A Survey on Data Driven Models for Crypto Price Forecasting

ANKITA SONI[1], DR. SACHIN PATEL[2]
[1, 2] *SAGE University Indore*

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.

Index Terms— Crypto price Forecasting, Data Drien Models, Regression Analysis, Machine Learning, Performance Metrics.

I. INTRODUCTION

Cryptocurrency price forecasting is a challenging task due to the volatile and unpredictable nature of the crypto market. Machine learning (ML) models have gained popularity for predicting cryptocurrency prices, offering the ability to process large datasets and uncover patterns in price movements that may be difficult for traditional models to detect. Various ML techniques such as time series analysis, supervised learning, and deep learning have been explored to develop accurate prediction models. This essay delves into the prominent machine learning models used for cryptocurrency price forecasting. 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 [1]. 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].

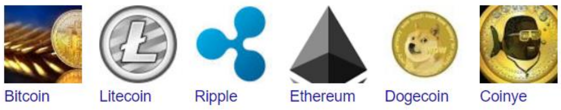

Fig.1 Common Crypto Currencies

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:

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

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 process 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 [8].

Statistical techniques are not found to be as accurate as the contemporary artificial intelligence and machine learning based approaches [9]. In this paper, a back propagation based scaled conjugate gradient algorithm is used in conjugation with the discrete wavelet transform (DWT) for forecasting crypto price trends. The evaluation of the proposed approach has been done based on the mean absolute percentage error (MAPE). A comparative MAPE analysis has also been done w.r.t. previously existing techniques [10].

II. MACHINE LEARNING AND DEEP LEARNING MODELS

Machine Learning and Deep learning have evolved as one of the most effective machine learning techniques which has the capability to handle extremely large and complex datasets [11]. The most common models used for prediction of crypto prices are:

Time Series Analysis Models: One of the most common approaches in cryptocurrency forecasting is time series analysis, which involves using historical data to predict future values. Autoregressive Integrated Moving Average (ARIMA) is a classical time series model that has been widely applied to cryptocurrency forecasting. ARIMA models assume that future price trends can be explained by past price patterns, but their simplicity may limit accuracy when dealing with high volatility. More advanced models, like Long Short-Term Memory (LSTM) networks, a type of recurrent neural network (RNN), address these limitations by learning long-term dependencies and capturing complex temporal dynamics in price movements [12].

Supervised Learning Models: Supervised learning models, such as regression models and decision trees, are also employed in cryptocurrency forecasting. Linear regression models predict prices based on a combination of features like historical prices, trading volume, and market sentiment. However, linear models may fall short in handling the non-linearity of the crypto market. Decision tree-based models like Random Forests and Gradient Boosting Machines (GBMs) offer more flexibility by capturing non-linear relationships between input features and price outcomes. These models work by building multiple decision trees that collectively improve the prediction accuracy [13].

Neural Networks and Deep Learning: Deep learning models, particularly neural networks, have become popular for cryptocurrency price prediction because of their ability to learn from large, complex datasets. Convolutional Neural Networks (CNNs) and LSTMs are often used in tandem for price forecasting, with CNNs extracting important features from the data and LSTMs modeling sequential dependencies in time series. More advanced architectures like Transformer models, which excel at sequence prediction tasks, are being increasingly explored for predicting crypto prices with a higher degree of precision [14]

Exogenous Data Integration: In addition to historical price data, external factors such as social media sentiment, news, and macroeconomic indicators significantly influence cryptocurrency prices. Machine learning models can incorporate these additional sources of information to improve forecasting accuracy. For instance, sentiment analysis models using natural language processing (NLP) techniques analyze social media posts or news articles to gauge market sentiment and predict price movements. Combining sentiment data with traditional time series or neural network models has been shown to enhance the predictive power of ML models [15].

III. LITERATURE REVIEW

The existing contemporary research in the domain is presented next:
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. examined 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. EXISTING CHALLENGES

The research gaps and existing challenges identified are:

While machine learning models offer promising results in cryptocurrency price forecasting, several challenges remain [16].

- The unpredictable nature of the crypto market, including sudden crashes and pumps, limits the accuracy of even the most advanced models.
- 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 [17].
- Overfitting, where models become too tailored to historical data and fail to generalize to new data, is another significant issue.
- Cryptocurrency data is often sparse and noisy, which presents a significant challenge for regression models. Noisy data can mislead regression models, causing inaccurate predictions.
- Often sentiment analysis based exogenous inputs may be heavily biased, rigged or even the sources may be fake resulting in falsified training [18].

The final performance metrics computed for system evaluation are:

1) Mean Absolute Percentage Error (MAPE)

$$MAPE \ = \frac{100}{M} \sum_{t=1}^{N} \frac{E - E_t|}{E_t} \qquad (2)$$

Here E_t and E_t˜ stand for the predicted and actual values respectively.

The number of predicted samples is indicated by M.

2) Regression

The extent of similarity between two variables is given by the regression.

The cost function J is typically computed as:

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

Here,
n is the number of samples
y is the target
pred is the actual output.

The performance of every model should be however tested and validated across a wide variery of authentic datasets so as to gauge the performance of the system [19].

CONCLUSION

It can be concluded from previous discussions that crypto price prediction is a category of time series prediction with high sensitivity and dependence on external factors. Hence it is often challenging to attain high levels of accuracy in prediction. Despite their simplicity and widespread use, regression models face significant challenges when applied to cryptocurrency price prediction. The non-linearity, volatility, and noise of the crypto market, along with issues like overfitting and the difficulty in selecting appropriate features, often limit the effectiveness of these models. To overcome these challenges, more sophisticated techniques like decision trees, neural networks, or hybrid approaches that integrate external data sources and handle complex market dynamics are typically needed.

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

[6] 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.

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

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

[9] 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.

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

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

[12] 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.

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

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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).

[19] A System for MANET to detect selfish nodes using NS2S, D Padiya, R Pandit, S Patel Int. J. Eng. Sci. Innov. Technol 1 (2), 25-30.