A Review on Forecasting Machine Learning Models for Predicting Crypto Prices

Vicky Soni[1], Prof. Anil Malviya[2]

Department of IT, PCST, Indore[1,2]

Abstract: Crypto price prediction is a category of time series prediction which extremely challenging due to the dependence of crypto prices on several financial, socio-economic and political parameters etc. Moreover, small inaccuracies in crypto price predictions may result in huge losses to firms which use crypto price prediction results for financial analysis and investments. Conventional statistical methods render substantially lesser accuracy compared to new age machine learning techniques. This machine learning based techniques are being used widely for crypto price prediction due to relatively higher accuracy compared to conventional statistical techniques. This paper presents a review on contemporary data driven approaches for crypto currency forecasting highlighting the salient attributes. Moreover, the identified non-trivial research gap in the existing approaches has been used as an underpinning for subsequent direction of research in the domain. The paper culminates with the performance metrics and concluding remarks.

Keywords: Crypto price Forecasting, Data Driven Models, Regression Analysis, Machine Learning, Performance Metrics.

1. Introduction

With increasing digitization and resource distribution, cryptocurrencies have gain significant importance. This has led to large scale investments in cryptocurrencies such as Bitcoin, Etherium etc.. However, crypto prices are extremely random, fluctuating and volatile in nature which makes investments risk prone. Moreover, previous crypto data often exhibits random fluctuations, volatility and deviation from a particular trend, which is often termed as noise [2]. This noisy behavior makes pattern recognition difficult leading to inaccuracies in forecasting results. Hence, it is necessary to filter out the baseline noise from the time series crypto data prior to applying the data to any machine learning or deep learning model for pattern recognition [3]. While crypto trend analysis is a time series regression problem, what makes it extremely complicated is the dependence on several non-numeric parameters such as socio economic conditions, political conditions of a country, political stability, financial crisis and trade wars, global slowdown and public sentiments pertaining to a company etc. This leads to variabilities in the stock trends often exhibiting non-coherent patterns with respect to historical data [4]. Figure 1.1 depicts the common crypto currencies.

Fig.1 Common Crypto Currencies

Global events, such as natural disasters, geopolitical tensions, or pandemics, can also impact crypto trends [5]. These events can create uncertainty in the markets and cause investors to become more risk-averse and hence it is essential to note that data trends are inherently variable and can be influenced by a wide range of factors [6]. It's also worth mentioning that past performance does not guarantee future results, so investors should exercise caution when making investment decisions based on historical trends [7]. Cryptocurrency prediction is basically a time series prediction problem. Mathematically

$$P = f(t, v) \qquad\qquad (1)$$

Here,

P represents crypto price

f represents a function of

t is the time variable

v are other influencing global variables

The dependence of crypto prices over time makes it somewhat predictable under similar other conditions of global influencing variables. However, even the slightest of changes can derail the prediction completely. Hence crypto market prediction extremely challenging due to the dependence of crypto prices on several financial, socio-economic and political parameters etc. Moreover, small inaccuracies in crypto price predictions may result in huge losses to firms which use crypto price prediction results for financial analysis and investments. Hence it has remained a serous research problem for researchers since [8]:

1) Inaccuracies in crypto market prediction can lead to large financial losses for individuals and firms aiming to leverage crypto market data.
2) It is a complex time series problem
3) It is difficult to find trends in such diverse parameters which affect crypto prices [9].

Crypto price movement is often very random and dynamic in nature; the prices are varying in fraction of seconds [10]. Additionally each time the changes in prices are unpredictable. These changes can be positive or can be negative according to the market outlooks [11].

II. Factors Affecting Crypto Price

A crypto price may depend on several factors operating in the current world and crypto market. We will try to take into account a combination of mainly two factors [12]:

- The impact of global influencing factors.
- The past performances and records of the target company.

Therefore in this proposed study the prediction of crypto price influencing feature is the key aim and objective. Additionally it is also required to utilize these features during the time prediction to improve the prediction accuracy of the systems [12]. Therefore the proposed work involves the study of machine learning and data mining techniques (supervised and unsupervised) by which the prediction is feasible. In addition of that need to implement the additional methodology that accurately analyze the crypto market influencing features and involve these factors to reduce the error in prediction data (i.e. error minimization techniques or optimization techniques) [13].

Finally after implementation of the proposed technique, it is required that the error rate has to be squeeze out to justify the proposed work. Therefore the proposed work involves the comparative study of the proposed technique with the similar available techniques [14]. In addition of a case study with a company crypto prices also involved with the proposed study work. This section provides the basic overview of the proposed crypto price prediction study the next section provides the core aim and objectives of the proposed work. Statistical techniques do not render high accuracy of prediction and hence are ineffective in prediction problems which need low errors in prediction [15].

III. Related Work

This section presents the existing work in the domain:

Rafi et al. proposed a price forecasting model based on three vital characteristics (i) a feature selection and weighting approach based on Mean Decrease Impurity(MDI) features. (ii) Bi-directional LSTM and (iii) with a trend preserving model bias correction (CUSUM control charts for monitoring the model performance over time) to forecast Bitcoin and Ethereum values for long and short term spans. On a new test-set collected from January 01, 2020 to January 01, 2022 for the two cryptocurrencies we obtained an average RSME of 9.17, with model bias correction, Comparing with the prevalent forecasting models we report a new state of the art in cryptocurrency forecasting.

Kim et al. proposed a novel framework that predicts the price of Bitcoin (BTC), a dominant cryptocurrency. For stable prediction performance in unseen price range, the change point detection technique is employed. In particular, it is used to segment time-series data so that normalization can be separately conducted based on segmentation. In addition, on-chain data, the unique records listed on the blockchain that are inherent in cryptocurrencies, are collected and utilized as input variables to predict prices. Furthermore, this work proposes self-attention-based multiple long short-term memory (SAM-LSTM), which consists of multiple LSTM modules for on-chain variable groups and the attention mechanism, for the prediction model. Experiments with real-world BTC price data and various method setups have proven the proposed framework's effectiveness in BTC price prediction. The results are promising, with the highest MAE, RMSE, MSE, and MAPE values of 0.3462, 0.5035, 0.2536, and 1.3251, respectively.

Shahbazi et al. showed that during recent developments, cryptocurrency has become a famous key factor in financial and business opportunities. However, the cryptocurrency investment is not visible regarding the market's inconsistent aspect and volatility of high prices. Due to the real-time prediction of prices, the previous approaches in price prediction doesn't contain enough information and solution for forecasting the price changes. Based on the mentioned problems in cryptocurrency price prediction, we proposed a machine learning-based approach to price prediction for a financial institution. The proposed system contains the blockchain framework for secure transaction environment and Reinforcement Learning algorithm for analysis and prediction of price. The main focus of this system is on Litecoin and Monero cryptocurrencies. The results show the presented system accurate the performance of price prediction higher than another state-of-art algorithm.

Ertz et al. proposed This study highlights the potential impacts of blockchain technology on the collaborative economy (CE), colloquially known as the sharing economy. This conceptual review first analyzes how the CE intersects with the blockchain technology. Collaborative consumption involves an intensification of peer-to-peer trade, underpinned by robust digital infrastructures and processes, hence an increased use of new technologies and a redefinition of business activities. As an inherently connected economy, the CE is, therefore, prone to integrating the most recent technological advances including artificial intelligence, big data analysis, augmented reality, the smart grid, and blockchain technology. This review then furthers the examination of the organizational and managerial implications related to the use of blockchain technology in terms of governance, transaction costs, and user confidence.

Mudassir et al. proposed a high-performance machine learning-based classification and regression models for predicting Bitcoin price movements and prices in short and medium terms. In previous works, machine learning-based classification has been studied for an only one-day time frame, while this work goes beyond that by using machine learning-based models for one, seven, thirty and ninety days. The developed models are feasible and have high performance, with the classification models scoring up to 65% accuracy for next-day forecast and scoring from 62 to 64% accuracy for seventh–ninetieth-day forecast. For daily price forecast, the error percentage is as low as 1.44%, while it varies from 2.88 to 4.10% for horizons of seven to ninety days. These results indicate that the presented models outperform the existing models in the literature.

Gyamerah et al. proposed that the uncertainties in future Bitcoin price make it difficult to accurately predict the price of Bitcoin. Accurately predicting the price for Bitcoin is therefore important for decision-making process of investors

and market players in the cryptocurrency market. Using historical data from 01/01/2012 to 16/08/2019, machine learning techniques (Generalized linear model via penalized maximum likelihood, random forest, support vector regression with linear kernel, and stacking ensemble) were used to forecast the price of Bitcoin. The prediction models employed key and high dimensional technical indicators as the predictors. The performance of these techniques were evaluated using mean absolute percentage error (MAPE), root mean square error (RMSE), mean absolute error (MAE), and coefficient of determination (R-squared). The performance metrics revealed that the stacking ensemble model with two base learner (random forest and generalized linear model via penalized maximum likelihood) and support vector regression with linear kernel as meta-learner was the optimal model for forecasting Bitcoin price. The MAPE, RMSE, MAE, and R-squared values for the stacking ensemble model were 0.0191%, 15.5331 USD, 124.5508 USD, and 0.9967 respectively. These values show a high degree of reliability in predicting the price of Bitcoin using the stacking ensemble model. Accurately predicting the future price of Bitcoin will yield significant returns for investors and market players in the cryptocurrency market.

Huang et al. examine whether bitcoin returns are predictable by a large set of bitcoin price-based technical indicators. Specifically, authors construct a classification tree-based model for return prediction using 124 technical indicators. Authors provide evidence that the proposed model has strong out-of-sample predictive power for narrow ranges of daily returns on bitcoin. This finding indicates that using big data and technical analysis can help predict bitcoin returns that are hardly driven by fundamentals.

Adcock et al. showed that Bitcoin is the largest cryptocurrency in the world, but its lack of quantitative qualities makes fundamental analysis of its intrinsic value difficult. As an alternative valuation and forecasting method we propose a non-parametric model based on technical analysis. Using simple technical indicators, we produce point and density forecasts of Bitcoin returns with a feedforward neural network. We run several models over the full period of April 2011–March 2018, and four subsamples, and we find that backpropagation neural networks dominate various competing models in terms of their forecast accuracy. We conclude that the dynamics of Bitcoin returns is characterized by predictive local non-linear trends that reflect the speculative nature of cryptocurrency trading.

Phillipas et al. showed that Bitcoin is a widely accepted payment system, among the so-called cryptocurrencies. This letter examines the jump intensity of Bitcoin prices, partially attributed to increasing media attention in social networks. Over the last decade that Bitcoin has been traded, many alterations have taken place from exchanges to the likelihood of closure. Nevertheless, the Bitcoin has unique default benefits and properties by its structure. It is fully decentralized and depends on a sophisticated cryptographic protocol that it is difficult to counterfeit It also has the benefits of security and anonymity for investors because banks, governments, or organizations do not issue it. Moreover, forecasting of Bitcoin prices is critically important for potential investors.

Shen et al. examines the link between investor attention and Bitcoin returns, trading volume and realized volatility. Unlike previous studies, authors employ the number of tweets from Twitter as a measure of attention rather than Google trends as we argue this is a better measure of attention from more informed investors. Authors find that the number of tweets is a significant driver of next day trading volume and realized volatility which is supported by linear and nonlinear Granger causality tests.

IV. Noteworthy Contribution in the Domain

The noteworthy contributions in the field of the proposed work can be summarized as:

1) Neural Networks: One of the most commonly adopted methodologies in the domain of crypto market forecasting is the use of neural networks which mimic the learning of the human brain. The most common neural network categories used off late have been:
   a) Long Short Term Memory (LSTM) neural networks: These neural networks are capable to weight how much

weightage is to be given to the recent data and the data of the far past. It is often shown to work effectively for time series prediction [16]-[17]..
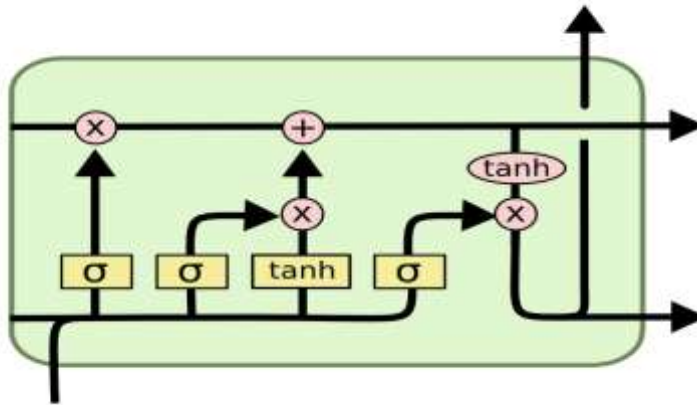

Fig.2 The Structure of the LSTM Neural Network

2)   Back Propagation: This neural network structure has a feedback path for errors in prediction. It is often effective in time series prediction problems since the estimation of future samples is based on both previous data as well as the errors in the previously prediction samples [18].
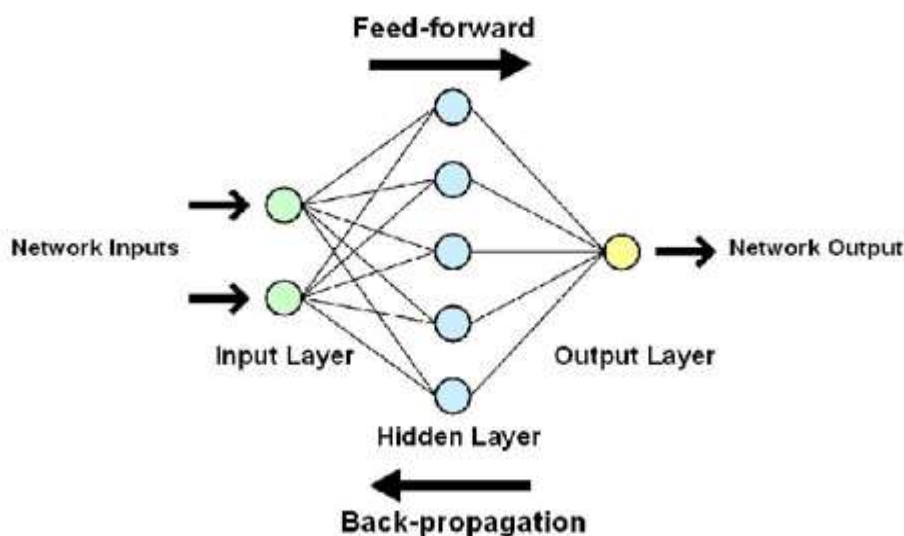

Fig.3 The Structure of the BP Neural Network

3)   Regression Models: A variety of regression models have been used for time series prediction among which the most common ones are:
   a)   Linear
   b)   Non-Linear
   c)   Polynomial
   d)   SVR (Support Vector Regression)

In this approach, the relationship between the independent and dependent variable is found using different mathematical or statistical models depending on the nature of the approach used [19]-[20]. The most common type of regression model can be thought of as the linear regression model which is mathematically expressed as [21]:

$$y = \theta_1 + \theta_2 x \qquad\qquad (2)$$

Here,

x represenst the state vector of inut variables

y rperesenst the state vector of output variable or variables.

Θ1 and Θ2 are the co-efficents which try to fit the regression learning models output vector to the input vector.

The aim of the approach is to attain the best fit regression line which is equivalent to saying that the co-efficient values θ1 and θ2 , should be adjusted such that to minimize the error between predicted y value (pred) and true y value (y). The cost function J is mathematically defined as [22]:

$$J = \frac{1}{n}\sum_{i=1}^{n}(pred_i - y_i)^2 \qquad\qquad (3)$$

Here,

n is the number of samples

y is the target

pred is the actual output.

While machine learning models offer promising results in cryptocurrency price forecasting, several challenges remain [23].

    1. The unpredictable nature of the crypto market, including sudden crashes and pumps, limits the accuracy of even the most advanced models.

    2. Cryptocurrency markets are highly non-linear, meaning price movements often don't follow a simple, predictable pattern. Traditional regression models like linear regression assume a linear relationship between variables, which may not hold true in the complex and volatile crypto market. As a result, these models struggle to capture the non-linear dependencies between features like historical prices, market trends, or trading volume, leading to poor prediction accuracy [24].

    3. Overfitting, where models become too tailored to historical data and fail to generalize to new data, is another significant issue.

    4. Cryptocurrency data is often sparse and noisy, which presents a significant challenge for regression models. Noisy data can mislead regression models, causing inaccurate predictions [25].

    5. Often sentiment analysis based exogenous inputs may be heavily biased, rigged or even the sources may be fake resulting in falsified training [26].

Conclusion:

It can be concluded that crypto price prediction is a category of time series prediction which extremely challenging due to the dependence of crypto prices on several financial, socio-economic and political parameters etc. Moreover, small inaccuracies in crypto price predictions may result in huge losses to firms which use crypto price prediction results for financial analysis and investments. Off late, soft computing techniques are being used widely for crypto market prediction due to relatively higher accuracy compared to conventional statistical techniques. This paper presents a comprehensive review on existing work in the domain of crypto price forecasting. Moreover, the existing challenges and research gaps have also been mentioned to aid future directions of research.

References

[1] M. Rafi, Q. A. K. Mirza, M. I. Sohail, M. Aliasghar, A. Aziz and S. Hameed, "Enhancing Cryptocurrency Price Forecasting Accuracy: A Feature Selection and Weighting Approach With Bi-Directional LSTM and Trend-Preserving Model Bias Correction," in IEEE Access, vol. 11, pp. 65700-65710, 2023

[2] G. Kim, D. -H. Shin, J. G. Choi and S. Lim, "A Deep Learning-Based Cryptocurrency Price Prediction Model That Uses On-Chain Data," in IEEE Access, vol. 10, pp. 56232-56248, 2022.

[3] Z. Shahbazi and Y. -C. Byun, "Improving the Cryptocurrency Price Prediction Performance Based on Reinforcement Learning," in IEEE Access, vol. 9, pp. 162651-162659, 2021

[4] M. Ertz and É. Boily, ''The rise of the digital economy: Thoughts on blockchain technology and cryptocurrencies for the collaborative economy,'' Int. J. Innov. Stud., vol. 3, no. 4, pp. 84–93, Dec. 2019.

[5] M. Mudassir et al., ''Time-series forecasting of Bitcoin prices using highdimensional features: A machine learning approach,'' Neural Comput. Appl., 2020.

[6] S. Asante Gyamerah, ''Are bitcoins price predictable? Evidence from machine learning techniques using technical indicators,'' 2019, arXiv:1909.01268.

[7] J.-Z. Huang, W. Huang, and J. Ni, ''Predicting bitcoin returns using highdimensional technical indicators,'' J. Finance Data Sci., vol. 5, no. 3, pp. 140–155, Sep. 2019.

[8] R. Adcock and N. Gradojevic, ''Non-fundamental, non-parametric bitcoin forecasting,'' Phys. A, Stat. Mech. Appl., vol. 531, Oct. 2019, Art. no. 121727.

[9] D. Philippas, H. Rjiba, K. Guesmi, and S. Goutte, ''Media attention and bitcoin prices,'' Finance Res. Lett., vol. 30, pp. 37–43, Sep. 2019.

[10] J. Abraham, D. Higdon, J. Nelson, and J. Ibarra, ''Cryptocurrency price prediction using tweet volumes and sentiment analysis,'' SMU Data Sci. Rev., vol. 1, no. 3, p. 1, 2018.

[11] D. Shen, A. Urquhart, and P. Wang, ''Does Twitter predict bitcoin?'' Econ. Lett., vol. 174, pp. 118–122, Jan. 2019.

[12] N. Aslanidis, A. F. Bariviera, and Ó. G. López, ''The link between cryptocurrencies and Google trends attention,'' Finance Res. Lett., vol. 47, Jun. 2022, Art. no. 102654, doi: 10.1016/j.frl.2021.102654.

[13] Y. B. Kim, J. G. Kim, W. Kim, J. H. Im, T. H. Kim, S. J. Kang, and C. H. Kim, ''Predicting fluctuations in cryptocurrency transactions based on user comments and replies,'' PLoS ONE, vol. 11, no. 8, Aug. 2016, Art. no. e0161197.

[14] D. Liang, F. Ma and W. Li, "New Gradient-Weighted Adaptive Gradient Methods With Dynamic Constraints," in IEEE Access, 2020, vol. 8, pp. 110929-110942.

[15] X. Huang, J. Guan, B. Zhang, S. Qi, X. Wang and Q. Liao, "Differentially Private Convolutional Neural Networks with Adaptive Gradient Descent," 2019 IEEE Fourth International Conference on Data Science in Cyberspace (DSC), Hangzhou, China, 2019, pp. 642-648.

[16] AC Wilson, R Roelofs, M Stern, "The marginal value of adaptive gradient methods in machine learning", Proceedings in Advances in Neural Information Processing Systems 30 (NIPS 2017).

[17] Y Xue, Y Wang, J Liang, "A self-adaptive gradient descent search algorithm for fully-connected neural networks", Neurocomputing, Elsevier 2022. Vol.478, pp. 70-80.

[18] I Khandelwal, R Adhikari, G Verma, "Time series forecasting using hybrid ARIMA and ANN models based on DWT decomposition", Procedia in Computer Science, vol.48, pp. 173-179.

[19] A Butorova, E Baglaeva, I Subbotina, "Application of the Wavelet Data Transformation for the Time Series Forecasting by the Artificial Neural Network", Proceedings in New Trends in the Applications of Differential Equations

[20] V. Buterin. (2013). Ethereum White Paper: A Next Generation Smart Contract & Decentralized Application Platform. [Online]. Available: https://github.com/ethereum/wiki/wiki/White-Paper

[21] D. Vujicic, D. Jagodic, and S. Randic, ''Blockchain technology, bitcoin, and ethereum: A brief overview,'' in Proc. 17th Int. Symp. INFOTEHJAHORINA (INFOTEH), Mar. 2018, pp. 1–6.

[22] S. Karasu, A. Altan, Z. Saraç and R. Hacioğlu, "Prediction of Bitcoin prices with machine learning methods using time series data," 2018 26th Signal Processing and Communications Applications Conference (SIU), Izmir, Turkey, 2018, pp. 1-4.

[23] Y. Indulkar, "Time Series Analysis of Cryptocurrencies Using Deep Learning & Fbprophet," 2021 International Conference on Emerging Smart Computing and Informatics (ESCI), Pune, India, 2021, pp. 306-311.

[24] M. Saad, J. Choi, D. Nyang, J. Kim and A. Mohaisen, "Toward Characterizing Blockchain-Based Cryptocurrencies for Highly Accurate Predictions," in IEEE Systems Journal, 2020, vol. 14, no. 1, pp. 321-332.

[25] P. Jay, V. Kalariya, P. Parmar, S. Tanwar, N. Kumar and M. Alazab, "Stochastic Neural Networks for Cryptocurrency Price Prediction," in IEEE Access, 2020, vol. 8, pp. 82804-82818.

[26] L. Duan, F. Duan, F. Chapeau-Blondeau and D. Abbott, "Noise-Boosted Backpropagation Learning of Feedforward Threshold Neural Networks for Function Approximation," in IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1-12, 2021, Art no. 1010612.