iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** XII **Month of publication:** December 2022

DOI: https://doi.org/10.22214/ijraset.2022.48286

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue XII Dec 2022- Available at www.ijraset.com

Cryptocurrency Price Prediction Using Linear Regression and Long Short-Term Memory (LSTM)

Atharva Dhande[1], Shoumyadeep Dhani[2], Shivang Parnami[3], K. P. Vijayakumar[4]
School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, India

Abstract: ***Predicting future events is difficult, particularly with regards to cryptocurrency, where the media, influential people and governments have a sharp and vital impact on worth. Cryptocurrency market analysis is a method through which the real-world data of the cryptocurrency market is used to predict where it will go next. If foretold accurately, it helps investors to invest when the value is low (purchasing in bulk when the price is dipping) and sell once it's high so as to gain a profit. This research provides two machine learning algorithms which are Long Short-Term Memory (LSTM) and Linear Regression for predicting the values of six different types of crypto currencies such as Bitcoin (BTC), Dash coin (DASH), Lite coin (LTC), Dogecoin (DOGE), Ethereum (ETH), and Monero (XMR). The accuracy of the models is analyzed using mean squared error.***
Keywords: Future; Market; Price; Prediction; Profit

I. INTRODUCTION

Today, all economies have embraced particular currencies (money) as a means of exchange. The money supply generates inflation and deflation in economies due to its excess supply and contraction, governments manage currencies in order to counteract inflation and deflation. Many governments throughout the world are focused on digital currencies and transactions these days. Also, majority of the people in the world don't want their transactions to be regulated by the government. This resulted in more innovation in a new currency, crypto currency, which is one of the most sophisticated, ambiguous, and regulation-free currencies. Transactions are rapid, digital, safe, and international, which essentially allows the preservation of records without the fear of data being pirated, as some may imagine, thus reducing fraud to its minimum.

The first step will be collecting the real-world information from the cryptocurrency market and plotting it the data to analyze the trend and predict whether it will be bullish or bearish. If foretold properly, this enables to invest when the value is low (buying on the dip) and sell once it's high in order to gain profit. Technical analysis of a cryptocurrency helps to read the market. It involves observation and analysis of price charts and graphs from various perspectives and finding a consensus within that information to help to predict where the market is going. Market prediction is done using machine learning techniques namely Long Short-Term Memory (LSTM) and Linear regression (LR), etc. These approaches can mimic the simultaneous dynamic interaction of several components, allowing for the study of complexity; they may also derive conclusions on an individual basis rather than as average trends.

In the digital market, there are hundreds of crypto currencies, but Bitcoin is the most prominent, and it is influenced by a lot of factors such as the news and social media. Bitcoin's usage of open-source code and a censorship-resistant architecture has led it to become the main source of reference for many cryptocurrencies and their developers. Many crypto currencies have gained importance other than bitcoin. Dogecoin, for example, was a meme-based joke coin that was popularized when CEO of Tesla Elon Musk promoted the crypto currency on social media. Other examples are Ethereum, Solana, monero, avalanche etc. In this paper, six well known cryptocurrencies are used to predict bitcoin and Ethereum as these are two of the most popular and largest cryptocurrencies in terms of volume. Also, monero, dash coin, lite coin is also used as these are the coins which were the first to enter the market are comparatively easier to predict than the newer coins and alt coins. Last but not the least, for a little challenge Doge coin is also used. Because the coin launched as a satire on the cryptocurrency space, and it is highly influenced by Elon musk tweets.

The aim of this paper is to predict the future prices of the Bitcoin, Ethereum. Dogecoin, Monero, Dashcoin and Litecoin with the help of Long Short-Term Memory and Linear Regression models and evaluate the results using mean squared error.

The motivation behind this paper is to facilitate cryptocurrency investors to invest at appropriate time by predicting future prices for six cryptocurrencies, thus improving their portfolio. Also, this paper compares the accuracy of the output of the LSTM and Linear Regression models, helping the reader to appropriately select model for prediction.

©IJRASET: All Rights are Reserved

The paper is organized into various sections as follows: Section 2 describes about the literature review. System model is explained in section 3. The proposed system is illustrated in section 4. Section 5 discusses about the dataset used in this paper. Result and evaluation of the proposed system is provided in section 6. Section 7 summarizes the paper with a conclusion and discusses about future work.

II. RELATED WORK

Paper [1] compares the performances of different Machine Learning algorithms such as SVM, Boosted NN, ANN and DL for a wide variety of cryptocurrencies for their forecasting and also talks about different time series data of cryptocurrency in detail. The paper [2] shows how incorporation of a cryptocurrency into a portfolio improves profit gain by: (i) reducing the standard deviation and, (ii)Use asset portfolio allocation for different cryptocurrencies. It also talks about the performances of cryptocurrencies with respect to stocks and market their returns as well. Paper [3] implements a combined ensemble of Random Forest and Stochastic gradient-based model. The ensemble is then used on a variety of coins such as bitcoin, ripple and Ethereum.

The paper [4] presents a hybrid form of cryptocurrency price prediction system based on Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU), mainly focused Litecoin and Monero cryptocurrencies. The paper [6] shows the implementation of traditional SVMs and linear regression methods to predict the price of Bitcoin. This research considers a forecast made up of closing price of Bitcoin every day for the creation of prediction models. The paper [7] implements algorithms such as linear regression, gradient descent, random forest and a classified deep learning algorithm for the price prediction of bitcoin.

Paper [8] analyses the price fluctuations of Bitcoin, Ethereum, and Ripple. The authors utilize multiple neural network frameworks such as ANN and LSTM. The authors found out that ANN relies more on the future history, whereas LSTM relies more on short-term data, The paper [9] uses a slew of different algorithms such as LSTM, SVM, random forest, XGBoost, Linear discriminant analysis (LDA), Quadratic discriminant analysis (QDA) for predicting the price of bitcoin and judges them on the parameters of precision and accuracy with LDA having the topmost accuracy of 66%.

The paper [10] studies the implementation of random forest (RF), neural networks (NN), and support vector machines (SVM). It is also found out that machine learning and sentiment analysis can also be used to know future of cryptocurrency markets and that neural network was found to be the best among the models mentioned previously. The paper [11] implements long short-term memory (LSTM) to predict and find ways to forecast price of Bitcoin on the stock market through Yahoo Finance. Thus, after the review of all the papers, analysts are still trying to find out the perfect algorithms which are suitable for forecasting by testing out new algorithms with robust mechanisms and modifying the old ones.

III. SYSTEM MODEL

A. Long Short-Term Memory

Long Short-Term Memory, also popularly known as LSTM is machine learning model based on Neural Network. It is an improvised version of Recurrent Neural Network. LSTM cell is pretty similar to RNN cell having three parts: Forget Gate, Input Gate and Output Gate, (shown in figure 1).

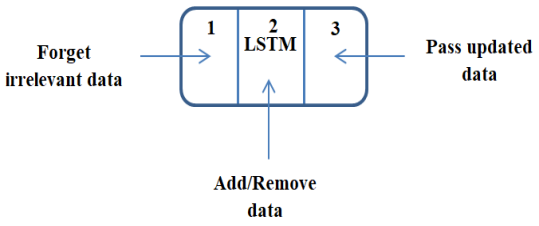


Fig. 1 The three parts of the LSTM cell

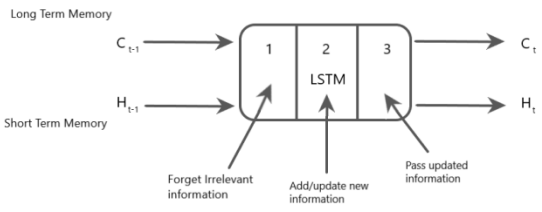


Fig. 2 Shows the short term and long-term memory

We can observe the construction of an LSTM cell from figure 2. The Forget Gate helps in removing useless information from the LSTM cell. The information is removed by multiplication of a filter. The Input Gate takes the responsibility of adding information to the LSTM cell through three steps : (i) Value regulation using sigmoid function, (ii) Vector creation using tanh function, and (iii) Multipliacation of the created vector with the regulatory function. The useful information from the current state is shown with the help of the output gate.

1) Working of LSTM

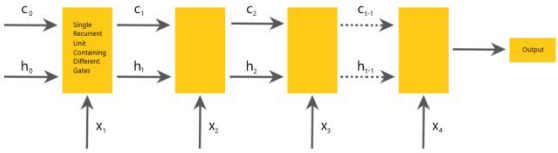


Fig. 3 Shows the workflow of LSTM

Recurrent Unit Working Principle (a diagrammatic representation of the working is shown in the figure 3):

a) Step1: Get the following inputs: Current Input, previous hidden state and previous internal cell state.
b) Step 2: The calculation of the values of the gates is given by: (i) Calculation of parameterized vectors for the previous hidden state and current input, (ii) application of respective activation function for every gate.
c) Step 3: Calculation of current internal cell state.
d) Step 4: Calculation of hidden state.

B. Linear Regression

The linear regression is one of the most fundamental and widely used algorithm for forcasting and prediction. It is used for modelling a relationship between two or more variables with the help of a linear equation. The researcher often attempts to comprehend or relate at least two independent (predictor) variables with a dependent variable to see the result. Both correlation and regression give this chance to comprehend the "risk factors-illness" relationship. While correlation gives a quantitative approach to estimating the degree or strength of a relation between two variables, regression analysis numerically depicts this relationship. The linear regression is represented by the following equation:- y = mx + c.

The dependent variable must be continuous whereas the independent variable may or may not be continuos. Generally, the relationship between continuous variables is represented by scatter plot. This sort of plot will show whether the relationship is linear or nonlinear as shown in Figures 4 and 5 respectively.

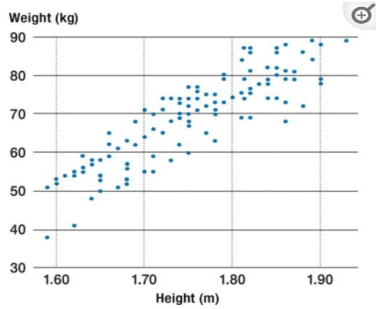


Fig. 4 A Scatter plot

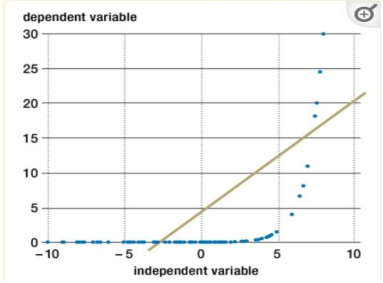


Fig. 5 A scatter plot showing an exponential relationship

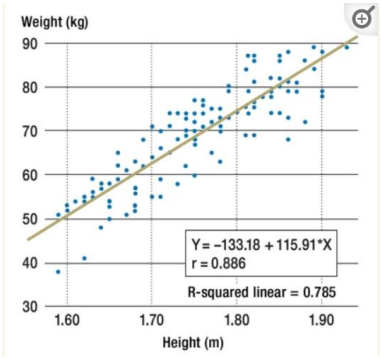


Fig. 6 A scatter plot showing the corresponding regression line and regression equation between the dependent variable (body weight in kg) and the independent variable (height in m).

In the figure 5, a univariable linear regression depicts the linear relationship between a single independent variable X and a dependent variable Y. The line of regression allows a person to predict the value of the dependent variable Y from the value of the independent variable X as shown in Figure 6.

IV. PROPOSED SYSTEM

Long Short-Term Memory and Linear Regression will be used to predict the future prices of the cryptocurrencies. Both the machine learning models are a series of algorithms that are used to forecast data, which are trained using time series data.

A. *Long Short-Term Memory*

In the proposed system using Long Short-Term Memory, following steps are carried out:

1) *Feature Selection:* From the dataset, only two features are considered for prediction using LSTM: Date and Closing price.
2) *Train-Test split:* After attribute filtering, the data is divided into training data and testing data with percentage records of 80% and 20% respectively.
3) *Formatting of Training Data:* The training data is formatted to the shape of window size of 5 and 2 features.
4) *Building LSTM Model:* The formatted data is used to build lstm model along with dropout as 0.2, density as 1, number of neurons as 100, active function as 'linear', loss as 'mse' and optimiser as 'adam'.
5) *Training the LSTM Model:* After building the model, it is trained with hyper parameters as follows: number of epochs as 20, batch size as 32, verbose as 1 and shuffling condition as true.
6) *Prediction and Error Analysis:* After training the model, the future data is predicted with the help of testing data and the error (mean squared error) is calculated between predicted data and test data.
7) *Plotting:* After prediction, the predicted data and testing data both are plotted on the same graph.
8) *Repeating for all Dataset:* The steps from 1 to 6 are repeated for every dataset in a loop.

B. *Linear Regression*

In the proposed system using Linear Regression, following steps are carried out:

1) *Feature Selection:* From the dataset, only two features are considered for prediction using LSTM: Date and Closing price.
2) *Date Formatting:* The date in the dataset is converted into timestamp as normal date notation don't work with Linear Regression algorithms.
3) *Train-Test Split:* Initially, the input data is split into training data and testing data with percentage records of 80% and 20% respectively.
4) *Building and Training Linear Regression Model:* After the data splitting, Linear Regression model is built with no extra hyper parameters. Then the model is trained using the training dataset.
5) *Prediction and Error Analysis:* After training the model, the future data is predicted with the help of testing data and the error (mean squared error) is calculated between predicted data and test data.
6) *Plotting:* After prediction, the predicted data and testing data both are plotted on the same graph.
7) *Repeating for all Dataset:* The steps from 1 to 6 are repeated for every dataset in a loop.

V. DATASET

The dataset was collected from [12]. The original dataset consisted 6 sheets of data (1 sheet for every coin), each sheet consisting of 384 records; for simplicity, every sheet was converted into a '.csv' (Comma Separated Value) file. Sample data of every cryptocurrency is shown in Figures 7-12. Table 1 shows the description of dataset.

TABLE I
DATASET SPECIFICATION

Feature Name	Feature Description	Feature Type
Unix Timestamp	Timestamp of the record generated in Unix format.	Number
Date	Date of the record generated.	Date
Symbol	Symbolic name of the coin.	String
Open	It refers to the price at 12:01 AM UTC of any given day for the quoted cryptocurrency.	Number
High	It refers to the highest price reached during the last 24 hours for the quoted cryptocurrency.	Number
Low	It refers to the Lowest price reached during the last 24 hours for the quoted cryptocurrency.	Number
Close	It refers to the price at 11:59 PM UTC of any given day for the quoted cryptocurrency.	Number
Volume (coin name)	Total quantity of a traded asset (disclosed in cryptocurrency)	Number
Volume USD	Total quantity of a traded asset (disclosed in US Dollar currency)	Number

	Unix Timestamp	Date	Symbol	Open	High	Low	Close	Volume BTC	Volume USD
0	1528675200	11/06/2018	BTCUSD	6806.12	6830.49	6727.39	6811.90	2591.46	17587448.97
1	1528588800	10/06/2018	BTCUSD	7496.81	7499.01	6732.03	6806.12	8199.93	58293789.94
2	1528502400	09/06/2018	BTCUSD	7620.77	7679.81	7472.82	7496.81	7288.46	55592105.24
3	1528416000	08/06/2018	BTCUSD	7708.73	7712.03	7543.91	7620.77	7176.63	54729919.27
4	1528329600	07/06/2018	BTCUSD	7660.93	7755.61	7660.03	7708.73	6832.76	52680059.51

Figure 7: Bitcoin sample data

	Unix Timestamp	Date	Symbol	Open	High	Low	Close	Volume DAS	Volume HUSD
0	1528675200	11/06/2018	DASHUSD	274.46	275.95	273.25	273.25	15.12	4160.83
1	1528502400	09/06/2018	DASHUSD	310.17	312.32	296.87	300.40	3593.55	1088510.33
2	1528416000	08/06/2018	DASHUSD	315.96	317.37	302.91	310.17	2892.42	893067.33
3	1528329600	07/06/2018	DASHUSD	315.30	319.63	313.46	315.96	10380.94	3280647.24
4	1528243200	06/06/2018	DASHUSD	318.97	320.03	308.00	315.30	9700.40	3050522.96

Figure 8: Dash coin sample data

	Unix Timestamp	Date	Symbol	Open	High	Low	Close	Volume LTC	Volume USD
0	1528675200	11/06/2018	LTCUSD	107.11	108.92	104.80	107.13	14274.0	1526468.78
1	1528588800	10/06/2018	LTCUSD	117.55	117.55	103.91	107.11	60408.4	6726829.36
2	1528502400	09/06/2018	LTCUSD	119.79	120.87	117.13	117.55	41443.2	4957296.18
3	1528416000	08/06/2018	LTCUSD	121.73	122.28	118.39	119.79	33688.1	4029067.06
4	1528329600	07/06/2018	LTCUSD	121.77	123.69	119.81	121.73	42450.5	5167380.41

Figure 9: Litecoin sample data

	Unix Timestamp	Date	Symbol	Open	High	Low	Close	Volume XMR	Volume USD
0	1528675200	11/06/2018	XMRUSD	140.28	141.00	135.17	137.56	6782.78	933012.27
1	1528588800	10/06/2018	XMRUSD	154.65	154.87	133.06	140.28	29018.28	4157984.43
2	1528502400	09/06/2018	XMRUSD	158.25	160.01	154.58	154.65	20684.70	3270066.43
3	1528416000	08/06/2018	XMRUSD	164.23	165.33	157.20	158.25	20580.41	3304829.10
4	1528329600	07/06/2018	XMRUSD	164.25	171.00	163.02	164.23	24194.08	4044497.29

Figure 10: Monero coin sample data

	Unix Timestamp	Date	Symbol	Open	High	Low	Close	Volume DOG	Volume EUSD
0	1528675200	11/06/2018	DOGEUSD	0.003102	0.003185	0.003074	0.003161	2061200	6413.43
1	1528502400	09/06/2018	DOGEUSD	0.003656	0.003700	0.003614	0.003627	71438190	261809.19
2	1528416000	08/06/2018	DOGEUSD	0.003726	0.003726	0.003450	0.003656	75571620	276957.36
3	1528329600	07/06/2018	DOGEUSD	0.003743	0.003856	0.003692	0.003726	68292630	257091.01
4	1528243200	06/06/2018	DOGEUSD	0.003641	0.003801	0.003600	0.003743	90495280	335175.68

Figure 11: Doge coin sample data

	Unix Timestamp	Date	Symbol	Open	High	Low	Close	Volume ETH	Volume USD
0	1528675200	11/06/2018	ETHUSD	532.83	536.20	529.82	533.32	282.12	150557.19
1	1528502400	09/06/2018	ETHUSD	599.77	608.81	597.20	598.85	6931.55	4183800.77
2	1528416000	08/06/2018	ETHUSD	606.28	609.50	593.61	599.77	10009.33	6009559.90
3	1528329600	07/06/2018	ETHUSD	606.40	616.48	601.76	606.28	11076.43	6741711.37
4	1528243200	06/06/2018	ETHUSD	610.02	611.48	594.70	606.40	11664.34	7048536.88

Figure 12: Ethereum coin sample data

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

Table 2 shows the training and testing split up of dataset considered in the model.

TABLE II
TRAIN-TEST SPLIT OF DATA

Model	Amount of Training data	Amount of Testing data
LSTM	80% (307 records for every cryptocurrency)	20% (76 records for every cryptocurrency)
Linear Regression	80% (307 records for every cryptocurrency)	20% (76 records for every cryptocurrency)

VI. RESULTS

A. *Output using LSTM*

Figure 13 shows the closing price of Bitcoin. In the start it increases but at around middle, it takes major drop in price as bitcoin and almost all cryptocurrencies fell due to China's crackdown on the cryptocurrency and Governments around the world proposing to ban cryptocurrencies. Eventually bitcoin and other coins will skyrocket in price due to the upcoming metaverse.

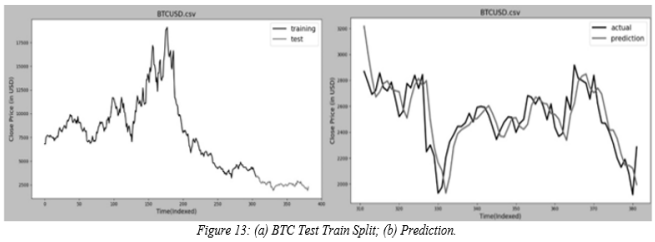


Figure 13: (a) BTC Test Train Split; (b) Prediction.

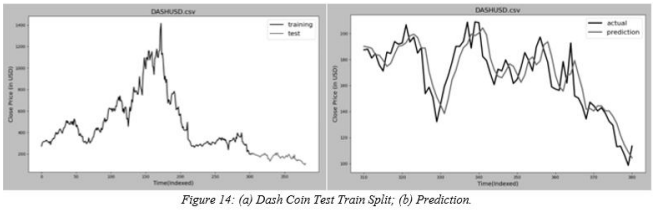


Figure 14: (a) Dash Coin Test Train Split; (b) Prediction.

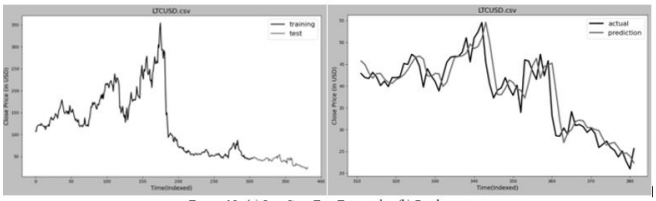


Figure 15: (a) Lite Coin Test Train split; (b) Prediction

Figures 14 and 15 show closing prices of DASH coin and Lite coin. These two are heavily influenced by BTC so the rise and fall of these two are very similar to BTC. DASH and LTC are also predicted to skyrocket in some months due to introduction of metaverse.

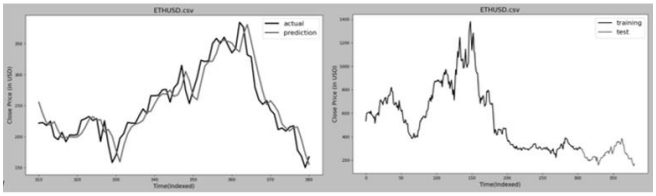


Figure 16: (a) Prediction; (b) Ethereum Test Train Split.

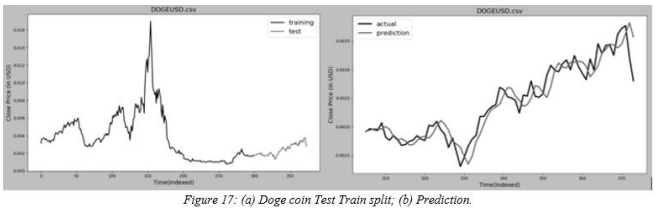


Figure 17: (a) Doge coin Test Train split; (b) Prediction.

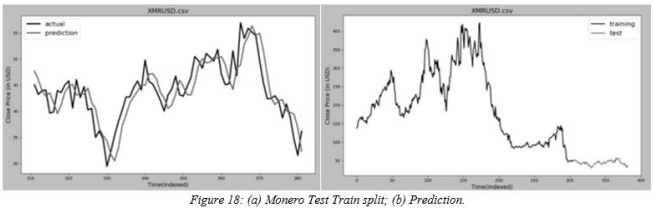


Figure 18: (a) Monero Test Train split; (b) Prediction.

Figure 16 shows closing price of the Ethereum coin within the targeted collected dataset. As ETH is the second biggest cryptocurrency in terms of volume so the trend of ETH is really similar to BTC. Figure 17 shows the closing price of the DOGE coin within the targeted collected dataset. It his highly influenced by social media mainly Elon musk tweets. So, the prediction of DOGE is really inconsistent. In 2021, Elon musk tweeted "Doge" which made the price of DOGE skyrocket. Figure 18 shows XMR closing price within the targeted collected dataset. Similar to LTC and DASH, XMR is also influenced by BTC and is showing similar trends to BTC.

B. Output using Linear Regression

Figures 19-21 show the output of linear regression applied on the dataset. A single linear regression line is plotted on graph for every cryptocurrency, from which the trend of data can be analysed and future closing prices of respective cryptocurrency can be predicted.

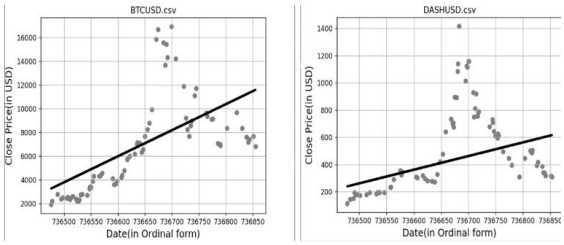


Figure 19: (a) Bitcoin Prediction; (b) Dash coin Prediction.

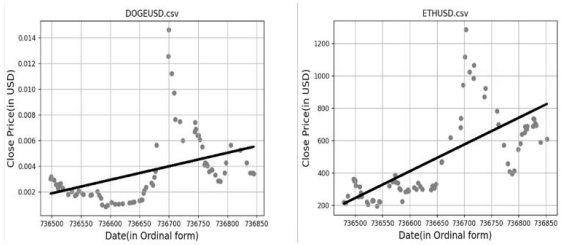


Figure 20: (a) Doge Coin Prediction; (b) Ethereum Prediction

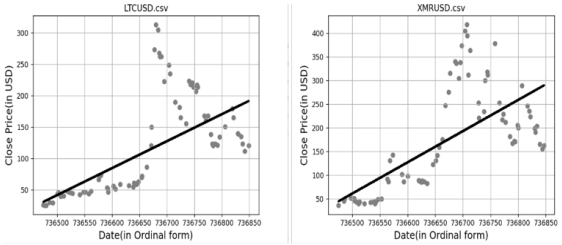


Figure 21: (a) Litecoin Prediction; (b) Monero Prediction.

From the evaluation metrics of LSTM and LR models, it is observed that LSTM is much more efficient in predicting the trend of each crypto.

The Mean Square Error (MSE) of each coin is given in Table 3, for most of them the MSE is quite average but for BTC it is really high due to high volatile and independent nature of BTC and for DOGE it's really low as there are not many external factors influencing DOGE that could predict. And also, it shows that the coefficient of determination is ultra-low because of linear regression not being as efficient as LSTM.

TABLE III
EVALUATION METRIC FOR ALGORITHMS

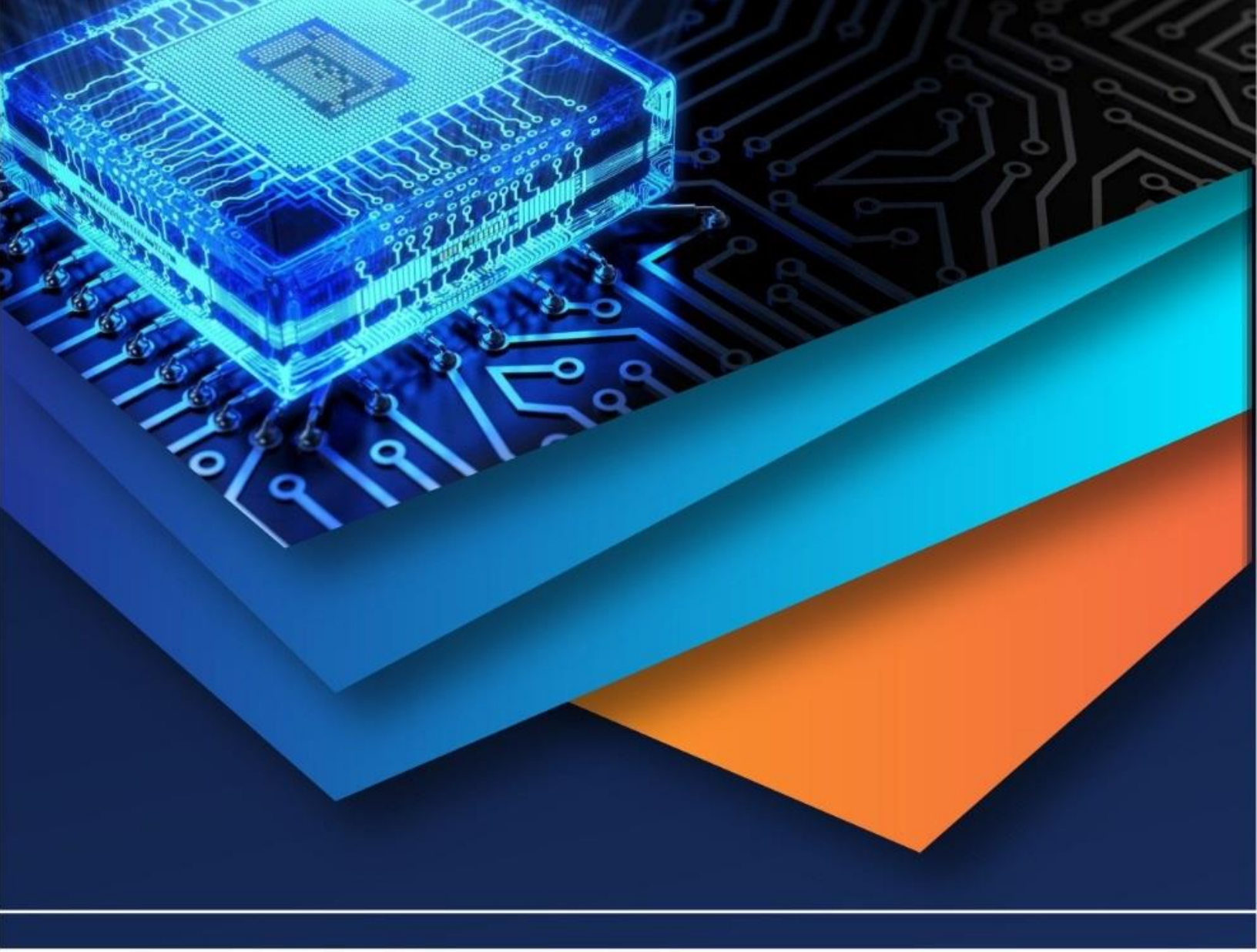
Coins	MSE Linear Regression	MSE for Long Short-Term Memory
Bitcoin (BTC)	9052114.61	0.003315
Dash coin (DASH)	87153.95	0.005315
Doge coin (DOGE)	0.00	0.007844
Ethereum coin (ETH)	34902.86	0.007146
Lite coin (LTC)	3711.05	0.007742
Monero coin (XMR)	6770.64	0.005358

VII. CONCLUSION

Prediction helps in making important economic decision cautiously and preventing any economic disaster due to miscalculation. Six cryptocurrencies are considered in this work by applying LSTM and Linear regression for predicting the trend. In future work, it is planned develop a model using various machine learning algorithms that will use huge amount of data and incorporate more cryptocurrencies for testing the accuracy of the model. In addition to this, attempt to build an application programming interface (API) to pipeline the real-time data to the model that can get real time predictions for cryptocurrency trades.

REFERENCES

[1] N. A. Hitam and A. R. Ismail, "Comparative per-formance of machine learning algorithms for cryptocurrency forecasting," Indonesian Journal of Electrical Engineering and Computer Science, vol. 11, no. 3, p. 1121, 2018.
[2] Y. Andrianto, "The effect of cryptocurrency on investment portfolio effectiveness," Jour-nal of Finance and Accounting, vol. 5, no. 6, p. 229, 2017. "Comparative performance of machine learn-ing ensemble algorithms for forecasting cryp-tocurrency prices," International Journal of Engineering, vol. 34, no. 1, 2021.
[3] M. M. Patel, S. Tanwar, R. Gupta, and N. Kumar, "A deep learning-based cryptocurren-cy price prediction scheme for Financial Insti-tutions," Journal of Information Security and Applications, vol. 55, p. 102583, 2020.
[4] R. Miura, L. Pichl, and T. Kaizoji, "Artificial Neural Networks for realized volatility pre-diction in cryptocurrency time series," Ad-vances in Neural Networks – ISNN 2019, pp. 165–172, 2019.
[5] S. Karasu, A. Altan, Z. Sarac, and R. Hacioglu, "Prediction of bitcoin prices with machine learning methods using time series data," 2018 26th Signal Processing and Com-munications Applications Conference (SIU), 2018.
[6] M. Saad and A. Mohaisen, "Towards charac-terizing blockchain-based cryptocurrencies for highly-accurate predictions," IEEE INFOCOM 2018 - IEEE Conference on Com-puter Communications Workshops (INFOCOM WKSHPS), 2018.
[7] W. Yiying and Z. Yeze, "Cryptocurrency price analysis with artificial intelligence," 2019 5th International Conference on Infor-mation Management (ICIM), 2019.
[8] Z. Chen, C. Li, and W. Sun, "Bitcoin price prediction using machine learning: An ap-proach to sample dimension engineering," Journal of Computational and Applied Math-ematics, vol. 365, p. 112395, 2020.
[9] F. Valencia, A. Gómez-Espinosa, and B. Val-dés-Aguirre, "Price movement prediction of cryptocurrencies using sentiment analysis and machine learning," Entropy, vol. 21, no. 6, p. 589, 2019.
[10] F. Ferdiansyah, S. H. Othman, R. Zahilah Raja Md Radzi, D. Stiawan, Y. Sazaki, and U. Ependi, "A LSTM-method for bitcoin price prediction: A case study yahoo finance stock market," 2019 International Conference on Electrical Engineering and Computer Science (ICECOS), 2019.
[11] F. Ferdiansyah, S. H. Othman, R. Zahilah Raja Md Radzi, D. Stiawan, Y. Sazaki, and U. Ependi, "A LSTM-method for bitcoin price prediction: A case study yahoo finance stock market," 2019 International Conference on Electrical Engineering and Computer Science (ICECOS), 2019.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)