A Survey on Cryptocurrency Price Prediction Techniques

Aurko Mitra
MT21166
aurko21166@iiitd.ac.in

Debarshi Dasgupta
MT21167
debarshi21167@iiitd.ac.in

Abstract

Blockchain technology has received a lot of attention from investors, speculators, developers and other retail participants in the past few years. There have been a number of academic studies on the same in the relatively recent past, especially on the most noticeable blockchain applications like Bitcoin, Ethereum, etc. A good amount of research has gone into predicting cryptocurrency prices using machine learning techniques. Some approaches make use of basic price history and some use social media sentiment while a few make use of on-chain metrics. This paper aims to present a brief survey on such developments in recent history.

1 Introduction

Up until a little more than a decade ago the asset class called cryptocurrencies didn't exist, until the Bitcoin white-paper was released (Nakamoto, 2008) which introduced a decentralised and permission-less network of interconnected nodes for transferring value. In the recent history, the rapid rise in cryptocurrency prices such as Bitcoin (BTC), Ethereum (ETH) and its counterparts has attracted a lot of attention from multiple communities of people. From investors to speculators, developers and even academia. Figure 1 shows the closing price history of Bitcoin. The graph depicts the meteoric rise of Bitcoin prices from 2011 till 2021 with the y-axis depicting BTC/USD closing prices in the logarithmic scale.

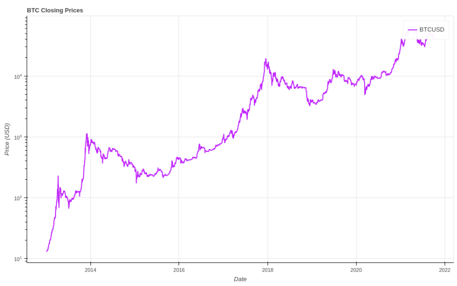

Figure 1: BTCUSD Closing Price Vs Date for last 10 years

Studies on price prediction of assets is not new but most research is focused on stocks, commodities, real estate and the foreign exchange markets. Certainly, this price rise in BTC has attracted similar interest in cryptocurrencies.

2 Literature Review

Amongst the currently available literature, a lot of work has been carried out in the field of price prediction for assets like commodities, stocks, real estate, etc. However, not much of published research has looked at price prediction for cryptocurrencies until recently. Figure 2 presents the BTC daily percentage returns since 2011 till 2021 in the form of a histogram. One can see that the price returns are like a normal distribution with a mean at zero. This makes the study for prediction of price trends even more interesting.

Based on the limited set of works carried out this aspect, their approaches can be divided into three main categories i.e., Firstly, those making use of social media sentiment for price prediction. Secondly, the ones that make use of time series data directly to predict price trend and lastly a few set of papers which look at on-chain metrics for predicting the direction of future price trend. The rest of the literature survey is arranged into subsections delving deeper into these three different approaches.

2.1 Sentiment Driven Price Prediction

Nowadays analysis of sentiments is an important tool for prediction of cryptocurrency price when it is combined with some other factors.Sometimes the cryptocurrency price also heavily depend on the sentiments people share and so the analysis of this relation is very crucial for price prediction.

(Kim et al., 2016) analyzed comments made by various kinds of users in different online platforms for prediction of the price and number of transactions of cryptocurrencies. The fluctuations in the price of cryptocurrencies were predicted by the proposed method at a very low cost. This proposed method was at par with any of the previous methods targeting the same task. It was seen that comments and replies made by users in different online platforms affected the number of transactions that took place. The method which

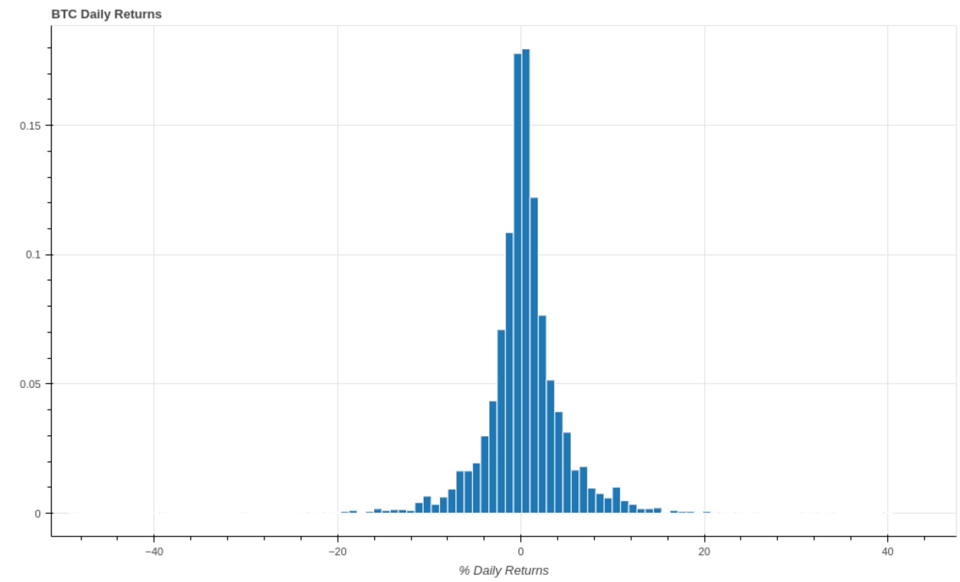

Figure 2: BTC daily percentage returns from 2011 to 2021.

was proposed is applicable for buying and selling cryptocurrencies and helped to determine various aspects which affected user opinions. In another research paper (Serafini et al., 2020) ARIMAX (similar to models of multivariate regression) and LongShortTermMemory based Recurrent Neural Network (a neural network with feedback where the training of the model takes place involving the generated outputs too) was trained with input features which involved several combinations of financial and sentiment features. However ARIMAX was found to be the clear winner which outperformed LSTM based recurrent neural network.

2.2 Time Series Approach

Time series approach is another way which can be used for prediction of cryptocurrency prices. A time series model is very different from other models like linear regression because the assumption that all the samples are linearly independent from each other does not hold good here. Also, most time series show certain highs and lows during specific intervals of time which depends on different types of factors as a result of which ordinary machine learning models cannot be used.

In the paper (Mudassir et al., 2020) short-term to mid-term Bitcoin price forecasting was done using machine learning models. It considered price indicators till December 31, 2019. The four models which were used were Artificial Neural Networks, Statistica Automata Neural Networks, Support Vector Machines and Long Short Term Memory. Artificial neural networks are similar to that present in the nervous system. The neural network consists of an input layer, output layer and hidden layers. The weights and biases associated with the neurons are optimized by backpropagation such that the loss calculated at the output is minimum. Statistica Automata Neural Networks use statistical methods for the same. Support Vector Machine is a supervised machine learning algorithm which can be used to classify linearly-non separable data by drawing a hyperplane which can separate them in a higher dimension. LSTM is a form of recurrent neural network which considers model output too while training. Out of all these models LSTM showed the best overall performance. (Hamayel and Owda, 2021) showed that bi-LSTM represents less accuracy than GRU and LSTM with differences between the actual and prices which are predicted. Gated Recurrent Unit is a similar to LSTM which aims to deal with the vanishing gradient problem. It is seen that Gated Recurrent Unit can predict cryptocurrency prices much better than LSTM and bi-LSTM. (Poyser, 2019) revealed that Bitcoin is linked with various financial and interest factors. But none of the internal factors had a relevant impact on the price, instead it depends macro-financial variables of the country.

2.3 On-Chain Metrics

Due to the public and permissionless nature of blockchains, they present a unique opportunity for analysis. One can analyse on-chain data to see how the blockchain network is being used by people and in turn help in prediction of price trends.

(Ciaian et al., 2016) presented one of the first articles that studies Bitcoin price formation by considering digital currency specific factors along with traditional parameters like market forces of supply and demand. This work made use of factors such as the total number of unique Bitcoin transactions per day, and the number of unique Bitcoin addresses used per day. They also came up with another parameter called ’number of days destroyed’ which along with the previous two parameters was found to be statistically significant for prices.

In another work by (Kurbucz, 2019), the Bitcoin network was modelled as a graph with individual addresses as nodes and the transactions as edges connecting them. Only a very small subset of the network was taken and used to find a correlation with future prices. Random forest was used to identify the most frequent edges and deemed important after which a single hidden layer feed-forward (SLFN) network was used. The authors reported an accuracy of 60.05% for prediction of price trend. It is noteworthy that this work did not make use of the entire graph but managed to produce the results with only a small subset of it.

A more recent work by (Chen et al., 2021) considered multiple on-chain metrics such as:-

Block size: the size of the data in the blockchain network containing the permanently recorded data.

Confirmation time: The average time required to confirm a transaction on the network and add it to the next block.

Difficulty: The mining difficulty of the network.

Hash rate: The value of the entire bitcoin network's hash rate. It is currently in exahashes per second.

Mining profitability: The value of the block reward and transaction fees for successfully mining a single block.

Average Transaction fee: The average fees per transaction on the bitcoin network in one day.

Average transaction value: The average value exchanged per transaction on the bitcoin network in one day.

Transaction volume: Considers the total volume of transactions on the network i.e., the number of transactions between users.

These metrics were used with a long-short-term-memory(LSTM) network. The results presented in the work demonstrated that using these parameters was effective for predicting the price trend for Bitcoin.

The works presented before in this section were focussed on price prediction for Bitcoin however in recent years another popular cryptocurrency has emerged. In fact, the 2nd most popular cryptocurrency in terms of market capitalisation after Bitcoin i.e., Ethereum. (Kim et al., 2021) showed that Ethereum, a popular cryptocurrency in the market, has blockchain information that is a bit different from that of Bitcoin. This is owing to a slightly different features offered such as smart contracts however the underlying decentralised architecture is mostly similar. Their work investigated the relationship between these metrics on the Ethereum blockchain with Ethereum prices.

Apart from common blockchain information such as the one's used by (Chen et al., 2021), here Ethereum specific blockchain information was also used. More metrics such as block gas limits, gas price, gas used, and uncle blocks along with other metrics like ones used in the previous paper were used. However no attention was paid to social media sentiment or other factors like google trends, etc. The authors even tried to look at the information of a few other popular coins such as Bitcoin, Dash and Litecoin to see if their information could also be an indication for Ethereum prices. Finally, the authors concluded that Ethereum specific blockchain information was the best contributor in terms of performance along with other factors like the blockchain information common to both Bitcoin and Ethereum.

In (Saad et al., 2020) (2020 Toward Characterizing Blockchain-Based Cryptocurrencies for Highly Accurate Predictions) even more on-chain metrics were used to predict price trends for both Bitcoin and Ethereum. The metrics used were block size, confirmation time, transaction volume both per day and per block, total bitcoins count, average transaction fees, difficulty, hash rate, total addresses on Bitcoin network, mining profitability, unspent transaction outputs which mark the spendable bitcoin on the network, the mempool size which can vary from one node to another. Along with all these again Ethereum specific factors like block gas limit, gas price was used. Finally with these metrics the authors made use of an LSTM network to predict price trend.

3 Scope of new research

For future work, even factors like the effect of government regulations in different jurisdictions could be studied. Furthermore modern natural language processing techniques could be leveraged to see effect of real time tweets on prices. From the set of works looked at in this review, we see that harnessing on-chain data for prediction is unique to cryptocurrencies. It represents activity on the blockchain network itself and is an indicator for price with good correlation. However, recent literature does not look at Dapp (Decentralised application) data on Ethereum based blockchains.

Ethereum based blockchains being a little more complex than Bitcoin-like chains allow the use of smart contracts. Due to this a number of applications have emerged on such chains and generate a lot of financial activity. Many of these Dapp's have transaction volumes in multiple millions of dollars and some have a total value locked of few billion dollars. As per (DefiLlama), as of April 2022, the present DeFi (Decentralised Finance) ecosystem has a total value locked of about one hundred fifty billion dollars. Not a lot of research has looked at this Dapp and data we believe making use of this data to predict cryptocurrency prices is surely worth exploring.

4 Conclusion

Bitcoin and Ethereum both are successful cryptocurrencies in their own ways, both have existed for a number of years and have garnered a sizeable amount of transaction traffic. Both of them have also been studied in the fields of economics and computer science. In this paper we have tried to present a brief survey on 3 different kind of approaches for price prediction of these assets. From the study one can conclude that making use of multiple on-chain metrics in conjunction with deep learning techniques has observed some success and presents a promising avenue for future research.

References

Wei Chen, Huilin Xu, Lifen Jia, and Ying Gao. 2021. Machine learning model for Bitcoin exchange rate prediction using economic and technology determinants. *International Journal of Forecasting*, 37(1):28–43.

Pavel Ciaian, Miroslava Rajcaniova, and d'Artis Kancs. 2016. The economics of BitCoin price formation. *Applied Economics*, 48(19):1799–1815.

DefiLlama. Defi Llama - A DeFi TVL aggregator.

Mohammad J. Hamayel and Amani Yousef Owda. 2021. A Novel Cryptocurrency Price Prediction Model Using GRU, LSTM and bi-LSTM Machine Learning Algorithms. *AI*, 2(4):477–496.

Han-Min Kim, Gee-Woo Bock, and Gunwoong Lee. 2021. Predicting Ethereum prices with machine learning based on Blockchain information. *Expert Systems with Applications*, 184:115480.

Young Bin Kim, Jun Gi Kim, Wook Kim, Jae Ho Im, Tae Hyeong Kim, Shin Jin Kang, and Chang Hun Kim. 2016. Predicting Fluctuations in Cryptocurrency Transactions Based on User Comments and Replies. *PLOS ONE*, 11(8):e0161197.

Marcell Tamás Kurbucz. 2019. Predicting the price of Bitcoin by the most frequent edges of its transaction network. *Economics Letters*, 184:108655.

Mohammed Mudassir, Shada Bennbaia, Devrim Unal, and Mohammad Hammoudeh. 2020. Time-series forecasting of Bitcoin prices using high-dimensional features: a machine learning approach. *Neural Computing and Applications*.

Satoshi Nakamoto. 2008. Bitcoin: A Peer-to-Peer Electronic Cash System.

Obryan Poyser. 2019. Exploring the dynamics of Bitcoin's price: a Bayesian structural time series approach. *Eurasian Economic Review*, 9(1):29–60.

Muhammad Saad, Jinchun Choi, DaeHun Nyang, Joongheon Kim, and Aziz Mohaisen. 2020. Toward Characterizing Blockchain-Based Cryptocurrencies for Highly Accurate Predictions. *IEEE Systems Journal*, 14(1):321–332.

Giulia Serafini, Ping Yi, Qingquan Zhang, Marco Brambilla, Jiayue Wang, Yiwei Hu, and Beibei Li. 2020. Sentiment-Driven Price Prediction of the Bitcoin based on Statistical and Deep Learning Approaches. In *2020 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, Glasgow, United Kingdom. IEEE.