Research Article

Prediction of Crypto Currency by Multiple Linear Regressions using Anaconda Software

[1]Anshey Singh, [2]Narendra Mohan,

[1]Anshey Singh, [2]Narendra Mohan,
Department of Computer Engineering and Application, GLA University, Mathura.
E-Mail: anshy.singh@gla.ac.in, narendra.mohan@gla.ac.in

ABSTRACT

Financial instruments traded in the stock markets attract various categories of investors across the world. The major objective of investment strategy is to minimize the risk and maximize the return and this can be achieved by choosing the right security at an appropriate time will result in loss reduction. The modern day investment alternative across the globe is crypto currency. A crypto currency is digital asset designed to work as a medium of exchange that uses strong cryptography to secure financial transactions, control the creation of additional units, and verify the transfer of assets. However, cryptocurrency are highly volatile due to their supply and demand factors as they are been designed to have a high degree of non availability. Hence, it is important to have a bitcoins price prediction model that can predict the bitcoins total supply value. In this paper, a multiple linear regression model for predicting the bitcoins total supply is proposed. The change in bitcoins total supply for definite time intervals like total supply per an hour, a day and a week were input for developing the linear regression model using the software package ANACONDA. This is significant improvement in comparison to traditional models. In the present scenario there is no such method of predicting the price of bitcoins. The proposed model can be used a decision making technique during the intense situations of bitcoins trading, where human predictions will not make appropriate price predictions. This paper comprises of various sections where the first section describes the introduction to Bitcoins, their importance followed by methodology and the strategy of algorithm applied. The third section describes the processing steps and implementation of algorithm and followed by validation of the project accuracy and efficiency checker.

Keywords: Bitcoin, Cryptography, Multiple Linear Regression, ANACONDA software package.

Introduction

Crypto currency uses cryptographically codes to execute the international financial transactions using internet as a standard of exchange [1]. This type of currencies mainly relies on block chain technology in order to have immutability, transparency and to gain the advantage of decentralization [2]. The major advantage with this type of currencies is they are controlled by any central authority or governments. Crypto currencies like bit coins are transferred directly between two parties using private and public keys. Bitcoins were firstly invented in the year 2008 as a peer- to- peer electronic payment system with an objective to prevent double spending. This financial product has become very popularized as it can be exchanged with other currencies issued by the central banks across the globe, products and services [3]. Due to this the price of the bitcoin is greatly affected by demand and supply factors. Apart from this certain global factor has direct relation on the price of bitcoin [4][6]. Demonetization action taken by the Government of India resulted in the high price volatility in the prices of cryptocurrencies and this is 20 percent on higher side. During recent time there is great volatility in the values of bitcoins, which has resulted in the huge losses to the investors across the globe [7]. The below table depicts the price volatility of the bitcoin during last one-year time period as there is high price fluctuations and there is much need to predict the price movements in order to minimize the loss to the investors [5][8].

Figure 1: Usage of crypto currency at digital marketing

METHODOLOGY

Data relating to price of bitcoins is collected from different websites. The data that is collected contains the attributes that are required for the prediction of desired output. Here the collected data contains percent change in hour of bitcoins total supply, its change in a day and a week. These inputs are used by applying different algorithms so that the desired output

is predicted. This output is nothing but the total supply of bitcoins[9]. This is taken as inputs for building the model that predicts the output. The dataset before applying any algorithm for the design of model should be pre-processed. In the next section we shall give short overview of the linear regression method and after that the predictive model will be presented, using ANACONDA software.

ARCHITECTURAL DIAGRAM

Architecture is nothing like a blue print for a process or a model. In the similar way every ML model will have an architecture that is to be followed the main steps remains same but the choosing of the dataset and the algorithm will be varied. The ML model that is to build follows some sequence of steps that is depicted as architecture. All these steps are to be followed for every dataset that is taken to build an ML model.

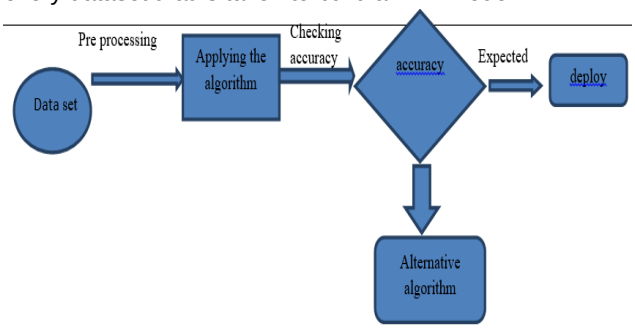


Figure 2: Architecture of the procedure of ML model

RESEARCH

Linear Regression

Regression is the process of estimating the relation between different variables. This type of estimation is used for the prediction of desired output. The regression method in Machine Learning is used for the computation of dependent variable from independent variable. It is based on the parameters. This method models the dependent variable with the help of independent variables. Here the independent variables may be many and the dependent variable is the unknown value that is to be predicted [10]. The function Y is called regression function. Every value of the independent variable x is associated with a value of the ss actual value of target variable and its predicted value, and this difference is the error of the estimation which is called “deviation” or “residual” [11].

Multiple linear regression

In the process of building a Machine Learning model for the prediction of bitcoins value an algorithm is to be applied [12]. The choosing of the machine learning algorithms will be dependent upon the variables present, the parameters assumed and also the number of independent variables present. Here in this model there are multiple independent variables and only single dependent variable is present which is obtained as output [13]. So based on these constraints Multiple Linear Regression algorithms is chosen to apply. Anaconda software is used in the building of this model which contains tools like Jupiter and spyder for building of the model [14]. In the process of processing the dataset the language used is python. In this project Jupiter is used. To build a ML model there are different steps to be followed,

1. Data collection
2. Data Wrangling
3. Data Cleaning
4. Data Analysis
5. Training an algorithm
6. Testing the Algorithm
7. Deployment

These are seven steps that are to be followed in the building of a machine learning model.

a. Data Collection

Here the dataset is collected from different sources it may be internet or direct sources**.**

b. Data Wrangling

In this step the data that is collected is discovered with its form. If the data is unstructured then that is made structured.

c. Data Cleaning

In the cleaning process the dataset is checked for any noisy or irrelevant data. If it is present such type of columns or rows are deleted. This is done to increase the efficiency of the model [15].

d. Data Analysis

The analysis is done for the variables that are assumed in the dataset. Based on these variables the type of algorithm is chosen. So from the analysing of the dataset the algorithm to be applied is determined.

e. Training of the algorithm

In this step the dataset is trained with the chosen algorithm. Here the algorithm chosen is Multiple Linear Regression So, the dataset that is collected is trained with the MLR functions and the accuracy of the model is found.

f. Testing the data

The trained data is to be tested for its working. So to test the model the dataset is divided into training data and testing data.The percentage of the training data taken is 80% of the dataset and the remaining 20% is made as testing data [16]. The dependent and the independent variables are portioned in the same proportion. Later the algorithm is applied and the obtained total supply value is compared with the actual value. The accuracy obtained is by finding the value of R2 score.The accuracy is high and so the model is accurate for the estimation of bitcoin total supply values.

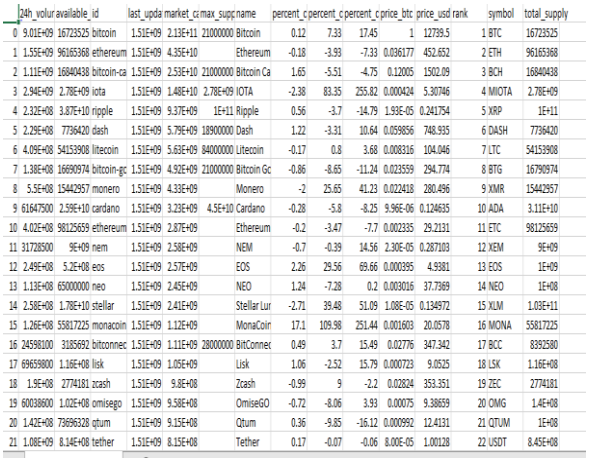
g. Deployment

The developed model is deployed for the use of clients. So a user interface is created which takes input as percent change _per 1h, percent change _per 24h, percent change per 7d.These values are processed at the backend with the model created and the total supply of the bitcoins is predicted.

IMPLEMETATION

Collecting the dataset

The dataset that is collected which has 14 columns and about 3000 rows. This dataset contains number of columns which are the parameters that cause changes in the values of bitcoins supply. But there may be some attributes that doesn't result in any change of the output to be predicted. The dataset that is collected that is searched from various sources is to be imported into the Jupiter platform.



	24h_volu	available_id		last_upda	market_c	max_supp	name	percent_c	percent_c	percent_c	price_btc	price_usd	rank	symbol	total_supply
0	9.01E+09	16723525	bitcoin	1.51E+09	2.13E+11	21000000	Bitcoin	0.12	7.33	17.45	1	12739.5	1	BTC	16723525
1	1.55E+09	96165368	ethereum	1.51E+09	4.35E+10		Ethereum	-0.18	-3.93	-7.33	0.036177	452.652	2	ETH	96165368
2	1.11E+09	16840438	bitcoin-ca	1.51E+09	2.53E+10	21000000	Bitcoin Ca	1.65	-5.51	-4.75	0.12005	1502.09	3	BCH	16840438
3	2.94E+09	2.78E+09	iota	1.51E+09	1.48E+10	2.78E+09	IOTA	-2.38	83.35	255.82	0.000424	5.30746	4	MIOTA	2.78E+09
4	2.32E+08	3.87E+10	ripple	1.51E+09	9.37E+09	1E+11	Ripple	0.56	-3.7	-14.79	1.93E-05	0.241754	5	XRP	1E+11
5	2.29E+08	7736420	dash	1.51E+09	5.79E+09	18900000	Dash	1.22	-3.31	10.64	0.059856	748.935	6	DASH	7736420
6	4.09E+08	54153908	litecoin	1.51E+09	5.63E+09	84000000	Litecoin	-0.17	0.8	3.68	0.008316	104.046	7	LTC	54153908
7	1.38E+08	16690974	bitcoin-gc	1.51E+09	4.92E+09	21000000	Bitcoin Gc	-0.86	-8.65	-11.24	0.023559	294.774	8	BTG	16790974
8	5.5E+08	15442957	monero	1.51E+09	4.33E+09		Monero	-2	25.65	41.23	0.022418	280.496	9	XMR	15442957
9	61647500	2.59E+10	cardano	1.51E+09	3.23E+09	4.5E+10	Cardano	-0.28	-5.8	-8.25	9.96E-06	0.124635	10	ADA	3.11E+10
10	4.02E+08	98125659	ethereum	1.51E+09	2.87E+09		Ethereum	-0.2	-3.47	-7.7	0.002335	29.2131	11	ETC	98125659
11	31728500	9E+09	nem	1.51E+09	2.58E+09		NEM	-0.7	-0.39	14.56	2.30E-05	0.287103	12	XEM	9E+09
12	2.49E+08	5.2E+08	eos	1.51E+09	2.57E+09		EOS	2.26	29.56	69.66	0.000395	4.9381	13	EOS	1E+09
13	1.13E+08	65000000	neo	1.51E+09	2.45E+09		NEO	1.24	-7.28	0.2	0.003016	37.7369	14	NEO	1E+08
14	2.58E+08	1.78E+10	stellar	1.51E+09	2.41E+09		Stellar Lur	-2.71	39.48	51.09	1.08E-05	0.134972	15	XLM	1.03E+11
15	1.26E+08	55817225	monacoin	1.51E+09	1.12E+09		MonaCoir	17.1	109.98	251.44	0.001603	20.0578	16	MONA	55817225
16	24598100	3185692	bitconnec	1.51E+09	1.11E+09	28000000	BitConnec	0.49	3.7	15.49	0.02776	347.342	17	BCC	8392580
17	69659800	1.16E+08	lisk	1.51E+09	1.05E+09		Lisk	1.06	-2.52	15.79	0.000723	9.0525	18	LSK	1.16E+08
18	1.9E+08	2774181	zcash	1.51E+09	9.8E+08		Zcash	-0.99	9	-2.2	0.02824	353.351	19	ZEC	2774181
19	60038600	1.02E+08	omisego	1.51E+09	9.58E+08		OmiseGO	-0.72	-8.06	3.93	0.00075	9.38659	20	OMG	1.4E+08
20	1.42E+08	73696328	qtum	1.51E+09	9.15E+08		Qtum	0.36	-9.85	-16.12	0.000992	12.4131	21	QTUM	1E+08
21	1.08E+09	8.14E+08	tether	1.51E+09	8.15E+08		Tether	0.17	-0.07	-0.06	8.00E-05	1.00128	22	USDT	8.45E+08

Figure 3: Dataset used in crypto currency prediction

Importing the dataset

The dataset is imported and then searched for noisy or irrelevant content.

Unnamed	Name	Percent_change_1h	Percent_change_24h	Percent change_7d
0	Bitcoin	0.12	7.33	17.45
1	Ethereum	-0.18	-3.93	-7.33
2	Bitcoin Cash	1.65	-5.51	-4.75
3	IOTA	-2.38	83.35	255.82
4	Ripple	0.56	-3.70	-14.79
5	Dash	1.22	-3.31	10.64

Table 1: Representing the required input columns of the dataset

Total supply
1.673252e+07
9.616537e+07
1.684044e+07
2.779350e+09
9.993909e+10

Table 2: Represents the output column of the dataset

Data Wrangling

The attributes that are not necessary for the prediction of output are removed and only the required columns are selected for the further processing. Even in every row and column checking of null values is done using some python commands.

df.is null ().any()

This is the command used for checking of null values in the dataset. The presence on null values will be represented using Boolean values. If any column has null values in it then the Boolean value will be represented as True or the value will be False.

Here isnull ()is a method that checks for null values which is present in python.

df represents the object that is created for representing the imported dataset.

Column	Boolean value
24h_volume_usd	False
available supply	True
Id	False
last_updated	False
Market_cap_usd	True
Max_supply	True
Name	False
Percent_change_1h	True
Percent_change_24h	True
Percent_change_7d	True
Price_btc	False
Price_usd	False
Rank	False
Symbol	False
Total_supply	True

Table 3: Represents all the columns of dataset and their presence of null values

Column	Boolean value
Name	False
percent _change_1h	True
percent _change_24h	True
percent _change_7d	True
total_supply	False

Table 4: Representing the required columns of the dataset for processing

Since there are null values in some columns these values are to be replaced by some mean value so that there will not be null values. So these null values are replaced with mean values of that corresponding column by using methods in python

Data Analysis

Based on the dataset and analyzation of inputs and outputs present in it Multiple Linear Regression algorithm is to be applied to get the desired output.

Method →mlr.LinearRegression() is the method used for implementing this algorithm.

Fitting of the data →mlr.fit(x_train,y_train) .This is the command used for sending the train data into the regression function.

Testing of the data

The testing of the data is done by giving some input values from the dataset and checking if the predicted values are in actual dataset.

Here in this project of predicting the total supply value the predicted values are more nearer to the actual values and the efficiency of the project is also high.

The accuracy obtained is measured based on the R2score value and is 0.88.So it says that the model is efficient for the dataset and the algorithm is also apt.

Deployment

The model that is built for the prediction of total supply of bitcoins is to be made user friendly by creating a user interface. This is done using Node Red platform. This contains input fields and an output field with a button.This is used for the validation of data that is being processed.

EXPERIMENTAL RESULTS

The results of validation show that the model is accurate and efficient in predicting the bitcoin total supply values. The experimental analysis shows the pattern in which the input parameters are given. Here in this experiment the values of the bitcoin total supply with respect to different time intervals is given and value of total supply is predicted. The analysis of the experiment is done by giving different types of inputs and checking the dataset for the validation. Most of the input values given predicted the accurate values to that of the actual values. Validation of this experiment can be checked by comparing the predicted values with the actual values in the dataset. On frequent input parameters passed the model predicted the accurate results.

Actual data values present in the dataset
Percent_change_1h - 0.12
Percent_change_24h - 7.33
Percent_change_7d - 17.45
total_supply - 929812565975

Table 5: Represents the actual data values of input and output

Percent_change_1h - 0.12
Percent_change_24h - 7.33
Percent_change_7d - 17.45
total_supply - 919812565974

Table 6: Represents the predicted value of total supply of bitcoins

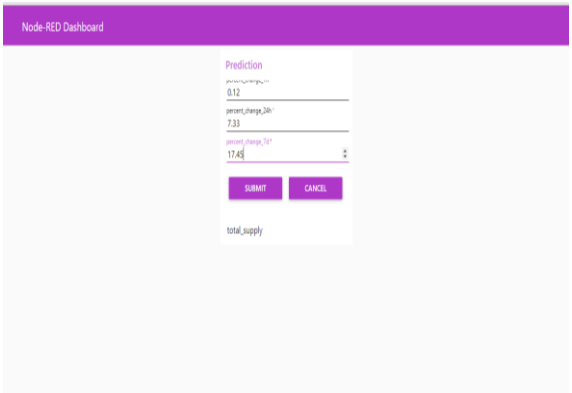


Figure 4: Inputs are given through graphical user interface

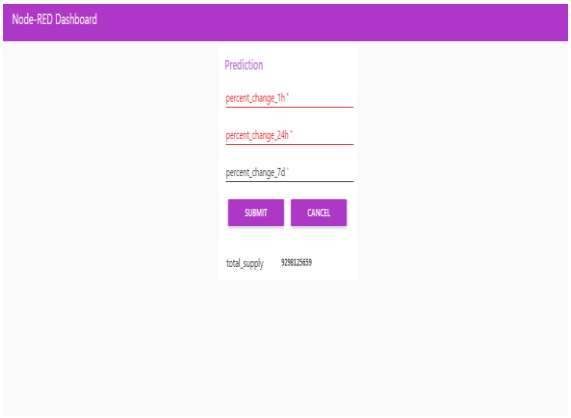


Figure 5: Output is predicted depends on input parameters

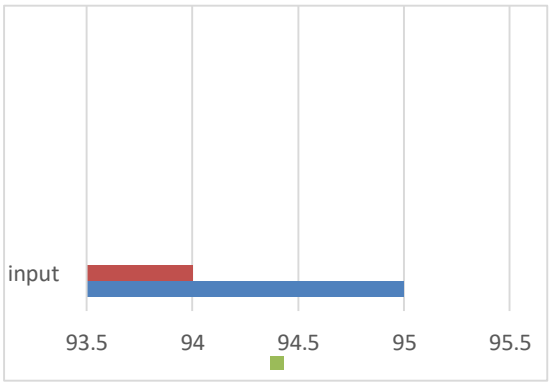


Figure 6: The analysis of the results obtained

Based on the obtained results it is clear that the model is efficient. The output obtained for each input value is approximately equal to the value in the dataset. The analysis shows that the predicted values and the actual values are shown in figure 4.3.

Conclusion

ANACONDA software package was used for creating the linear regression model for predicting the total supply of bitcoin is more efficient in estimating the demand and supply. The accuracy of the model was estimated from the statistics of validation data. R2 score is used as estimator of the model. Expecting that the presented model can be successfully used in the practice as very accurate, efficient and fast forecasting model for total supply of bitcoins prediction, especially in the stock markets. In directly this model is used for estimating the price of the bitcoins by the investors, So that if the price is known there will be a no chance of incurrence of loss for the investors during the investment.

References

1. Laura Alessandretti, 1 AbergElBahrawy, 2 Luca Maria Aiello, 3 and Andrea Baronchelli , Research Article Anticipating Crypto currency Prices Using Machine Learning 2,41 Technical University of Denmark, 2800 Kgs. Lyngby, Denmark 2 City, University of London, Department of Mathematics, London EC1V 0HB, UK 3 Nokia Bell Labs, Cambridge CB3 0FA, UK 4 UCL Centre for Block chain Technologies, University College London, UK
2. An Empirical Study on Modeling and Prediction of Bitcoin Prices With Bayesian Neural Networks Based on Blockchain Information HUISU JANG AND JAEWOOK LEE Department of Industrial Engineering, Seoul National University, Seoul 151742, South Korea
3. Automated Bitcoin Trading via Machine Learning Algorithms Isaac Madan Department of Computer Science Stanford University Stanford, CA 94305 imadan@stanford.edu
4. Bitcoin Price Prediction using Machine Learning Siddhi Velankar*, Sakshi Valecha*, Shreya Maji* *Department of Electronics & Telecommunication, Pune Institute of Computer Technology, Pune, Maharashtra, India
5. Price Fluctuations and the Use of Bitcoin: An Empirical Inquiry, The Open University's repository of research publications and other research outputs
6. 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.
7. C. Gunduz and B. Donald, "A Theoretical Foundation for Technical Analysis", Journal of Technical Analysis, Vol. 59, No. 5, pp. 1-22, 2003.
8. 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.
9. 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.
10. 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.
11. R. Grinberg, "Bitcoin: An Innovative Alternative Digital Currency", Hastings Science
12. Kumar, Manoj, and Ashish Sharma. "Mining of data stream using "DDenStream" clustering algorithm." 2013 IEEE International Conference in MOOC, Innovation and Technology in Education (MITE). IEEE, 2013.
13. Sharma, Ashish, Anant Ram, and Archit Bansal. "Feature Extraction Mining for Student Performance Analysis." Proceedings of ICETIT 2019. Springer, Cham, 2020. 785-797.

14. Sharma, Ashish, and Dhara Upadhyay. "VDBSCAN Clustering with Map-Reduce Technique." Recent Findings in Intelligent Computing Techniques. Springer, Singapore, 2018. 305-314.
15. Sharma, Ashish, Ashish Sharma, and Anand Singh Jalal. "Distance-based facility location problem for fuzzy demand with simultaneous opening of two facilities." International Journal of Computing Science and Mathematics 9.6 (2018): 590-601.
16. Varun K L Srivastava, Dr. Anubha Shrivastava, N. Chandra Sekhar Reddy, " A Study on Maintainability and Availability Parameters using Code Metrics ", J.Mech.Cont.& Math. Sci., Vol.-14, No.2, March-April (2019) pp 100-111.