzzy rough set theory with application to tumor classification, Appl. Soft Comput. 13 (1) (2013) 211–221.
[33] J. Wang, H. Hou, C. Wang, L. Shen, Improved v-support vector regression model based on variable selection and brain storm optimization for stock price forecasting, Appl. Soft Comput. 49 (2016) 164–178.
[34] S.B. Kim, P. Rattakorn, Unsupervised feature selection using weighted principal components, Expert Syst. Appl. 38 (5) (2011) 5704–5710.
[35] M.A. Hall, Correlation-based feature selection for machine learning The University of Waikato, 1999.
[36] M. Qiu, Y. Song, Predicting the direction of stock market index movement using an optimized artificial neural network model, PLoS ONE 11 (5) (2016) 1–11.
[37] Y. Kara, M.A. Boyacioglu, O.K. Baykan, Predicting direction of stock price index movement using artificial neural networks and support vector machines: the sample of the istanbul stock exchange, Expert Syst. Appl. 38 (5) (2011) 5311–5319.
[38] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, John Wiley & Sons, 2012.
[39] C.M. Bishop, Pattern recognition and machine learning, 2006, Korean Soc. Civil Eng. 60 (1) (2012) 78–78.
[40] J. Platt, Sequential minimal optimization: a fast algorithm for training support vector machines, Adv. Kernel Methods – Support Vector Learn. (1998) 1–21.
[41] A.K. Jain, Data clustering: 50 years beyond K-means, Pattern Recognit. Lett. 31 (8) (2010) 651–666.