International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

Cryptocurrency Prediction using Deep Learning

Abusufiyan Athani[1], Samarth Kumar[2], Kirankumar K R[3], Abhishek Sonavane[4]

[1]Student, PES University, Bangalore, India
abuathani26[at]gmail.com

[2]Student, PES University, Bangalore, India
samarthkumar762[at]gmail.com

[3]Student, PES University, Bangalore, India
kirankumarkr5901[at]gmail.com

[4]Student, PES University, Bangalore, India
abhisheksonavane1718[at]gmail.com

Abstract: ***Cryptocurrencies are considered to be big fastest growing players in the financial sector in the current scenario of the world. Many individuals, institutions, and corporate enterprises had invested heavily in this. In the last 5years, there is so much fluctuation in almost every crypto-currency price due to government regulations, public sentiment, media hype, and supply and demand of investors and users. So, We have decided to suggest cryptocurrency prices using deep learning techniques which make links between changes in the price of the cryptocurrency, their historical data, and other factors which will affect cryptocurrency price. We have also provided a technique which is suggesting a price of a few days in the future and helping in taking investment decisions about which cryptocurrency is going to give more benefit in the future. Also, provide currency conversion techniques.***

Keywords: Crypto-Currency, Bitcoin, Ethereum, Dash Coin, Litecoin Correlation, Conversion, LSTM, GRU, Prediction Price, Flask & Recommendation.

1. Introduction

Satoshi Nakamoto made the first cryptocurrency i.e. Bitcoin. It was going to be available for the public on January 3, 2009. It is decentralized digital or virtual money which doesn't rely on middlemen like banks or government institutions to verify transactions. It is not even controlled by any of the government or central institutions but it is managed by peer-to-peer networks and transactions are done online, near-instantly, 24/7 at low fees. But cryptocurrency is very volatile and stored in a digital wallet if we lost our wallet then lost our entire cryptocurrency investment. Also, so many different cryptocurrencies came onto the market such as Ethereum, Litecoin, Dash coin, and so on. There are also many developed countries that considered many crypto-currency as legal currencies such as Canada, Japan, United States, South Korea, United Kingdom, Germany, and so on. In our country, almost 14 lack trading accounts were opened during the lockdown period and more than half of trading account owners came from a non-financial background and they had very less or almost no knowledge about the trading world.

Over the last 5years, there are so many fluctuations in the price of almost every crypto-currency due to various reasons such as government regulations, public sentiment, media hype, and supply and demand of investors and users. So, there is a need for applications that will help you to analyse cryptocurrency prices in the future days using deep learning techniques such as a hybrid model which takes link between price changes for cryptocurrencies, their historical data, and other factors. Also, suggest which is best cryptocurrency to invest in so we get maximum profit and best cryptocurrency to sell at that point of time from which we hold. It also provides cryptocurrency conversion in different countries currencies for better visualization of users.

2. Literature Review

We had the impression, thanks to the thorough literature review, that numerous machine-learning methods had been employed to forecast the price of cryptocurrencies. All of these papers' price predictions are either too high or too low in relation to the cryptocurrency's real price. These studies also employed a stochastic neural network, where each layer's output is supplemented with a random model. The current activation function and the random activations from the previous timestep are both inputs to the module. In neural networks, non-linear dependencies between market variables, blockchain data, and social sentiment are captured using LSTM and MLP. MLP and LSTM models will be trained on both normalised and non-normalized samples. The MLP model has six layers, each with an activation function, whereas the ReLU is trained over 700 iterations using the Adam method. Using flattening to create a vector of length 161, the data consists of 23 features from the previous 7 days. In input-to-output directions, layers input comprises 130, 100, 50, 25, 10, and 1 neuron. In the Normalized dataset, the average gain from stochastic neural networks over traditional neural networks is 1.565 at gamma=0.1, 1.73% at gamma=0.12, and 1.76% when gamma is made a learnable constant. For the Unnormalized dataset, the average increase is 0.19% at gamma=0.1, 4.525 at gamma=0.12, and 7.41% at gamma=0.13.

Suraj Bhatt (2019) suggested a paper in which it is scraped data from Webhouse.io and Twitter using the Python Twitter crapper library. They employed long short-term memory to analyse the mood from twitter data (LSTM). Then, sentiment scores from Twitter and news are mapped with previous Bitcoin data using UNIX timestamps. We have employed the Random Forest Regression technique to forecast cryptocurrency prices. Based on which forecasts earned the most votes, the random forest analyses the predictions from each decision tree to make an educated guess as to what will happen. By breaking up all data into digestible chunks and training on various trees, random forest regression has the advantage of preventing overfitting. The random forest algorithm employs a structure like a tree to calculate, remember, and store the previous value. But, when large prices are present, tree computations become quite large, and the random forest is unable to calculate accurately. As a result, the output of the random forest method is less precise when predicting the prices over more ranges.

Franco Valencia (2020) suggested a paper that uses information from Twitter and market data to compare neural networks, support vector machines, and random forests. Since Random Forest is an ensemble method, it can be found under Sklearn. The model will be provided data as a variable, and it will then be fitted to the data. A text classification tool called sentiment analysis determines if sentiment is favourable, neutral, or negative. The results would be less effective if the public attitude were neutral, as determined by VADER sentiment analysis, as neutral sentiment typically does not reveal a pattern for purchasing or selling. Regardless of future price adjustments, Twitter sentiment on cryptocurrency seems to be positive. Overfitting is avoided with Random Forests since the entire dataset is divided into manageable portions. RFs anticipate prices using historical data and an ensemble technique. Price forecasting is not always accurate.

Arti Jain (2019) proposed a paper which had employed two cryptocurrencies—Bitcoin and Litecoin—to make predictions, collected data over a 30-day period, and scraped Twitter data in JSON format before converting it to CSV. The training phase, which is a one-time task, and the detection phase are the two stages in which the system operates (real-time tweets are inputted into the model, and the model predicts the average price for the duration of two hours). The response from "Textblob, sentiment, polarity" ranges from -1 to 1. The sentiment polarity of the tweets in the data is examined. Positive tweets are those with a polarity greater than 0. Neutral tweets are those that have a polarity of zero. Tweets with polarity less than 0 are those employed two cryptocurrencies—Bitcoin and Litecoin—to make predictions, collected data over a 30-day period, and scraped Twitter data in JSON format before converting it to CSV. The training phase, which is a one-time task, and the detection phase are the two stages in which the system operates (real-time tweets are inputted into the model, and the model predicts the average price for the duration of two hours). The response from "Textblob.sentiment.polarity" ranges from -1 to 1. The sentiment polarity of the tweets in the data is examined. Positive tweets are those with a polarity greater than 0. Neutral tweets are those that have a polarity of zero. Negative tweets are those with a polarity of less than 0. All of the tagged tweets are saved, and the data that has been saved is divided into chunks that contain tagged tweets that were posted over the course of two hours. The proportion of each type of tweet—positive, neutral, and negative—that makes up a given chunk is counted. The average price for the accompanying two-hour time period is then projected onto these counted figures. Future price predictions are made using multiple linear regression (MLR) models. Model of the MLR equation. Y = b0 + b1 X1 + b2 X2 + b3 X3 +... + bn Xn • Here, X1, X2, X3,..., Xn are the independent variables, Xn is the dependent variable, and b0, b1, b2,..., bn are the coefficients. It is used to track changes in the dependent variable for cryptocurrency predictions as the independent variable changes. Because it is more efficient every two hours, the price of cryptocurrencies is anticipated every two hours. Bitcoin's R2 score is 59%, while Litecoin's is 70%. The stats model is used to determine whether or not the MLR assumption is true. To find different findings, the statistics model uses the Ordinary Least Square (OLS) Method. The Stats model verifies that all of the MLR's presumptions are satisfied as well as the linearity and lack of correlation of the independent variables. More social aspects need to be measured in order to generate more accurate predictions of bitcoin pricing.

3. Proposed Workflow

A. Dataset Description

We have extracted data of four cryptocurrencies Bitcoin, Dash, Ethereum, and Litecoin from bitinfocharts.com using a web scraping technique.

These extracted datasets contain features such as:

- **Market Capitalization**: Total value of cryptocurrency present in the market in USD.
- **Price:** Value of one cryptocurrency in the market in USD.
- Transaction Value: Amount of cryptocurrency required to add a transaction to the next block in the blockchain.
- **Transaction Fee:** It is the average transaction fee paid by the parties for the confirmation of their transactions. The fee is required in order to process the transaction in the network.
- **Profitability:** Profitable income to the miner against the use of resources for consuming power and time. With the increase in the number of miners, the reward per miner is decreased exponentially.
- **Hash rate:** The amount of computational power required to verify and add the block of transactions to the cryptocurrency's blockchain.
- **Reward:** The amount of cryptocurrency that miners get after successfully validating the last transaction of a blockchain network's transaction block.
- **Difficulty:** It is the computational difficulty required to mine a single block of the coin.

4. Methodology

The project is divided into parts such as frontend and backend. In Frontend, we are using Django, HTML templates, and sqlite3 to implement features of the project such as login page, signup page, home page, about page, conversion page, and analysis page.

International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

Figure 1: Master class diagram of Cryptocurrency Prediction System

We have taken market cap, difficulty, profitability, hash rate, and transaction fee in predicting cryptocurrency price whose R-value is greater than 0.5 for all four cryptocurrencies such as Litecoin, Dash, Bitcoin, and Ethereum coin. The features having a correlation great than 0.5 are used to train the model, all these were normalized to a range between 0 and 1 using a MinMaxScaler to avoid any dominance by any of the features. We are using previous 15 days price of the coin values along with other features of currencies to capture the price trends if any. All the normalized values of previous price and other features are passed to the deep learning model for training. Here X_train contains the independent values and y_train is the dependent variable. We have used a sequential model trained for 100 epochs and 50 units. The loss function used is mean squared error and optimizer used is Adam optimizer. We have also used a dropout of 0.2 to avoid the problem of overfitting the model. After training the model when passed with the previous 15 days price values along with other features gives the output of 50 probable values for that day. We are predicting price for next 60 days and to predict the price of a particular day, our model needs the values of the features for that day which is not available at present. To solve this we are training a model for each and every feature. These models are simple models which will be using the previous 15 days values and predicting next day's value. We will be needing values for 60 days, so we will be feeding the predicted prices to the model to get all 60 day's values. Now to predict the price for next 60 days, we are predicting all the features of the next first day, and using those features we are predicting the price of next first day. To get all features of the second next day, we need previous 15 day's values among them one is the value of first next day which we have already predicted. This will be used predict the features of the second next day, those features will be used to predict second day's price and this will be continued to get prices of all 60 days. To get features of a day, we have defined functions which take input the day number and give the predicted value of that particular feature on that day. The model returns 50 probable values for each day, so we take the average of these values. The values are between the range 0 to 1, to get back the actual values we inverse transform the values. These values will be the predicted values in actual dimension.

5. Results

We have plotted a graph in the x-y axis for all four coins bitcoin, Ethereum, Dash coin, and Litecoin comparing actual and predicted prices.

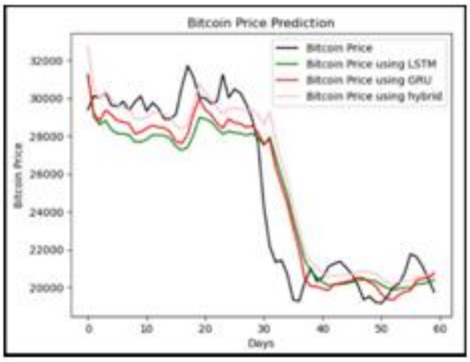

Figure 2: Plotting actual price vs predicted price using different model of bitcoin

We got Mean Absolute Error (MAE) for bitcoin using lstm model as 0.078, gru model as 0.056, and hybrid model (lstm+gru) as 0.057. So, the GRU model is performing as compared to all the above three models.

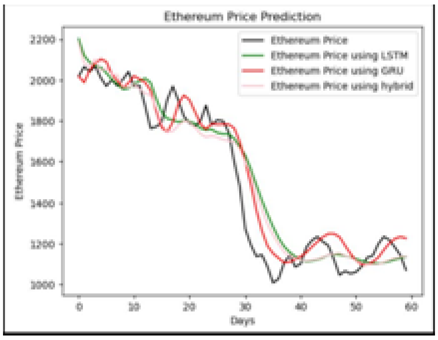

Figure 3: Plotting actual price vs predicted price using different model of Ethereum

We got Mean Absolute Error (MAE) for ethereum using lstm model as 0.0715, gru model as 0.0596, and hybrid model(lstm+gru) as 0.0671. So, the GRU model is performing as compared to all the above three models.

Figure 4: Plotting actual price vs predicted price using different model of dashcoin

We got Mean Absolute Error (MAE) for dashcoin using lstm model as 0.069, gru model as 0.049, and hybrid model (lstm+gru) as 0.0605. So, the GRU model is performing as compared to all the above three models.

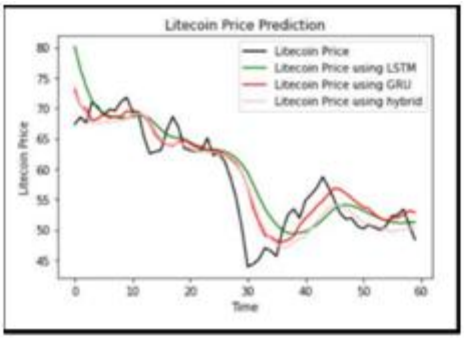

Figure 5: Plotting actual price vs predicted price using a different model of litecoin

We got Mean Absolute Error (MAE) for litecoin using lstm model as 0.062, gru model as 0.04, and hybrid model(lstm+gru) as 0.0533. So, the GRU model is performing as compared to all the above three models.

6. Conclusion

The project began with research on features that will affect cryptocurrency prices and then we will get to know there are basically seven features that will affect cryptocurrency prices such as market cap, transaction value, transaction fee, rewards, profitability, hash rate, and difficulty. We started to collect datasets using web scraping from the website bitinfocharts.com. At the end of the literature survey, we have a clear picture of the dataset, technologies, and algorithms that are important for us and what we need to explore. Then explore in model LSTM and GRU and made a hybrid model (LSTM + GRU) but this model has more mean absolute error than the GRU model. Thus, we conclude that the GRU model is the best model for predicting cryptocurrency price. As part of Future work, the project can be further extended by adding sentiment features using tweets or news article.

References

[1] D.L.K. Chuen, Handbook of Digital Currency: Bitcoin, Innovation, Financial Instruments, and Big Data, Academic Press, 2015.

[2] Zheshi Chen, Chunhong Li, Wenjun Sun, Bitcoin price prediction using machine learning: An approach to sample dimension engineering, Journal of Computational and Applied Mathematics, Volume 365, 2020, 112395, ISSN 0377-0427.

[3] S. McNally, J. Roche and S. Caton, "Predicting the Price of Bitcoin Using Machine Learning," 2018 26th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP), 2018, pp. 339-343, DOI: 10.1109/PDP2018.2018.00060.

[4] Phaladisailoed, Thearasak&Numnonda, Thanisa. (2018). Machine Learning Models Comparison for Bitcoin Price Prediction. 506-511. 10.1109/ICITEED.2018.8534911.

[5] C. Wu, C. Lu, Y. Ma, and R. Lu, "A New Forecasting Framework for Bitcoin Price with LSTM," 2018 IEEE International Conference on Data Mining Workshops (ICDMW), 2018, pp. 168-175, DOI: 10.1109/ICDMW.2018.00032.

[6] Patrick Jaquart, David Dann, Christof Weinhardt, Short-term bitcoin market prediction via machine learning, The Journal of Finance and Data Science, Volume 7, 2021, Pages 45-66, ISSN 2405-9188.

[7] Zaiqiao Meng, Richard McCreadie, Craig Macdonald, IadhOunis, “Exploring Data Splitting Strategies for the Evaluation of Recommendation Models”, arXiv:2007.13237, 2020

[8] Zahraa Said Abdallah, Lan Du, and Geoffrey I Webb, “Data Preparations”, Encyclopedia of Machine Learning and Data Mining, Springer, 2017.

[9] Gökhan Aksu, CemOktayGüzeller and Mehmet Taha Eser, “The Effect of the Normalization Method Used in Different Sample Sizes on the Success of Artificial Neural Network Model”, International Journal of Assessment Tools in Education, 2019, Vol. 6, No. 2, 170–192, DOI: 10.21449/ijate.479404

[10] Yun Xu and Royston Goodacre, “On Splitting Training and Validation Set: A Comparative Study of Cross-Validation, Bootstrap and Systematic Sampling for Estimating the Generalization Performance of Supervised Learning”, Journal of Analysis and Testing (2018), DOI: 10.1007/s41664-018-0068-2

[11] D. E. Birba, “A Comparative study of data splitting algorithms for machine learning model selection”, Dissertation, 2020.

[12] ZulfadzliDrus, Haliyana Khalid, Sentiment Analysis in social media and Its Application: Systematic Literature Review, Procedia Computer Science, Volume 161, 2019, Pages 707-714, ISSN 1877-0509.

[13] Stephan A. Curiskis, Barry Drake, Thomas R. Osborn, Paul J. Kennedy, An evaluation of document clustering and topic modelling in two online social networks: Twitter and Reddit, Information Processing & Management, Volume 57, Issue 2,2020, 102034, ISSN 0306-4573.

[14] M. Briaere, K. Oosterlinck, and A. Szafarz,Virtual currency, tangible return: Port- folio diversification with Bitcoins, Tangible Return: Portfolio Diversification with Bitcoins , 2013.

[15] S. Nakamoto, ”Bitcoin: A Peer-to-Peer Electronic Cash System”.

[16] McNally, Sean & Roche, Jason & Caton, Simon. (2018). Predicting the Price of Bitcoin Using Machine Learning. 339343. 10.1109/PDP2018.2018.00060