Grenze International Journal of Engineering and Technology, Jan Issue

Machine Learning based Framework to Predict the Bitcoin Price

Sujana S[1] and Dr. Bhat Geetalaxmi Jairam[2]
[1-2]Department of Information Science andEngineering, The National Institute of Engineering, MysoreKarnataka, India
Email: 2021scn_sujanas_a@nie.ac.in, bgj@nie.ac.in

***Abstract*—Bitcoin has recently attracted a lot of media and public attention as a result of its recent price boom and crash. Similarly, numerous researchers have explored many aspects influencing the Bitcoin price and the patterns underlying its fluctuations, specifically utilizing various machine learning methods. In this work, authors investigate and analyze various cutting-edge machine learning algorithms for Bitcoin price prediction, such as a logistic regression and a long short-term memory (LSTM) model. Although LSTM-based prediction models outperformed other prediction models for Bitcoin priceprediction (regression), a simple profitability analysis revealed that classification models were more effective than regression models for algorithmic trading. Overall, the suggested ML learning-based prediction models performed similarly. This work performs an in-depth investigation on the evolution of Bitcoin, as well as a thorough review of several machine learningmethods used for price prediction. It's included is a Bitcoin priceprediction. The Existing work fails to predict day to day bitcoinprice changes. And has given less accuracy compared to the proposed work. Authors have overcome this limitation in the proposed work by predicting the day to day BTC price for the upcoming month.**

***Index Terms*— Bitcoin, machine learning, crypto currency.**

I. INTRODUCTION

Bitcoin is a crypto currency that is utilized all over the world for investing or for making digital payments [1]. Bitcoin is decentralized, meaning that nobody owns it. Bitcoin transactions are simple because they are not governed by anyone nation. The term "bitcoin exchanges" refers to a variety ofonline markets where investments can be made. These enablethe buying and selling of bitcoins using various currencies. Mt Gox is the biggest Bitcoin exchange. A digital wallet, which functions somewhat similarly to a virtual bank account, is where bitcoins are kept. A site called Blockchain is where the timestamp data and transaction history are kept. A block is a single entry in a blockchain. A pointer to the previous block of data is present in every block.
Blockchain data is encrypted. The user's name is not made public during transactions; just their wallet ID is. The bankingsector has undergone a change thanks to the decentralized digital money known as Bitcoin [2-4]. Bitcoin has drawn the attention of investors, traders, and researchers who want to comprehend and forecast its price movements because of its decentralized nature and potential for large rewards. Due to itsintrinsic volatility and the impact of several factors like market mood, trading volume, macroeconomic data, and regulatory changes, accurately predicting Bitcoin values is a difficult process. Figure 1 displays the history of bitcoin prices from 2018 through 2023.

Grenze ID: 01.GIJET.10.1.502_1

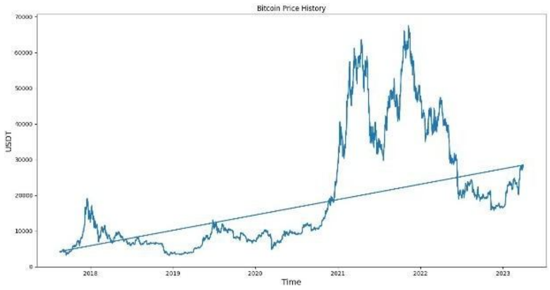


Fig. 1. Bitcoin Price history

The intricate linkages and temporal dependencies found in Bitcoin price data are frequently difficult to grasp using conventional financial forecasting techniques. On the other hand, machine learning approaches have demonstrated promise in time series data analysis and prediction due to theircapacity to recognise complex patterns and connections. [5].

The Long Short-Term Memory (LSTM) model, a kind of recurrent neural network (RNN) intended to handle sequence data, is one such potent machine learning approach. The LSTM model's capacity to learn long-term dependencies andadjust to shifting market dynamics makes it especially well- suited for predicting Bitcoin price [6]. The LSTM model maybe able to find patterns and relationships that conventional techniques might miss by taking into account past price data and pertinent attributes. Making informed decisions about risk management, market analysis, and investing strategies can be aided by these insights [7]. This study's objective is to create a model for predicting the price of bitcoin using machine learning methods, specifically the LSTM model and LogisticRegression.

The capacity of the LSTM model to capture temporal dependencies and complicated interactions between Bitcoin price data and other influencing factors is one of our goals. Authors may test the model's effectiveness in predicting futureprice movements by training it on previous data and evaluating its performance.

II. Literature Work

When considering the problem of Bitcoin price formation, the literature mostly consists of empirical works to analyze the determinants.

Ladislav Kristoufek [8] first viewed the Bitcoin market as entirely speculative, with no fundamentalists, and assessed the relationship between Bitcoin and Wikipedia and Google Trends search volume. His findings revealed a substantial asymmetry in the relationship between Bitcoin price and search queries, implying that speculation and trend-chasing drive Bitcoin price dynamics in the cryptocurrency market.

Adam Hayes [10] used cross-sectional empirical data toexamine 66 "coins" to determine what factors drive the valueof cryptocurrencies in the technical area, and he identified three main drivers: the aggregate computational power used in mining for units of cryptocurrency, the rate of unit production, and the cryptologic algorithm used for the protocol.

In [10] did an empirical investigation on the price drivers of Bitcoin within the framework of a Barro [11] model by implementing time-series analytical processes with daily Bitcoin data. Their findings found that, among the traditional determinants and elements related to digital currencies, market forces and investor attractiveness influence Bitcoin price the most strongly, indicating that macro-level financialchanges are the long-term drivers of Bitcoin.

Balcilar et al. discovered that volume can predict Bitcoin returns using a non-parametric causality-in-quantiles test,implying the need of modelling nonlinearity and accounting for tail behaviour [12].

The authors of propose a taxonomy for blockchains andsystems based on them in [13]. The taxonomy seeks to make comparisons and analyses of blockchain-based software systems easier by addressing important architectural characteristics and the impacts of various decisions. Thetaxonomy aids in architectural discussions on performance, quality, and the overall evaluation of blockchain-based systems. The report also cites blockchain technology's decentralised nature and its potential for facilitating shared state agreement without relying on a central integration point.Although financial transactions are emphasised, other applications are also covered. The first section of the study discusses data collection, normalisation, and smoothing approaches for Bitcoin price data.

The authors provide a methodology for analysing a "stress test" denial-of-service (DoS) attack against Bitcoin using spam. Their cluster-based research suggests that during the peak of the spam campaign, a considerable share (23.41%) of Bitcoin transactions were classified as spam. For non- spam transactions, the attack raised transaction prices and processing time. The report emphasises the importance of conducting research into Bitcoin transaction spam filtering technologies as well as DoS mitigation strategies. Following DoS attacks on Bitcoin, several tactics such as "money drops" and transaction malleability are mentioned [14].

The author's main goal is to create a machine learning method (using random forests and generalised linear models) to forecast the future price of bitcoin. The gathered dataset contains 25 different features over five years. It was discovered that they have reached an accuracy of 50 to 55 percent when anticipating future prices. Data was collected for ten minutes [15].

Socially created notions about virtual money on Twitter, according to the authors [16], have a direct or indirect impact on all market evaluations of virtual currencies. The goal of this study is to use sentiment analysis to forecast the fluctuating value of bitcoin and to determine the relationshipbetween positive and negative thoughts. In bitcoin priceprediction, several attribute selection processes and machinelearning approaches such as artificial neural network (ANN),support vector machine (SVM), recurrent neural network (RNN), and k-means clustering were used to acquire the mostessential features [17].

Using a Bayesian optimised recurrent neural network and LSTM, authors predicted the direction of the Bitcoin price inUSD. They also used the ARIMA model to compare deep learning approaches [18].

Based on the cryptocurrency market and macromarket index(stock market index, crude oil price, exchange rate, etc.) andsearch index, the authors developed a total of 40 explanatoryfactors for Bitcoin price prediction. SDAE algorism outperforms BPNN, PCA-SVR, and SVR in prediction performance [19].

[20] predicted the price of Bitcoin using four deep learning algorithms (Theil-Sen regression, Huber regression, LSTM, and GRU). The LSTM algorithm has the highest accuracy (52.78%). According to the same explanatory variables.

In this work, the authors included the factors difficulty and hash rate connected to Bitcoin features; the 52.78% prediction accuracy of LSTM is also greater than the accuracy of RNN and ARIMA. [7] predicted the Bitcoin price using four algorithms: LSTM, SVR, ANFIS, and ARIMA. Chen included eight different types of Bitcoin attribute factors, public attention variables (Google Trends and Twitter data), and economic category variables. LSTM performed better than the other three models in all foursubsample periods.

The following table 1gives a comparison about the algorithm discussed in various paper and the accuracy determined by them:

TABLE I. COMPARISON OF VARIOUS STUDY

Sl No	Paper Title	Algorithm	Accuracy
1.	Deep Learning EthereumToken Price Prediction with Network with motif Analysis.[8]	LSTM	55%
2.	Ether Price Prediction Using Advanced Deep Learning Models.[10]	LSTM	70%
3.	Bitcoin price prediction using ARIMA model.[11]	LSTM	52%
4.	Forecasting Price of Cryptocurrencies Using Tweets Sentiment Analysis. [12]	LSTM	52.78%
5.	Proposed model has got the accuracy of 98 % with the use ofLSTM		
6.	Bitcoin Price Prediction: AMachine Learning SampleDimension Approach. [15]	Logistic Regression	64.8%
7.	Statistical Machine Learningfor Bitcoin Prediction. [17]	Logistic Regression	54.0%
8.	Proposed model got the accuracy of 56% with the use ofLogistic Regression		

The Figure 2 refers to the comparison chart considering the LSTM and Logistic Regression algorithms that are mentioned in this paper. The bar chart is referring to the above table 1 which is mentioned with the paper title, algorithm used and the accuracy mentioned in each of them. The chart is been designed in two colors blue and brown to differentiate our work with the comparison of other papers. The accuracy of LSTM in this work in 98% and Logistic Regression is 56% which are seen in the above graph.

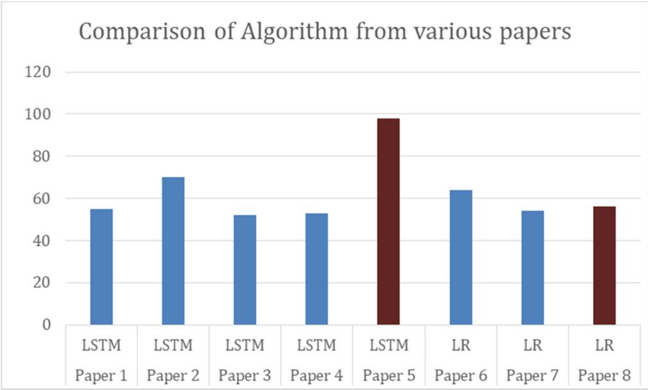


Fig. 2. Bar chart comparing the LSTM and Logistic Algorithms fromvarious papers

III. METHODOLOGY

The proposed methodology uses two separate deep learning-based prediction models to forecast daily bitcoin prices by recognising and analyzing key features. Authors can identify which model is much more accurate for the future fulfillment of our aim after applying both models for bitcoin prediction and selecting relevant parameters to attaina higher performance. In this proposed paper, authors have used deep learning mechanisms including LSTM and Logistic Regression, which are the most recent and efficient strategies for forecasting bitcoin prices. Because bitcoin is the most popular cryptocurrency, the price fluctuation issue should be resolved quickly. Figure 3 depicts the prediction process, from data collection through bitcoin price predictions.

A. Dataset

Machine learning models in this study make use of a dataset collected via the binance package's api get_historical_klines. The following parameters are passed tothe API get_historical_klines: symbol, interval, start and endtime. The symbol represents bitcoin's name. It returns a list of OHLC values, including the following: Open time, Open,High, Low, Close, Volume, Close time, Quote asset volume,Number of trades, Taker buy base asset volume, and Taker buy quote asset volume. The data for our prediction task is acquired from 1 January 2010 to the most current month, andthe output from the get_historical_klines function is saved asa csv file and utilized as the input dataset. CSV files contain the datasets.

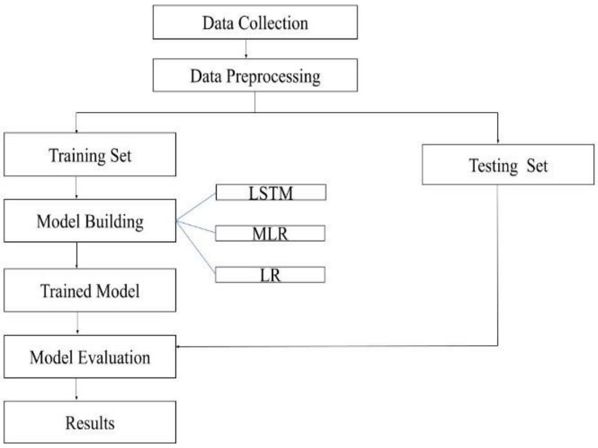


Fig. 3. Architecture of BTC price prediction

B. LSTM

This study's machine learning models make use of a dataset obtained via the binance package's api get_historical_klines. The API get_historical_klines takes the following parameters: symbol, interval, start and end time. The symbol depicts the name of bitcoin. It returns a list of OHLC values such as Open time, Open, High, Low, Close, Volume, Close time, Quote asset volume, Number of trades, Taker buy base asset volume, and Taker buy quote asset volume. The data for our prediction task is collected from January 1st, 2010 to the present month, and the output of theget_historical_klines function is stored as a csv file and usedas the input dataset. The datasets are stored as CSV files.

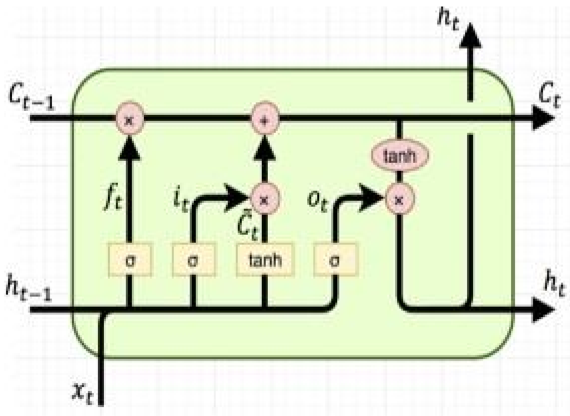


Fig. 4. The structure of LSTM

C. Logistic Regression

Logistic Regression is a straightforward machine learning technique used to solve categorization problems. It is a predictive analysis tool that validates the probability idea. It employs a complicated cost function known as the 'Sigmoid function', sometimes known as the 'logistic function,' rather than a linear function. The logistic regression equation can bedefined by the following equation.

$$\pi(\mathbf{X}) = \frac{\exp(\beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k)}{1 + \exp(\beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k)}$$

The Accuracy, Precision and Recall for the proposed workcan be calculation using following formulas:

$$\text{accuracy} = \frac{\text{correctpredictions}}{\text{allpredictions}} \quad (1)$$

$$\text{precision} = \frac{\text{truepositives}}{\text{truepositives} + \text{falsepositives}} \quad (2)$$

$$\text{recall} = \frac{\text{truepositives}}{\text{truepositives} + \text{falsenegatives}} \quad (3)$$

IV. Results and Discussions

Authors can anticipate Bitcoin prices for the next 30 days now that authors have an LSTM model that has been trainedon historical data. According to the statistics that authors have used to develop the model, the final historical price forBitcoin is on March 31, 2023. Authors are now going beyondthat exact day in order to anticipate Bitcoin prices for the next30 days. It should be noted that authors are estimating the price for the following day using data from the previous day.The lookback time is set to five days in this case, implying that authors are estimating the price of bitcoin for the following day using data from the previous thirty days.
Figure 5 depicts the Logistic Regression confusion matrix forbitcoin price prediction. Figure 5 shows that if both the actualand anticipated values are No, the result is a True Negative. If the actual value is less than the projected value, the result is a False Positive. If the actual value is greater than the anticipated value, the result is a False Negative. If both the actual and anticipated values are yes, the result is a True Positive.
Figure 6 depicts the graphical representation of accuracy which is obtained by LSTM in predicting the bitcoin price. In predicting the bitcoin price the LSTM model is able to achieve 98% accuracy.
The graphical representation of Next 30 days Bitcoin price prediction of April month is depicted in Figure 7. Table 2 shows the accuracy comparison of proposed work with existing work. After comparison authors can conclude our LSTM model has achieved good accuracy compared to otherexisting models.

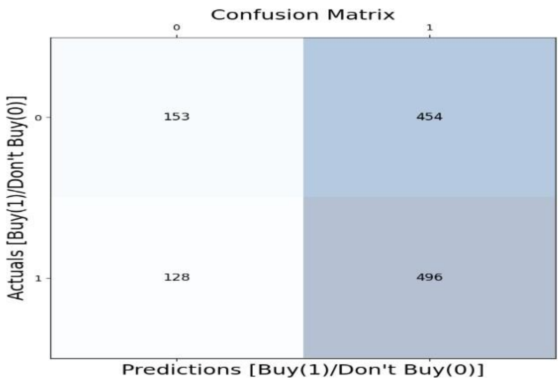


Fig. 5. The Confusion matrix of Logistic Regression

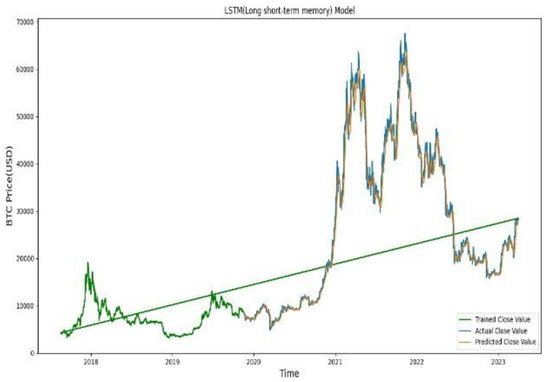


Fig. 6. The LSTM Accuracy

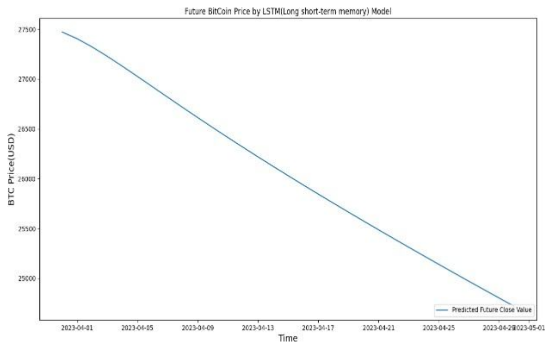


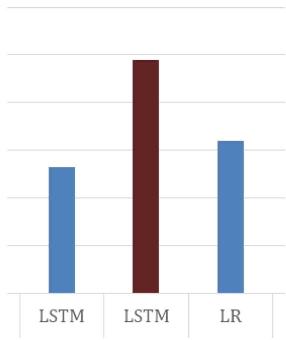
Fig. 7. The Future LSTM

V. COMPARISON ANALYSIS

Authors have compared the accuracy results of proposed work with the existing work. The existing work doesn't givemore accuracy using RNN with LSTM and Linear regressionalone. It has given the accuracy of 69%. In this work authorshave used LSTM which has given 98% accuracy and predictsday to day BTC values.

TABLE II. SUMMARIZE RESULTS

Models	Accuracy
Linear Regression Regression	69%
RNN with LSTM	72
Random Forest	55%
Logistic Regression (Proposed)	56%
LSTM (Proposed)	98%



x-axis: algorithms
Y axis: Accuracy level

VI. Conclusion

Two types of machine learning algorithms are built and utilised to predict prices in this paper. The accuracy of several models was tested using performance measures. The real and forecasted prices were then compared. The results reveal that LSTM outperformed the other algorithms by 98%. Based on the results, the LSTM model for the cryptocurrencies under consideration can be regarded efficient and dependable. This model is regarded as the best. The LSTM model has been shown to perform effectively with time series data. The results show that the accuracy of the LSTM model is significantly higher than that of logistic regression. LSTM models can provide useful insights into bitcoin price prediction, allowing investors to get a sense of market movements. The current approach is unable to forecast daily fluctuations in the price of bitcoin. and has provided results with less accuracy than the suggested job. In proposed work authors have circumvented this limitation by projecting the day-to-day BTC price for the upcoming month

Future Work

Authors in future can design a model which works for shorter time interval such as predicting bitcoin price for the next hour which this model fails to predict. Using the proposed work users can only predict the daily changes of BTC price for the upcoming one month. This can be improved by designing the LSTM model which can predict BTC price every hour. Future work can be modified to investigate other factors that may affect cryptocurrency market prices, with a particular focus on the effect that social media in general, and tweets, can have on cryptocurrency price and trading volume by analysing tweets using natural language processing techniques and sentiment analysis.

References

[1] Ukhopadhyay, U.; Skjellum, A.; Hambolu, O.; Oakley, J.; Yu, L.; Brooks, R. A brief survey of Cryptocurrency systems. In Proceedings of the 14th Annual Conference on Privacy, Security and Trust (PST), Auckland, New Zealand, 12–14 December 2016; pp. 745–752.

[2] Eyal, I. Blockchain Technology: Transforming Libertarian Cryptocurrency Dreams to Finance and Banking Realities. *Computer* **2017**, *50*, 38–49.

[3] Rehman, M.U.; Apergis, N. Determining the predictive power between cryptocurrencies and real time commodity futures: Evidence from quantile causality tests. *Resour. Policy* **2019**, *61*, 603–616.

[4] Edwin Sin , Lipo Wang.Bitcoin price prediction using ensembles of neural networks[10] .2017.publised 978-1-5386-2165-3/17 by IEEE.

[5] D. Shah and K. Zhang, "Bayesian regression and Bitcoin," in 52nd Annual Allerton Conference on Communication, Control, and Computing (Allerton), 2015, pp. 409-415.

[6] Karasu, S.; Altan, A.; Sarac, Z.; Hacioglu, R. Prediction of Bitcoin prices with machine learning methods using time series data. In Proceedings of the 26th Signal Processing and Communications Applications Conference (SIU), Izmir, Turkey, 2–5 May 2018.

[7] Jain, A.; Tripathi, S.; Dwivedi, H.D.; Saxena, P. Forecasting Price of Cryptocurrencies Using Tweets Sentiment Analysis. In Proceedings of the 11th International Conference on Contemporary Computing (IC3), Noida, India, 2–4 August 2018; pp. 1–7.

[8] L. Kristoufek, Bitcoin meets google trends and wikipedia: quantifying the relationship between phenomena of the

internet era, Sci. Rep. 3 (2013) 3415.
[9] L. Kristoufek, What are the main drivers of the Bitcoin price? evidence from wavelet coherence analysis, PLoS One 10 (4) (2015) e0123923.
[10] A. Hayes, What factors give cryptocurrencies their value: An empirical analysis, 2015.
[11] P. Ciaian, M. Rajcaniova, d.A. Kancs, The economics of bitcoin price formation, Appl. Econ. 48 (19) (2016) 1799–1815.
[12] R.J. Barro, Money and the price level under the gold standard, Econ. J.89 (353) (1979) 13–33.
[13] M. Balcilar, et al., Can volume predict bitcoin returns and volatility? aquantiles-based approach, Econ. Model. 64 (2017) 74–81.
[14] J. Bukovina, M. Martiček, Sentiment and bitcoin volatility, Mendel University in Brno, Faculty of Business and Economics, 2016.
[15] Y.B. Kim, et al., Predicting fluctuations in cryptocurrency transactions based on user comments and replies, Plos One 11 (8) (2016).
[16] D. Shah, K. Zhang, Bayesian regression and bitcoin, in: Communication, Control, and Computing (Allerton), 2014 52nd Annual Allerton Conference on, IEEE, 2014.
[17] Awoke, T., Rout, M., Mohanty, L., & Satapathy, S. C. (2020). Bitcoinprice prediction and analysis using deep learning models. In *Communication Software and Networks: Proceedings of INDIA 2019* (pp. 631-640). Singapore: Springer Singapore.
[18] Dutta, A., Kumar, S., & Basu, M. (2020). A gated recurrent unit approach to bitcoin price prediction. *Journal of Risk and Financial Management*, *13*(2), 23.
[19] Poongodi, M., Vijayakumar, V., & Chilamkurti, N. (2020). Bitcoin price prediction using ARIMA model. *International Journal of Internet Technology and Secured Transactions*, *10*(4), 396-406.