International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 10 | October - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

Cryptocurrency Price Prediction using Deep Learning

Keerthi Aluvala, Uppu Venkat Teja, Gadiputi Madhava Kalyan, Dheeraj Choudhary, Mohd Abid Ali, Swarupa Murala, Sharath Kumar Reddy Guvvala, Partha Sai Aurangabadkar

***Abstract*— Cryptocurrency, a virtual or digital currency that are decentralized networks based on blockchain technology, have become increasingly popular in the recent years. Some of the more famous cryptocurrencies include Bitcoin, Ethereum, YFI Ripple, Dash, Iota, Monero, Stratis, Waves etc. Cryptocurrencies have risen from a gray and black market transaction medium to a more vibrant way of investment, becoming increasingly famous among investors and researches alike. This section of the financial market can be labelled as "high risk high reward". As the crypto market is seeing a 200% growth every year, soon cryptocurrency trading will become a trend in itself. This report documents how our team predicts prices of cryptocurrencies using a Recurrent Neural Network architecture called Long Short Term Memory (LSTM) Algorithm which works on the principle of backpropagation, avoiding long term dependency problem while also optimizing and enhancing the accuracy of the algorithm.**

***Index Terms*— Cryptocurrency, price prediction, Machine Learning, Deep Learning, LSTM, RNNs**

I. INTRODUCTION

EVIDENTLY Machine Learning (ML) has been around for quite some time now. At the same time, we've seen its diverse and wide range of applications in many branches which include, but are not limited to social media, manufacturing, healthcare, gaming, transportation, finance etc. One of machine learning's most sought after subset which is Deep Learning (DL) is also being extensively practiced these days and it is expected that it's use will only increase many thresholds in the years to come. Talking about the world of finance, projects about Stock-price prediction is something which is not unheard of, and many investment firms like Tata Asset Management, DSP Mutual [1] have already started investing into these ventures they call "Quants", which is believed will help them and companies like them in asset-related management services. But, there are many other application spaces of ML and DL in the finance sector that is waiting to be explored like - Cryptocurrency.

The market is diverse and has room for high profit margins if intelligent strategies are taken. The market capitalization of cryptocurrencies in the likes of Bitcoin, YFI and such have been reported to exceed values of $300 billion or 210 hundred Thousand crores in Rupees in 2018 [2]. Although, there was a complete regulatory blanket ban on cryptocurrencies straight up until March 2020 [3], the transactions now are considered legal, but still not considered a legal tender in India, we hope that changing Financial scenarios and an international market for trading in cryptocurrencies will encourage the Government to reform the policies in the near future.

Investing in Stocks, Bonds, Gold, Forex, Real Estate, Crypto has become more popular world- wide. Among these is one of the highest profitable form of investment, which is in the Crypto-land which is a comparatively high risk. As the crypto market is seeing a 200% growth every year [4], soon cryptocurrency trading will become a trend in itself. Currently, select people in the world know about crypto being a form of investment. We, as a team are focusing on this market segment as it is one of the most unexplored division of investment business, utilizing the applications of Machine Learning and Deep learning.

We wish to pursue this project with the goal of predicting the prices of cryptocurrencies which are considered as virtual currencies and are backed on the basis of social consensus, political and technological factors The world has seen a financial boom in the last decades, because of that, more and more people are looking to move their money into soft financial institutions.

One of our team members is an active investor in crypto and knows the potential of this market which will explode by 2025 parabolically. The main and top assets of the crypto space are Bitcoin(BTC) and Ethereum (ETH). Bitcoin is an asset with only 21 million supply in total. There are no more bitcoins than 21 million. Think tanks and Billionaires predict a bitcoin could be worth $100000 by 2025[5]. Bitcoin being the main asset and the driving force in crypto there will be more new people coming in crypto, amid rising popularity, which will eventually make crypto trading among the top 3 investment options.

Future financial price prediction is not at all an easy thing. To state it rather blatantly, back in 2011 Bitcoin(BTC) was worth INR 250, and currently it is close to INR 13,00,000 as of today [6]. We really don't want to miss the next moonshot but we don't know when that will happen. We firmly believe that Machine Learning and Deep Learning can tell the answer.

The high volatility of crypto price can be used to be taken advantage of, and high profits can be offered if some intelligent invent-ful strategies are to be followed.

(a) We are going to train and test real time crypto data and predict the price using neural networks and machine learning. (b)Not only we will be predicting the price of Bitcoin which is the main asset of the crypto space but we will be predicting prices of various Alternate Digital currencies like Ethereum, YFI. (c)With high risk comes high rewards, we will examine various machine learning models and choose the one which suits our needs and requirements, also some traditional algorithms which support prediction of gold and stocks using machine learning.
Companies like Matic, WazirX, Coinbase & Zebpay are paving way for Crypto investment in India [7]. I think our project will be helpful for future investors in crypto to gain insights into this field.

II. LITERATURE REVIEW

A. *Cryptocurrency*

The past of cryptocurrency begins in the 1980s when a computer scientist and cryptographer, David Chaum, suggested a new scheme to build a cryptographic architecture when a sender sends a message that is blinded in such a way that the content of the message cannot be interpreted by the signer. These blind signatures, such as digital signatures, have been used to validate material. Chaum proposed such a view that a third party could not track the contents of the message in any way [8].
In 1998, Wei Dai, a Computer Engineer proposed an anonymous and distributed cash system known as B-money [9]. It is often said that if Bitcoin is today's Cryptocurrency, then B-money can be considered its draft. In his work, he stated a model where two protocols exist in the network that cannot be traced, where only their public keys can identify the senders and receivers, and each message sent is verified by those keys.

A dying light in the line of the new generation cryptocurrency network, which laid the foundation for future study, known as Hashcash[10], a network that relied on cryptographic hash functions to extract evidence of computational work to authenticate the systems, was proposed by Adam Back.

Finally, we have our most common digital currency, Bitcoin [11], which was developed by Satoshi Nakamototo as the first decentralized cryptocurrency. Nakamototo, a pseudo name for a Japanese Computer scientist, around 2008 and 2009 published the first Bitcoin whitepaper around the same time.
Nowadays, it is one of the most highly valued cryptocurrency and a household name, besides other competing cryptocurrencies like YFI and Ethereum, being valued at almost 14 lakh crore Rupees at the current time.

B. *Related Work*

There have been researchers who have previously worked upon in this area and all of them have tried their own methodologies. Some models have been able to predict well while some models proved to be less accurate.

Greaves et al proposed using Logistic Regression, Support Vector Machine (SVM) and analysis using graphs to predict cryptocurrency prices [12]. But only 55% accuracy was achieved with this method because it did not take into account the exchange behavior which affects the prices of cryptocurrencies directly. A later research paper strongly recommends taking into account the exchange behavior which directly affects the prices of these digital currencies.
Almeida et al proposes artificial neural networks for this purpose. The models were generated from Theano library from MATLAB [13]. This model was applied for 2 years in a real time scenario and a profit of $8000 or 4 lakh Rupees was gained from it in that spam.
Shah et al represented Bayesian regression algorithm [14] for generation of source models, using the algorithms was able to register a 67% gain with a sharp immediate prediction ratio. This model was better for immediate prediction, but long term position strategies couldn't have been taken in the market.

McNally et al presented various models that were based on Recurrent Neural Networks, and Autoregressive Integrated Moving Average (ARIMA). The models were generated using traditional datasets containing high price, low price, opening and closing data. The results forecasted a 52.78% accuracy.
ARIMA networks are ideal for short term data forecasting [15], and it works better when the data we are trying to predict has a stable or consistent pattern over time. But this is tedious with real time cryptocurrency price data when the data being processed is really volatile while depending on many complex factors as the data fluctuates violently.

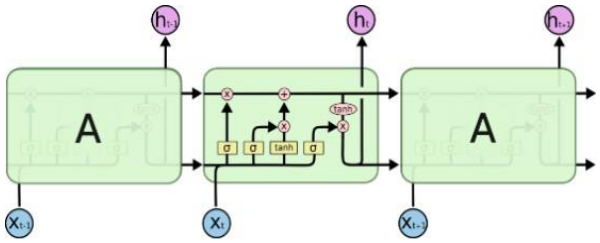
There have been such models tried and implemented, we are going to analyze and predict the prices using a LSTM network.

C. *Long Short Term Memory*

The Neural Network which will be used in this project is the LSTM (Long Short Term Memory) neural network [16]. It is widely used in machine and deep learning and has many advantages. It is a recurrent neural network(RNN) and unlike the standard networks, it also possesses feedback connections. They also have the ability to process data sequences (like speech and video) in addition to processing singular data points (like images). That is why it is widely used in recognition of expression and associated handwriting. Based on time series data, LSTMs find application in classifying, processing and making predictions, as time series may have lags between significant events. This made it unique as an architectural RNN.

The LSTM network, while similar to RNNs, also consists of modules with recurrent consistency. RNNs and LSTMs have a comparable structure, with the memory cell of the secret layer being the only difference. Fig. 1 shows a schematic diagram of the LSTM module. It consists of the three special gates, which is very useful in solving gradient problems.

The biggest advantage of LSTM over a common recurrent network is it's memory cell unit which has the ability to encapsulate part of the previously stored information and also add part of the new information available. This means that it is suitable to avoid long term dependency problem which is often the case in traditional recurrent neural networks [17].



The repeating module in an LSTM contains four interacting layers.

Fig. 1. LSTM Layers representation

D. Data and Methodology

There are many cryptocurrencies functioning in the crypto market, these include, Bitcoin, Ethereum, Tether, Ripple, Dash, Iota, Monero, Stratis, Waves etc. In this paper the currencies taken for research are Bitcoin, YFI and Ethereum. The historical price data for experimentation and processing are taken from yahoofinance.com. The parameters in the dataset include date, open, close, high, low, adj (adjusted close) price and volume.

Features of the datasets are as follows:

Features	***Definition***
Close	Closing trade price
Open	Opening trade price
High	Highest trading price
Low	Lowest trading price
Adjusted Close	Closing price after adjustments
Volume	Total traded volume in a day
Date	Recorded day

Similar to RNNs, the LSTM network also comprises of modules with recurrent consistency. RNNs and LSTMs possess a similar structure, the only difference being the memory cell of the hidden layer. The design of the three special gates is quite implicit in solving the problem of gradients.

The entire tasks involve several processes like collection of data and processing it, feature extraction, training the LSTM recurrent neural network and making predictions with the help of the model trained by us. The preprocessing comprises of steps like data reduction, normalization and cleaning to obtain the required dataset. The data is further split into test and train data. Feature extraction feeds the most important features to the network by combining correlated features and removing less important ones. In our case, it involves parameters like opening, high, low and closing price and date. For accuracy price prediction, random biases and weights are assigned. The LSTM network contains a sequential input layer and a dense output layer where a linear activation function (RELU) has been applied. Using Adam Optimizer and setting the batch values in the LSTM functions will give us the best possible accuracy if applied correctly. The model predicts the value and to check its efficiency the statistic mean absolute error is considered.

	Date	Open	High	Low	Close	Adj Close	Volume
0	2014-09-17	465.864014	468.174011	452.421997	457.334015	457.334015	2.105680e+07
1	2014-09-18	456.859985	456.859985	413.104004	424.440002	424.440002	3.448320e+07
2	2014-09-19	424.102997	427.834991	384.532013	394.795990	394.795990	3.791970e+07
3	2014-09-20	394.673004	423.295990	389.882996	408.903992	408.903992	3.686360e+07
4	2014-09-21	408.084991	412.425995	393.181000	398.821014	398.821014	2.658010e+07
...	...	...	...	...	...	...	...
1927	2019-12-27	7238.141113	7363.529297	7189.934082	7290.088379	7290.088379	2.277736e+10
1928	2019-12-28	7289.031250	7399.041016	7286.905273	7317.990234	7317.990234	2.136567e+10
1929	2019-12-29	7317.647461	7513.948242	7279.865234	7422.652832	7422.652832	2.244526e+10
1930	2019-12-30	7420.272949	7454.824219	7276.308105	7292.995117	7292.995117	2.287413e+10
1931	2019-12-31	7294.438965	7335.290039	7169.777832	7193.599121	7193.599121	2.116795e+10

Fig. 2. Sample Data

The sample data used is shown in Fig. 2. Based on Fig. 3, we get our data from yahoofinance.com. After reading the csv file, we apply feature extraction to it, and retrieve the important features. We split it into train and test datasets, apply LSTM layers, train the model and post process it to show the prediction.

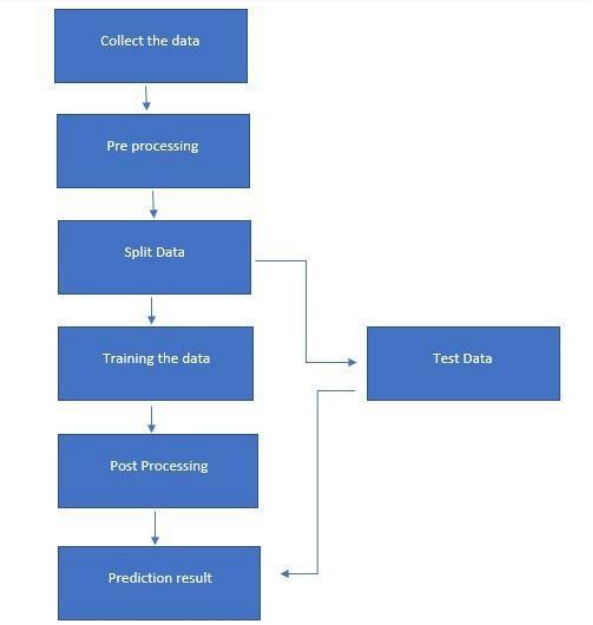


Fig. 3. Flowchart of process

III. RESULTS AND DISCUSSION

Fig. 4 shows the pre-processing stage when we load the dataset into the algorithm, and show the opening and trading prices before we train and test the datasets [18].

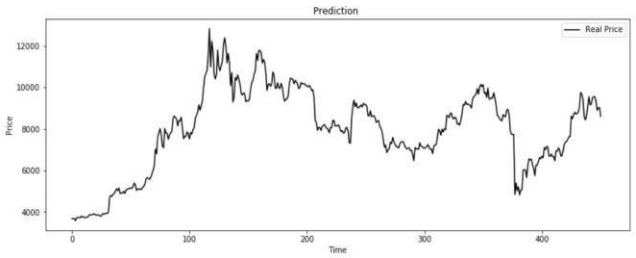


Fig. 4. BTC Price fluctuation in dataset

Total data timestamps in our dataset are from May 2019 to May 2020, we have made the datasets split into two sets: for testing and training [19]. Division has been done along March, and the data before March is used as training and after is used for testing our model.

Before we finalize our results, we measure the loss as a mean absolute error function [20], because it is more interpretable than using mean squared error.

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
lstm (LSTM)                  (None, 60, 50)            11200
_________________________________________________________________
dropout (Dropout)            (None, 60, 50)            0
_________________________________________________________________
lstm_1 (LSTM)                (None, 60, 60)            26640
_________________________________________________________________
dropout_1 (Dropout)          (None, 60, 60)            0
_________________________________________________________________
lstm_2 (LSTM)                (None, 60, 80)            45120
_________________________________________________________________
dropout_2 (Dropout)          (None, 60, 80)            0
_________________________________________________________________
lstm_3 (LSTM)                (None, 120)               96480
_________________________________________________________________
dropout_3 (Dropout)          (None, 120)               0
_________________________________________________________________
dense (Dense)                (None, 1)                 121
=================================================================
Total params: 179,561
Trainable params: 179,561
Non-trainable params: 0
_________________________________________________________________
```

Fig. 5: LSTM layers

Fig. 5 shows the LSTM layers which we have applied in the network. The number of Trainable parameters is 1,79,561 here. Now, by setting the batch size and epochs for each dataset, we strive to optimize our algorithm to provide us the closest prediction.

As we have used the MinMaxScaler function for scaling, we have scaled down values which we revert to bring it to a scaler form. After visualizing the data, we get the graph in the form of a matplotlib function. Here we have shown how we optimized the dataset for Bitcoin, Ethereum and YFI in our model. The projections of predicted prices yielded different accuracy in each case and we had to play with our layers, epochs and batch size to find the most optimized prediction.

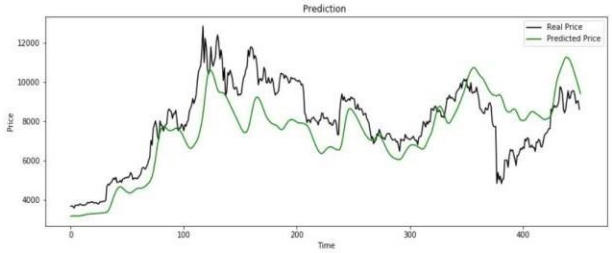


Fig. 6. BTC Prediction(green)

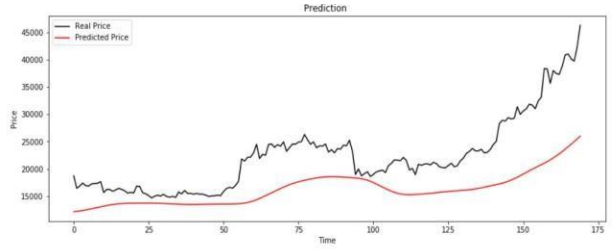


Fig 7. ETH Prediction(red)

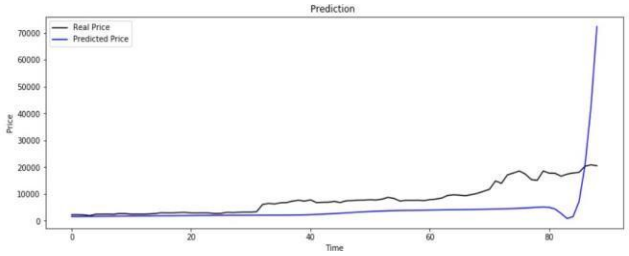


Fig. 8. YFI Prediction(blue)

Fig. 6 to 8 show the calculated price predictions for Bitcoin, Ethereum and YFI respectively. Final observations from these experiments imply that Bitcoin analysis yielded not such a bad outcome, but it isn't enough if you are riding your money on it. The Bitcoin (BTC) data was more predictable than Ethereum (ETH) as more loss in prediction is experience in Fig. 7.

We experience a huge unaccounted spike in our predicted price in the case of Yearn.Finance (YFI), as depicted in Fig. 8, in the end dates of our dataset. This is due to the uncertainty that occurred in our feature extraction, while YFI also being highly volatile. The next price days according to our algorithm is around $15000 (Fig. 9).

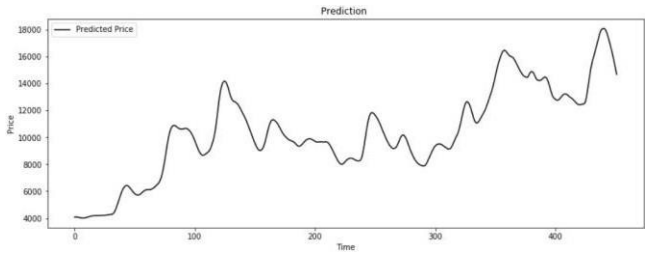


Fig. 9. Quarterly results if our LSTM Algorithm was followed

CONCLUSION

We have been successful in building a model capable of predicting crypto prices taking low, high, opening and closing values, also volume of the currency traded as inputs. It was a good effort on our part. We explored different algorithms and the one with best fit was selected to work upon. Our model has shown good results and accuracy. However, we cannot completely rely on this model's predictions before investing in any cryptocurrency like Bitcoin, Ethereum and YFI. It can be helpful to quite an extent to a new person who wants to invest in the crypto market. Many times crypto and stocks prices have been seen to change significantly due to uncertain factors like political issues, social issues or economic matters.

There is space in this region for further exploration. It is also possible to calculate the effect of unknown events on crypto prices either by using time series analysis or by changing the LSTM model by adjusting the LSTM layers, adding dropout, increasing the number of epochs and using a different dataset of instability.

ACKNOWLEDGEMENTS

1) Teachers and Faculty

2) Teachers and Faculty

REFERENCES

[1] Ronojoy Mazumdar, "Quant funds spread in India where stocks keep hitting records", Economic Times, 28 Jan (2020), https://economictimes.indiatimes.com/markets/stocks/news/quant-funds-spread-in-india-where-stocks-keep-hitting-records/articleshow/73684845.cms.

[2] Rakesh Sharma, "Crypto Investor Novogratz Sees $800B Market Cap In Next Year", Investopedia, 26 Jul (2020), https://www.investopedia.com/news/crypto-investor-mike-novogratz-predicts-800-billion-market-cap-next-12-months.

[3] Manish Singh, "India lifts ban on cryptocurrency trading", 04 March (2020), https://techcrunch.com/2020/03/03/india-lifts-ban-on-cryptocurrency-trading.

[4] Anandi Chandrashekhar & Ashwin Manikandan, "Time to 'HODL' on bitcoin, experts say", Economic Times, 14 May (2020), https://economictimes.indiatimes.com/markets/stocks/news/time-to-hodl-on-bitcoin-experts-say/articleshow/75742625.cms.

[5] Olumide Adesina, "Bitcoin could go for $100,000 by 2025", Nairametrics, 06 Oct (2020), https://nairametrics.com/2020/10/06/bitcoin-could-go-for-100000-by-2025/#:~:text=As%20the%20fastest%20ever%2Dchanging,hit%20a%20valuation%20of%20%24100%2C000.

[6] Yahoo Finance, BTC-INR, https://in.finance.yahoo.com/quote/BTC-INR?p=BTC-INR.

[7] Shailak Jani, "The Growth of Cryptocurrency in India: Its Challenges & Potential Impacts on Legislation", Parul University, 2-3, April 2018.

[8] A. Judmayer, N. Stifter, K. Krombholz, and E. Weippl, "Blocks and Chains: Introduction to Bitcoin, Cryptocurrencies, and Their Consensus Mechanisms," Synth. Lect. Inf. Secur. Privacy, Trust, 2017.

[9] W. Dai, "B-money proposal," White Pap., 1998.

[10] A. Back, "Hashcash." 1997.

[11] S. Nakamoto and others, "Bitcoin: A peer-to-peer electronic cash system," 2008.

[12] A. Greaves and B. Au, "Using the bitcoin transaction graph to predict the price of bitcoin," No Data, 2015.

[13] Almeida, J., Tata, S., Moser, A., Smit, V.: Bitcoin prediciton using ANN. Neural networks. 1±12 (2015).

[14] Amjad, M., Shah, D.: Trading bitcoin and online time series prediction. In: NIPS 2016 Proceedings. pp. 1±15 (2017).

[15] Box, G.; Jenkins, G. Time Series Analysis: Forecasting and Control; Holden-Day: San Francisco, CA, USA, 197

[16] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Comput., vol. 9, no. 8, pp. 1735– 1780, 1997.

[17] F. Qian and X. Chen, "Stock Prediction Based on LSTM under Different Stability," in 2019 IEEE 4th International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), 2019, pp. 483–486.

[18] AISHWARYA SINGH, "Stock Prices Prediction Using Machine Learning and Deep Learning Techniques (with Python codes)," OCTOBER 25, 2018, 2018. [Online]. Available: https://www.analyticsvidhya.com/blog/2018/10/predic ting-stock-price-machine-learningnd-deep-learningtechniques-python/.

[19] Jang, H., Lee, J.: An Empirical Study on Modeling and Prediction of Bitcoin Prices with Bayesian Neural Networks Based on Blockchain Information. IEEE Access. 6, 5427±5437 (2017). https://doi.org/10.1109/ACCESS.2017.2779181.

[20] D. Shah and K. Zhang, "Bayesian regression and Bitcoin," in 2014 52nd Annual Allerton Conference on Communication, Control, and Computing (Allerton), 2014, pp. 409–414.