DOI: 10.21917/ijsc.2018.0245

AUTOMATED CRYPTOCURRENCIES PRICES PREDICTION USING MACHINE LEARNING

Ruchi Mittal, Shefali Arora and M.P.S Bhatia
Division of Computer Engineering, Netaji Subhas Institute of Technology, India

Abstract

Currently, Cryptocurrency is one of the trending areas of research among researchers. Many researchers may analyze the cryptocurrency features in several ways such as market price prediction, the impact of cryptocurrency in real life and so on. In this paper, we focus on market price prediction of the number of cryptocurrencies based on their historical trend. For our study, we tried to understand and identify the daily trends in the cryptocurrency market which analyzing the features related to the price of cryptocurrency. Our dataset consists of over nine features relating to the cryptocurrency price recorded daily over the period of 6 months. We applied some machine-learning algorithms to predict the daily price change of cryptocurrencies.

Keywords:

Cryptocurrency, Bitcoin, Decentralization, Network, Price Prediction

1. INTRODUCTION

The exponential growth of Internet access has triggered new technologies and techniques in real life [6]. Cryptocurrency is one of the emerging Internets technology uses as currency over the traditional monetary system [2]. The term cryptocurrency means the digital currency or the virtual currency, which is works as a mode of exchange or transfer of assets digitally. The market of cryptocurrency has evolved at an exponential speed in a short span of time. The first cryptocurrency was introduced in 2009 named as Bitcoin by Satoshi Nakamoto [1]. Later on, there are thousands of other cryptocurrencies are running in the market. Unlike centralized banking system and electronic money, cryptocurrencies follow the decentralized system, which means it supports blockchain transactional databases [2] [4]. The centralized banking system means there is the hierarchy of network exist and government controls the overall currency system, on the other hand, there is no hold of government or any other agency exists on cryptocurrencies [5]. A sample blockchain is shown in Fig.1.

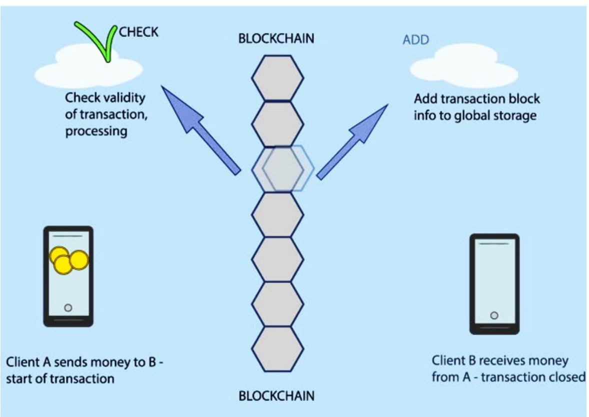

Fig.1. Cryptocurrency workflow using blockchain mechanism

Digital coins or cryptocurrencies are named so due to the use of encryption techniques in regulating transfers creation of coins. It is essential to understand the social and financial factors that determine the price of a bitcoin so that we can understand its impact on the economy of a nation.

Bitcoin as well as other cryptocurrencies have not gone down well with governments due to avoidance of financial systems and increased impossibility to allow cash movements and fight against activities that are not legal. These include the decision of Chinese government to get rid of Bitcoins in 2013, the bankruptcy of Mt. Gox which is one of the heads of Bitcoin trading and gain legitimacy after Brexit vote. These and many other issues have led to the need of studying about digital currencies intensely.

In this paper, we are predicting daily price changes for multiple cryptocurrencies. A few of them are bitcoin, ripple, NMC and so on. We propose an approach for the price prediction using one of the famous machine learning algorithms, i.e. multivariate linear regression. Our plan starts with data pre-processing, in which we clean up the dataset by removing rows with the missing value. Next, we examine the independent features in the dataset, which help us in predicting the highest price of the cryptocurrency. Next we find out the correlation between dependent and independent variables, and at last, we can predict the costs.

Paper outline: In section 2 we discuss the related work in the area of cryptocurrencies' price prediction. In section 3, we present our approach or methodology followed for the price prediction. In section 4, we discuss our dataset briefly, and in section 5, we discuss our experimental results.

2. RELATED WORK

Authors have described the working of Bitcoin in [6]. Unlike traditional payment systems dominated by US dollar, Bitcoin has its metric value called bitcoin. A bitcoin's value is derived from its use for making payments in the Bitcoin system. Authors have questioned the economists on whether bitcoins meet standard attributes of money or not [7].

As shown in Fig.2, these nodes or entities signify the payments made using Bitcoin [8]. These entities make transactions directly and do not need any central banks or networks. Every transaction is recorded chronologically in a blockchain by the participants in this network. Participants in this network compete to get rewards which are obtained by recording transactions in the Bitcoin system. Each participant keeps a copy of the public ledger or blockchain, and a well-defined process is used to choose the winning participant. There is decentralization of verifications and transactions. The procedure of giving awards to participants leads to many economic incentives and thus drives the system of Bitcoin. The reward essentially includes a voluntary fee and fresh bitcoins. In [9], authors discuss the fact that many cryptocurrencies have been making their mark like Bitcoin. This competition is healthy and is leading to more innovations in the field of security. Alternative cryptocurrencies have been suggested, and their areas are explored in this paper. Authors assume that price fluctuations of cryptocurrencies are based on a specific set of patterns. Cagnialp and Balenovich [10] showed that presence of different traders, as well as valuation assessment methods, also lead to the difference in models. Future changes in price can also be predicted using geometric patterns like double-top-and-bottom, triangle, heads-and-shoulders, etc. Lo et al. [11] used kernel regression to identify some of these patterns, and it did not lead to a very profiTable.trading strategy. Classification algorithms have also been used by various authors to predict the changes in stock price, and these algorithms can be made use in cryptocurrency analysis as well.

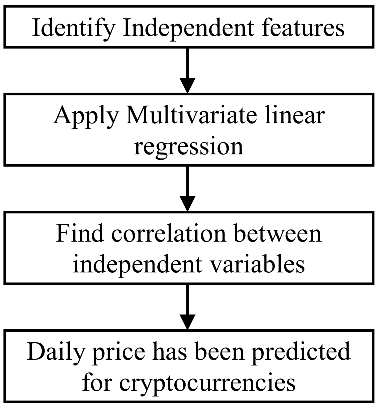

Fig.2. Working of bitcoins

3. APPROACH

Multivariate linear regression has been used to predict the highest and lowest prices of cryptocurrency. In this model, multiple independent variables contribute to a dependent feature with the help of multiple coefficients.

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2,\ldots,\theta_n x_n \quad (1)$$

Here, $h(\theta)$ is the estimate of output, based on n independent features and corresponding parameters θ. The cost function can be calculated as follows:

$$J(\theta_1, \theta_2,\ldots,\theta_n) = \frac{1}{2m}\sum_{i=1}^{m}\left(h_\theta\left(x_i\right)-y_i\right)^2 \quad (2)$$

Considering there are m data points in training data and $h_\theta(x_i)$ is the predicted output and y observed data of dependent variable. Thus linear regression model involving multiple features becomes a multivariate linear regression model. Computation of parameters is the most critical factor in this model. For computing the parameters, all independent variables are not taken up at the same time to minimize the error function. The best possible independent variables should be taken up, and the calculating correlation between independent and dependent variables could do this. This would decide which variable holds significance and which does not.

Given n features and m training examples, x has m rows and n+1 columns, where first term is the 0^{th} term added to every vector with a value of 1 (this is the coefficient of the constant term θ_0).

In our given dataset, we have made use of the following independent features to determine the highest price of cryptocurrencies. These features are:

- Open: Price at which cryptocurrency opened
- Low: Lowest price achieved
- Close: Closing price of cryptocurrency

Correlation is found between the highest price parameter High and these features. By these, a multivariate linear regression model is created, and predictions are made for the highest price. The values of correlation between dependent and independent variables are as shown in Table.1 below and a brief workflow of our approach is shown in Fig.2.

Table.1. Correlation Values

Correlation	Value
Open	0.972253
Low	0.9655539
Close	0.9797648

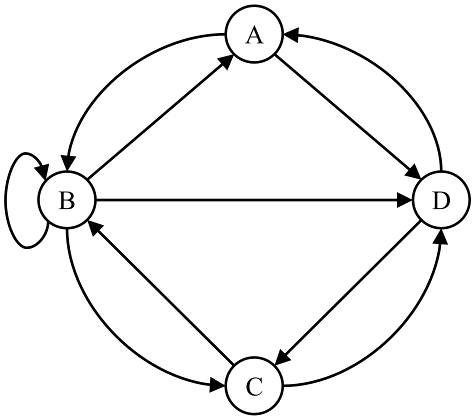

Fig.3. A brief workflow of our approach

4. DATASET DESCRIPTION

For our experiments, we pick the CryptocoinHistoricPrice dataset from popular databank Kaggle. This dataset is of size 20MB and has the historical price of around 1384 types of cryptocurrencies running currently. The data consists of 659360 rows and nine columns. Each column represents a parameter of the data such as date, open price, high price, low price, close price, volume, market cap, type of coin and delta. We choose this dataset to analyze behavior and structure of the market price of each of the cryptocurrency. A brief description of the dataset is shown in Table.2.

Table.2. Dataset description

Parameter	Description
Date	Date to which data is picked
Open	Open price of the day
High	High price of the day
Low	Low price of the day
Close	Closing price of the day

Volume	Total Volume of the currency
Market Cap	Market capital of the currency
Type	Type of the coin
Delta	Delta of the currency

5. EXPERIMENTAL RESULTS

R is used as the platform to predict the price of cryptocurrency based on dependent features. We determine the counts of different kinds of cryptocurrency present in the dataset.

Various statistical measures have been used to describe the model parameters (coefficients) and dependent variables. F-score or F-measure is a test of the model's accuracy. It considers precision and recall to compute the score. Precision gives the number of correct positive results divided by all positive results returned by the classifier. Recall is the number of positive results divided by all relevant samples. F-score is a harmonic average of both precision and recall. The p-value gives a statistical significance within a statistical hypothesis test, representing the probability of occurrence of an event. It is a number between 0 and 1, providing the smallest level of significance at which null hypothesis would be rejected. The t-value or t-statistic can be used to find the ratio of the departure of the estimated value of the parameter from the value obtained in the hypothesis to its standard error. It is used in finding population mean from the sample when the standard distribution is unknown.

Table.3. Types of cryptocurrency

Type of cryptocurrency	Count in dataset
BTC (Bitcoin)	1713
EXCL (Exclusive coin)	1192
XRC (Ripple)	1615
LTC (LiteCoin)	1713
LSK (Lisk)	639
XMR (Monero)	1324
SC (SiaCoin)	862
CNX (Cryptonex)	86
ADX (AdEx)	86
NLG (Guldencoin)	1368

The summary of linear regression model is given as follows:

Table.4. Summary of LR model

Parameter	Values
Residual standard error	184.2
Multiple R-squared	0.9701
Adjusted R-squared	0.9701
F-statistic	$7.122e^{+06}$
p-value	$< 2.2e^{-16}$

Table.5. Summary of LR model

	Estimate	Std. value	Error	$Pr>\|t\|$
Intercept	-0.2159282	0.227109882	-0.9507652	0.3417239
Close	0.9733615	0.003539152	718.9037505	0.0000000
Low	-0.6317852	0.2218223	-284.8159234	0.0000000
High	0.7283718	0.001559572	467.0331106	0.0000000

These prices are predicted based on the prices seen in the past three years i.e. 2017, 2016 and 2015. For example, given values of Open, Low and Close the High prices of cryptocurrency are predicted as follows for these cases:

Table.6. Results of linear regression

Open	Close	Low	Predicted-High
15123.7	14424	14595.4	16109.1906
16476.2	14208.2	15170.1	17790.0436
6777.77	6758.72	7078.5	7556.4003

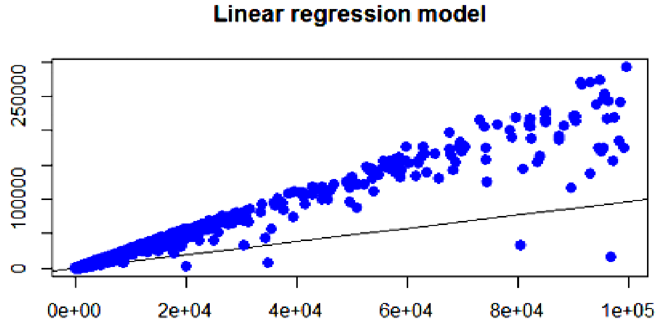

Fig.4. Visualization of linear regression model

Thus our model makes use of specific features as mentioned, to predict the highest price for Bitcoin on a given date. This can be shown in Fig.5, where authors show the variation between the open price, close price, low price, predicted price and actual price of three days for the bitcoin cryptocurrency.

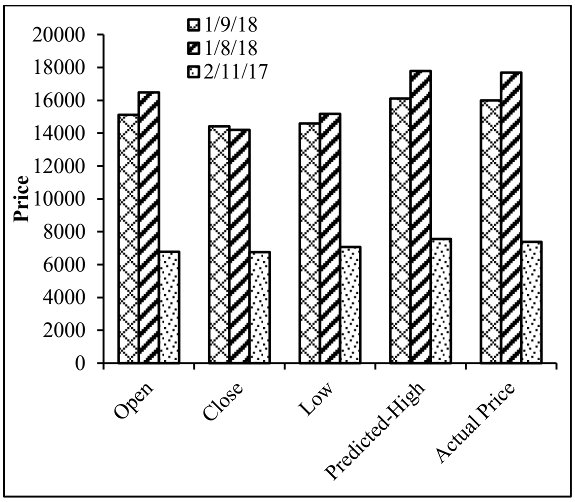

Fig.5. Variation in prices for three days

The above shown metrics are also modeled in Table.7, from where authors validate their approach by computing the accuracy of their approach. From Table.7, we see, there is 99% of accuracy obtained.

Table.7. Results of accuracy

Date	**1/9/2018**	**1/8/2018**	**2/11/2017**
Open	15123.7	16476.2	6777.77
Close	14424	14208.2	6758.72
Low	14595.4	15170.1	7078.5
Predicted-High	16109.19	17790.04	7556.4
Actual Price	15997.1	17680.2	7390.04
Accuracy	99.30452	99.38277	97.79842

6. CONCLUSION AND FUTURE WORK

In this paper, we have done few experiments to predict the price of various cryptocurrencies based on their open, high and low cost. This price prediction helps the number of users who are using cryptocurrencies for multiple types of transactions. These experiments allow us to get much more in-depth knowledge about the various aspects of the cryptocurrencies. In this paper, we applied our approach to the comparatively smaller dataset. For the sake of future work, we are planning to extend our analysis to the more significant dataset and may collaborate big data technologies with it. We believe that it will be interesting to study the impact of cryptocurrency in real life scenario and this become one of the accessible areas among researchers for researchers. We would also make use of deep learning models like LSTM to analyze various cryptocurrencies. Using LSTM, we can also make use of previous day's data to predict the highest price for the next day. We can decide the number of days by which we wish to predict the price. This is done by creating small data frames, with some days from the training set. The drawback of using this approach is that we might not have enough information for long-term analysis.

REFERENCES

[1] S. Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System", Available at: https://bitcoin.org/bitcoin.pdf, Accessed on 2008.

[2] D. Garcia, C.J. Tessone, P. Mavrodiev and N. Perony, "The Digital Traces of Bubbles: Feedback Cycles between Socio-Economic Signals in the Bitcoin Economy", *Journal of the Royal Society Interface*, Vol. 11, No. 99, pp. 1-28, 2014.

[3] D. Ron and A. Shamir, "Quantitative Analysis of the Full Bitcoin Transaction Graph", *Proceedings of International Conference on Financial Cryptography and Data Security*, pp. 6-24, 2013.

[4] F. Reid and M. Harrigan, "An Analysis of Anonymity in the Bitcoin System", *Proceedings of IEEE International Conference on Privacy, Security, Risk, and Trust*, pp. 1318-1326, 2013.

[5] R. Grinberg, "Bitcoin: An Innovative Alternative Digital Currency", *Hastings Science and Technology Law Journal*, Vol. 4, pp. 160-165, 2012.

[6] M. Levine, "Bitcoin Is an Expensive Way to Pay for Stuff", Available at: https://www.bloomberg.com/view/articles/2014-01-02/bitcoin-is-an-expensive-way-to-pay-for-stuff.

[7] D. Yermack, "Is Bitcoin a Real Currency?", *The Handbook of Digital Currency*, 2013.

[8] A. Badev and M. Chen, "Bitcoin: Technical Background and Data Analysis", Available at: https://www.federalreserve.gov/econresdata/feds/2014/files/2014104pap.pdf.

[9] M. Iwamura, Y. Kitamura and T. Matsumoto, "Is Bitcoin the Only Cryptocurrency in the Town? Economics of Cryptocurrency and Friedrich A. Hayek", Discussion Paper Series, Institute of Economic Research, Hitotsubashi University, pp. 1-13, 2014.

[10] C. Gunduz and B. Donald, "A Theoretical Foundation for Technical Analysis", *Journal of Technical Analysis*, Vol. 59, No. 5, pp. 1-22, 2003.

[11] L. Andrew, H. Mamaysky and J. Wang, "Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation", *Journal of Finance*, Vol. 55, No. 4, pp. 1705-1765, 2000.