IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
IJARSCT
Impact Factor: 6.252
Volume 2, Issue 6, May 2022

Crypto Currency Price Prediction Using LSTM

Karan Kumar R[1], Karthik Raja R[2], Kavimukilan M[3], Dr. B Muthu Senthil[4]

UG Scholar, Department of Computer Science & Engineering[1,2,3]
Associate Professor, Department of Computer Science and Engineering[4]
SRM Valliammai Engineering College, Chengalpattu, India

Abstract: *Crypto currencies have attracted extensive attraction from investors, researchers, regulators and the media. A well known and unusual feature is that crypto currencies price often fluctuates significantly, which has however received less attention. This project learns how to create model prediction of crypto currencies using LSTM. LSTM (Long Short-Term Memory) is another type of module provided for RNN (Recurrent Neural Network) later developed and popularised by many researchers. We investigate the impact of parameters on crypto currency price fluctuation prediction which can be further used in a real trading environment by investors. As a plan for future work, there is a recommendation to investigate other factors that might affect the prices of cryptocurrency market such as social media and tweets.*

Keywords: LSTM, RNN, Crypto Currency, Bitcoin, Predictions

I. INTRODUCTION

A cryptocurrency is a digital currency or a virtual currency that is secured by cryptography, which makes it more secured and popular. Many cryptocurrencies are noting but decentralized networks based on blockchain technology which is a distributed ledger enforced by a disparate network of computers. One of the true features of cryptocurrencies is that they are not generally issued by any central authority, rendering them theoretically immune to government interference or manipulation.

Here in our project, we have chosen Bitcoin to predict the future value of that crypto currency. Created in 2009, Bitcoin is a digital asset or currency that uses a P2P (Peer-to-Peer) network to facilitate the transfer of value without intermediation from banks or central authority.

Predicting the following day's price of bitcoin is impossible with normal calculations. There are lot of research work in process for developing models and algorithms for predicting the prices of such cryptocurrencies. In this study one such model is been developed using machine learning. The RNN plays the base role with the LSTM (Long Shot- Term Method) algorithm. The model is developed using Python programming language. Artificial Intelligence and Machine Learning has made the ease to predict the future. The act of intelligence is that knowing the future in today's world. There are many fields where predictions can be made like crypto currencies, stock markets, weather, business analysis, crime rate, flood prediction, etc.

II. RELATED WORKS

Chih-Hung Wu, Chih-Hung Lu, Ruei-Shan Lu, Yu-Feng Ma [1] gave an idea of how to predict the price of the following day using LSTM with time series and optimal input variable to build the LSTM model without trial and error processes which was able to predict the model without strict assumptions of data distribution. Baur, Dirk G.et al [2] says that crypto currencies are assets. A. H. Dyhrberg [3] says that crypto currencies are digital assets and it is not regulated by the government or other organisation and finally concluded that crypto currencies are between assets and currency.

J. Bartos [4] said that crypto investment will be highly efficient and the market prices are immediately affected by new information whereas A. Urquhart [5] concluded that crypto returns are inefficient.

Shah et al [6] used latent source model to predict the price of bitcoin by the method of Bayesian regression. The results of the prediction were in the ratio of 1:4. Pavel Ciaian et al [7] predicts the price of Bitcoin based on Linear regression model, SVM and ANN. The model gave an accuracy of 55% in predicting the price of Bitcoin.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
IJARSCT
Impact Factor: 6.252
Volume 2, Issue 6, May 2022

K. Hegazy and S. Samumford [8] used Multi-Layer Autoregressive (MLP) with Exogeneous inputs (NARX) and forecasted the prices of Bitcoin they found that their model was gave accuracy of 58.21% by testing all model validation tests.

III. IMPLEMENTATION

The model was developed using Python programming language after knowing that the best suited for machine learning where predictions are involved. Python gives an easy access to tensorflow which is an open source artificial intelligence library. Tensorflow has comprehensive and flexible ecosystem which is mainly used in prediction, understanding, creation, discovering, perception and classification. The model imports sklearn which is also known as scikit- learn. This provides all the machine learning library for the development of the model. Apart from this the model uses some in-built libraries in python like numpy, pandas, matplotlib.pyplot and linear regression. In this model the following are the steps involved to look for the best of it.

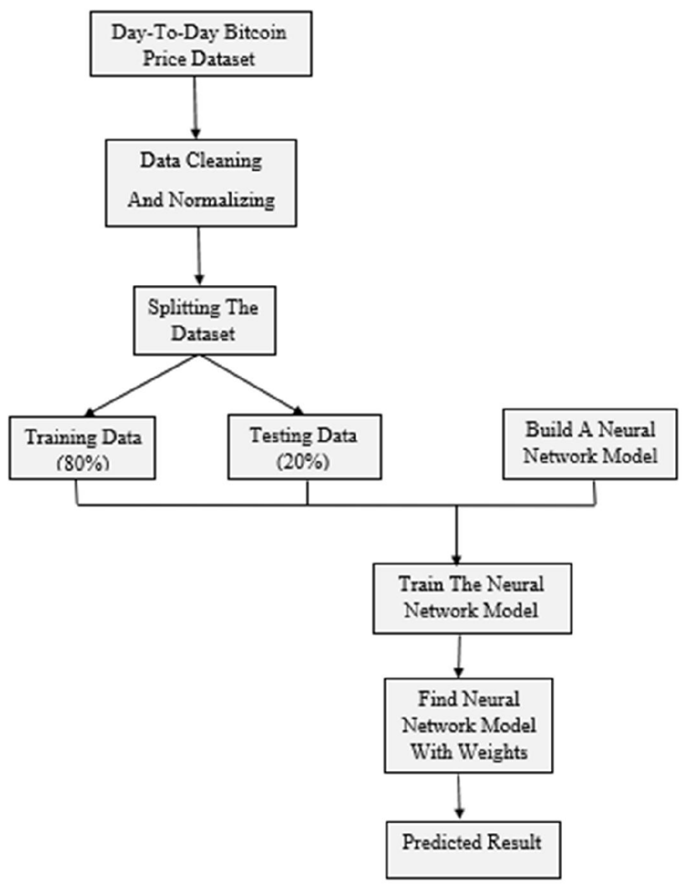

Figure 1: System Flow Diagram

A. DATASET

The data set used in this model is the historical data of bitcoin prices ranging from 2016 to 2012. The data set has six columns specifying the date, open price, high price, low price, close price, adjacent close price and volume.

B. DATA CLEANING AND NORMALIZING

The data we need might have some extra information which is not needed for our model and some other information which is not needed for training the network. Such information or columns must be deleted and the process is called as data cleaning or cleaning the data. The next step we take to our dataset is to normalize the values of the dataset.

Normalizing the values mean to change the numeric columns into a common scale, without distorting differences in the ranges of values. In this model MinMaxScalar is used to normalize the data. The MinMaxScalar is an estimator which scales and translates each feature individually such that it is in the given rage of the dataset. Now we get the normalized values which is shown below in figure 2.

```
array([[0.00081996],
       [0.00080802],
       [0.00076031],
       ...,
       [0.70163085],
       [0.68934511],
       [0.71004211]])
```

Figure 2: Cleaned and Normized Data

C. SPLITTING THE DATASET

Once the dataset is imported, the next step is splitting the data set into training set and test set. For this model the data before 01-01-2016 was taken as the training dataset and the data after 01-01-2020 was taken as the test dataset.

D. BUILDING THE MODEL

The next step is to feed the neurons to the layer and the activation function used for this model is ReLU (Rectified Linear Activation Function) as it shows better efficiency when compared with other activation functions because it allows the network to learn non-linear dependencies which helps the network better fit the algorithm. The first layer in the network is given 100 neurons with ReLU activation function and the input shape, the dropout rate for this layer is 20% that is 0.2. The dropout is a technique which uses to prevent our model from overfitting. For the next layer the number of neurons is 60 with the same activation function and the dropout rate is 40%. Likewise, another 2 layers with 80 and 100 neurons is feed with the dropout rate 40% and 50% respectively.

E. TRAIN THE MODEL

Now we start to feed the data to the recurrent neural network by using LSTM algorithm. First step is to split the training dataset into X_train and Y_train. Form the data set we use all others columns except the close column as X_train and the close column as Y_train which is the output. The timestamp was decided to be 60 for this model and the number of features was 5.

F. COMPILING AND FITTING

The model is compiled by using the 'Adam' optimizer which uses the weighted average of the gradients to accelerate the gradient descent algorithm, along with the loss Mean Square Error (MSE). The loss measures how closely the model's output matches the original output signal. The next step is to import the datasets X_train and Y_train followed by the number of epochs, batch size and the validation loss. The epochs are is the number of iteration or the number of times the model uses the given data, in this model it was assigned to be 20. The batch size is nothing but the tamp stamp, the number of values the model is going to take from a single step which was given to the model as 50. After each epoch the performance of the model is monitored by using the validation of 10%.

G. PREDICTED RESULT

The model is now successfully trained. The next step is to work with the test data set. All steps which were done to the training dataset follows. First, we clean the data set by removing the unwanted columns and the normalize the values by converting them to scalar, finally split the data set into X_test and Y_test. The future values of bitcoin are predicted by using linear regression. Regressor was a variable which was used to perform the linear regression followed by reshaping the array of the testing dataset. The next step is to fit the regressor by using the reshaped X_test and Y_test and by calling the predict function the model predicts the value of Y_pred. To normalize the values of the predicted

IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

Impact Factor: 6.252

Volume 2, Issue 6, May 2022

values and the data set, Standard Scalar is used which will transform the data such that its distribution has a mean 0 and standard deviation 1. Finally, the results were plotted.

```
Epoch 1/10
1695/1695 [==============================] - 64s 27ms/step - loss: 1.8906e-04
Epoch 2/10
1695/1695 [==============================] - 40s 24ms/step - loss: 8.8332e-05
Epoch 3/10
1695/1695 [==============================] - 42s 25ms/step - loss: 6.0315e-05
Epoch 4/10
1695/1695 [==============================] - 43s 25ms/step - loss: 5.7786e-05
Epoch 5/10
1695/1695 [==============================] - 79s 47ms/step - loss: 5.3753e-05
Epoch 6/10
1695/1695 [==============================] - 45s 26ms/step - loss: 5.0823e-05
Epoch 7/10
1695/1695 [==============================] - 44s 26ms/step - loss: 4.1108e-05
Epoch 8/10
1695/1695 [==============================] - 41s 24ms/step - loss: 4.5869e-05
Epoch 9/10
1695/1695 [==============================] - 43s 25ms/step - loss: 4.0217e-05
Epoch 10/10
1695/1695 [==============================] - 42s 25ms/step - loss: 4.3637e-05
```

Figure 3 - Compiled Result

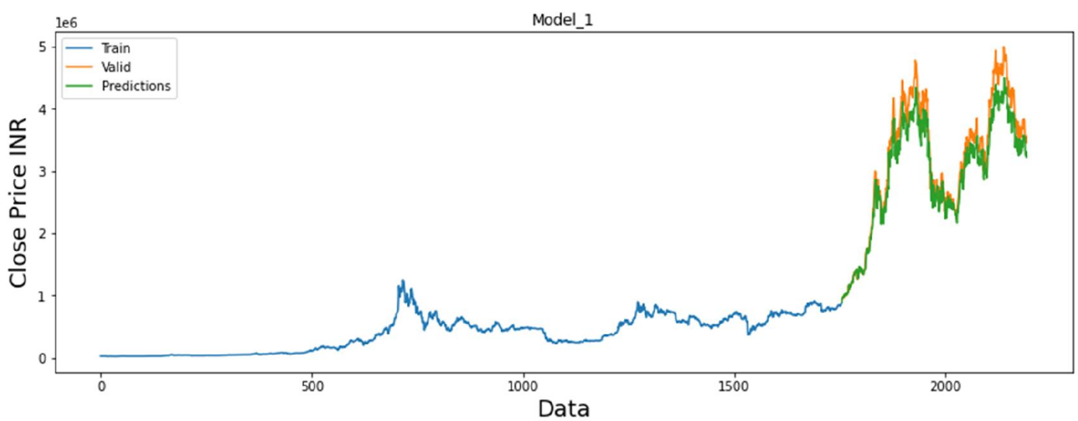

Figure 4 - Predicted Result

H. RESULT AND DISUCSSION

The LSTM model performed well but it was understood that the model was overfitting. To resolve this the number of epochs was increased which never showed much difference. Finally, to solve the problem the number of inputs was reduced in the model and then trained which showed better results without overfitting. The results were analyzed by using the graph shown in figure 4.

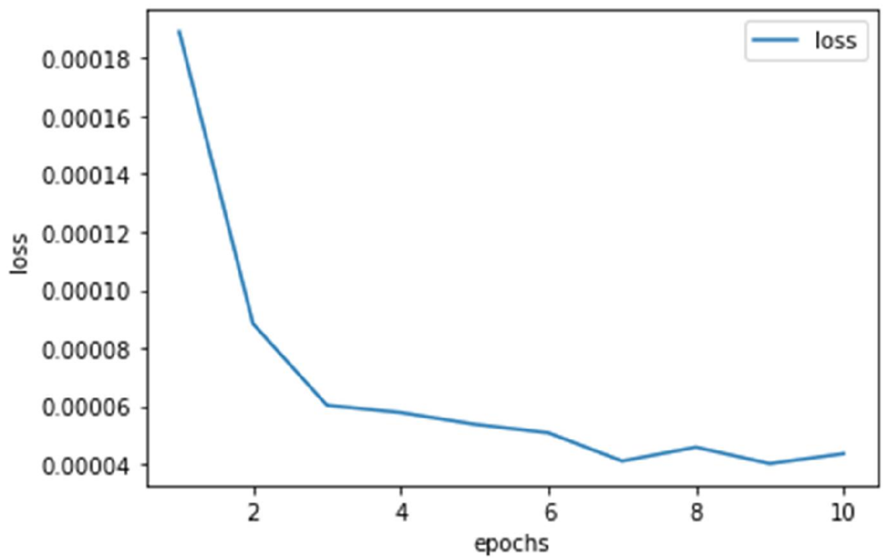

Figure 5 - Epoch Losses

From the above figure 4 we can see that the loss has been reduced after the 5th epoch and after which it doesn't show much variations in values. Close to 9 the epoch loss is the minimum which shows the model is well fit to predict predictions. In this model the number of epochs can also be reduced to 6 since the loss is not much after the 6th epoch Finally, after the prediction of the test data its values almost border the actual bitcoin prices but shows some deviation as shown in the graph in figure 4 in which the blue line in the graph indicates the trained values whereas, the orange line in the graph indicates the test values and the orange line in the graph indicates the predicted values of the model.

IV. CONCLUSION

Artificial Intelligence and deep learning are the base for the improving technology in today's world and in such a developing environment where everybody wants to know the happening of future. Many facilities have paved the way to predict the future. Predicting the values of cryptocurrencies which fluctuate with no particular fashion was impossible but is made possible with the help of machine learning.

The developed model shows a clear prediction of the future prices which has very less deviations from the true prices by using LSTM in tensorflow and keras in python. There is always a thin line between the overfitting of the model and its best performance. The model can also be used for other time series data with some small modifications. This paper helps a lot to learn about the developed model and the algorithm. This study concludes that the machine learning model LSTM (Long Short-Term Method) predicts the future price of crypto currencies with bordering the actual price of the crypto currency.

V. FUTURE WORK

As a plan for future work a recommendation has been made to consider other factors like social media, tweets, natural disasters and war that affects the prices and volume of crypto currencies.

REFERENCES

[1] Chih-Hung Wu,Chih-Hung Lu, Ruei-Shan Lu, Yu-Feng Ma "A New Forecasting Framework for Bitcoin Price with LSTM".

[2] F. Glaser, K. Zimmermann, M. Haferkorn, M. C. Weber and M. Siering, "Bitcoin - Asset or Currency? Revealing Users' Hidden Intentions", ECIS 2014 (Tel Aviv), 2014.

[3] A. H. Dyhrberg, "Hedging capabilities of bitcoin. Is it the virtual gold?", Finance Research Letters, vol. 16, pp. 139-144, 2016.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Impact Factor: 6.252
Volume 2, Issue 6, May 2022

[4] J. Bartos, "Does Bitcoin follow the hypothesis of efficient market?", International Journal of Economic Sciences, vol. IV, no. 2, pp. 10-23, 2015.
[5] A. Urquhart, "The inefficiency of Bitcoin", Economics Letters, vol. 148, pp. 80-82, 2016.
[6] G. H. Chen, S. Nikolov and D. Shah, "A Latent Source Model for Nonparametric Time Series Classification", Advances in Neural Information Processing Systems (NIPS 2013), pp. 1088-1096, 2013.
[7] P. Ciaian, M. Rajcaniova and D.A. Kancs, "The economics of BitCoin price formation", Applied Economics, vol. 48, no. 19, pp. 1799-1815, 2016.
[8] K. Hegazy and S. Samumford, Comparitive Automated Bitcoin Trading Strategies, 2016.