

Elevating Cryptocurrency Predictions: Bidirectional LSTM Methodology

Om Koli

Department of Computer Science and
Engineering, DYPCET
Kolhapur, India
omkoli00@gmail.com

Yash Nikam

Department of Computer Science and
Engineering, DYPCET
Kolhapur, India
yashnikam1804@gmail.com

Pranav Irlapale

Department of Computer Science and
Engineering, DYPCET
Kolhapur, India
pranavirlapale@gmail.com

Shrikrishna Katavare

Department of Computer Science and
Engineering, DYPCET
Kolhapur, India
katavare11@gmail.com

Rajvardhan Kesarkar

Department of Computer Science and
Engineering, DYPCET
Kolhapur, India
rajvardhankesarkar16@gmail.com

Abstract—The system proposed in this paper aims to predict cryptocurrency prices using Bi-Directional Long Short-Term Memory (LSTM), leveraging historical data obtained from Yahoo Finance and CoinGecko APIs. The goal is to assess LSTM models effectiveness in forecasting cryptocurrency prices and offer an interactive interface for users to visualize historical and forecasted prices. Several research works have been conducted on the prediction of cryptocurrency prices through various Deep Learning (DL) based algorithms. This project comprises two main approaches : one involves data analysis, LSTM modeling, and change point detection using Yahoo Finance data, while the other focuses on LSTM model training and price prediction using CoinGecko API data. The paper suggests that the prediction models it presents are useful for traders, investors, [6] and finance academics and are close to accurate at predicting the values of cryptocurrencies. Future research will examine more advanced deep learning architectures, primarily Transformer-based models like the GPT series, to improve pattern detection in bitcoin data. Integrating other data sources, such as sentiment analysis or blockchain measurements, may increase the accuracy of forecasting. With further research into cutting-edge techniques, cryptocurrency forecasting will get better and provide stakeholders with more information to help them make informed decisions.

Keywords— *Cryptocurrency ; forecasting ; Bi-Directional LSTM Model ; Time-series forecasting ; Machine learning*

I. INTRODUCTION

Nowadays, it seems like everyone is going through a digital transition. Unlike traditional currencies, cryptocurrency is based on cryptography. A virtual or digital currency used in financial systems is called cryptocurrency. Cryptocurrency, which has been around for years, has gained popularity and generated controversy due to innovative developments. Cryptocurrencies are a new type of digital currency that uses cryptography to secure the transaction process and prevent counterfeiting [19]. One important fact about cryptocurrencies is that they are independent of traditional banks, as they are not issued by any central authority, which makes them distinguishable from traditional centralized currencies [6]. But the tricky part is that the value of this money keeps changing, just like regular money does in different countries. So, predicting or figuring out how much a cryptocurrency will be worth in the future. Nowadays, cryptocurrency forecasting is widely regarded as one of the most difficult time-series prediction problems due to the large number of unpredictable factors involved and the

significant volatility of cryptocurrency prices, resulting in complicated temporal dependencies [18]. To find a better way to predict the future value of cryptocurrencies. To do this, we're using a smart computer technique called bidirectional long-term short memory (Bi-LSTM), which is like a super smart brain that learns from patterns and is capable of finding long-term as well as short-term hidden dependency sequential structures in data such as natural language[5]. Instead of making direct human investments, generating profit through algorithms is a common practice in the stock market. Several case studies have been conducted to conclude that mathematical models produce better results than humans. As cryptocurrency's history is relatively short as compared to the stock market, it's like exploring new and uncharted territories, where there's plenty to discover. Basically, both the stock market and cryptocurrency price data exhibit time series characteristics, but the latter is characterized by high volatility, with prices fluctuating dramatically. A cryptocurrency market differs from a traditional stock market in that the former includes many new features. We use a Long Short-Term Memory (LSTM) model to forecast cryptocurrency price trends. Our model is designed to predict the price of cryptocurrencies like Bitcoin and Ethereum for the next year. On the basis of this study, our future work will focus on improving the accuracy of the model, taking into consideration financial threats that may be experienced in the crypto world, in order to control the rise in such illegal activity.

Our main motivation comes from paper [1], and we used it as a base paper. The base paper has used the vanilla Bi-LSTM model, which is more efficient due to an improved time-saving and automated feature selection approach known as the Mean Decrease Impurity (MDI) approach, which selects optimal indicators completely by itself, resulting in highly accurate price forecasts and better results. In our study, the project comprised two approaches: the first involved data analysis, LSTM modelling, and change point detection; the second involved LSTM model training and cryptocurrency price forecasting. This paper is organized as follows: In Section II, we review existing studies on predicting cryptocurrency prices, covering various methods such as statistics, machine learning, and deep learning. In Section III, we explain our approach, starting with collecting and preparing data, then processing it before implementing and validating our models. Section IV showcases the outcomes of our experiments. Sections V and VI discuss the insights gained from the results and conclude our study, respectively.

II. RELATED WORK

Bitcoin is a decentralized cryptocurrency, a kind of digital currency that serves as the foundation for blockchain-based peer-to-peer financial transactions [2]. In recent years, investors have been paying close attention to the rise in Bitcoin prices. Literature has studied various metrics such as BTC-related historical technical indicators [2], [11] to investigate the predictability of BTC returns. Reference [1] leverage the LSTM model to forecast Bitcoin and Ethereum prices for short-term and mid-term. It implements a deep learning model to anticipate remarkably precise prices of Bitcoin and Ethereum at the end of the day, short-term (1, 7 and 14 days) and mid-term (30, 60 and 90 days). The wrapper and MDI techniques are the two choice of features strategies that are used. Using the same model, the indicators that the MDI technique identified produce better accuracy. For further improvisation, used the CUSUM control chart for monitoring the shift in the prices of both cryptocurrencies mentioned above. The overall accuracy for forecasting Bitcoin prices using these improvement measures was up to 73.3%. The RMSE and MAPE are as low as 6.67% and 0.66%, respectively. Ethereum price forecasting resulted in accuracy up to 54%. The RMSE and MAPE are as low as 3.03% and 2.98%, respectively. The results of performance evaluation metrics show significant improvement in the recent literature in predicting both, fluctuation in prices and daily closing price. In reference [4], it is studied that to forecast the Bitcoin price precisely considering different parameters that influence the Bitcoin price. In order to provide information regarding Bitcoin price patterns, this study first identifies the daily variations in the price of the cryptocurrency. All prices of the value of Bitcoin are included in the collection of data as of right now. The machine learning module is introduced to anticipate price values by utilizing the collected data. The aim of this work is to derive the accuracy of Bitcoin prediction using different machine learning algorithms and compare their accuracy. Reference [7] proposed (RNN) algorithms based on GRU, LSTM, and LSTM (bi-LSTM models). The results of experiments using regression and decision tree models are compared in order to predict the prices of different cryptocurrencies like Ethereum, Bitcoin, and Litecoin. The GRU model outperformed the long short-term memory (LSTM) and bidirectional LSTM (bi-LSTM) models in terms of prediction for all forms of bitcoin. With MAPE percentages of 0.2454%, 0.8267%, and 0.2116% for BTC, ETH, and LTC, respectively, GRU gives the most accurate prediction for LTC. Reference [13] explores the application of neural networks to predict financial market trends. The novel hybrid recurrent neural network (RNN) and its training algorithms aim to solve the prediction problem for time series with seasonal components. The proposed method is applied to various seasonal time series, demonstrating its effectiveness compared to other ANN models in the literature.

In reference [6] three types of deep learning techniques - LSTM, GRU, and Bi-LSTM, were used to predict the prices of three major cryptocurrencies, as measured by their market capitalization : Bitcoin, Ethereum, and Litecoin. The performance of the models was evaluated using two scores, RMSE and MAPE. The outcomes of the research

demonstrated that the GRU model and the Bi-LSTM model had the highest accurate forecasts for each of the three currencies. This suggests that the combination of forward and backward flows in bidirectional models improves the performance of time-series prediction. Reference [8] considered improving the forecasting performance and reliability of deep learning models utilizing three widely utilized ensemble strategies, i.e., averaging, bagging, and stacking. The authors utilized hourly prices of Bitcoin, Ethereum, and Ripple from 1 January 2018 to 31 August 2019. Furthermore, they used multiple Conv-based and LSTM-based learners as base models to perform a thorough performance evaluation of different ensemble models. Their study showed that, although there would be a large computational cost involved, deep learning and ensemble learning may be effectively modified to create reliable and trustworthy cryptocurrency prediction models.

III. METHODOLOGY

Implemented in phases, the project preprocesses data with MinMaxScaler, deploying PyTorch for LSTM modeling based on Yahoo Finance data and TensorFlow for similar tasks using CoinGecko API data. Trained on historical cryptocurrency prices, LSTM models predict future trends. The Streamlit library constructs a user-friendly web interface, facilitating the visualization of historical data, predicted prices, moving averages, and change points, enhancing accessibility and comprehension of cryptocurrency price forecasts. The MultiApp class manages the different pages and provides a sidebar menu for easy navigation between them. Based on Figure 1, The process flow starts from collecting the data, dataset is collected from yahoo finance API which is further described in data and preprocessing steps.

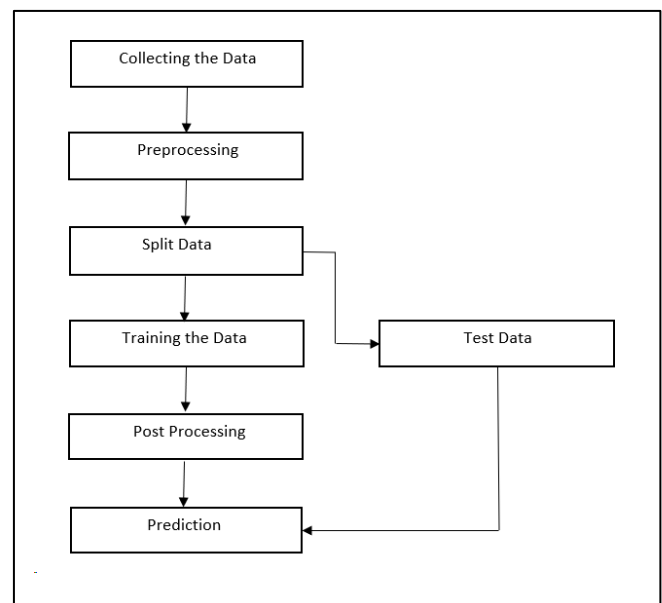


Figure 1. Forecasting / Prediction Model

We provided a high-level overview of the entire system, showing the main processes, data stores, and external entities involved. External users interact with the system to input cryptocurrency data and receive forecasted results. The system utilizes data stores to manage historical data, temporary data, model parameters, and predicted prices as

shown in the below class diagram i.e. Figure 2. :

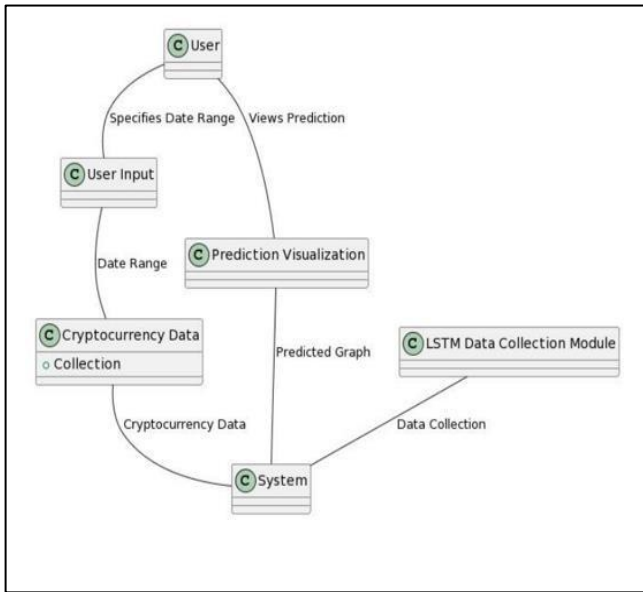


Figure 2. Class Diagram for system

This can delve deeper into the main processes outlined in the figure 2 of class diagram for the system. Each subprocess, such as fetching historical data, preprocessing, training the LSTM model, making predictions, and visualizing data, is further detailed. These subprocesses interact with specific data stores to access and store relevant data, enhancing the understanding of data flows and storage mechanisms within the system as shown in figure 3. This provides a comprehensive overview of the activities and data exchanges involved in the cryptocurrency forecasting process.

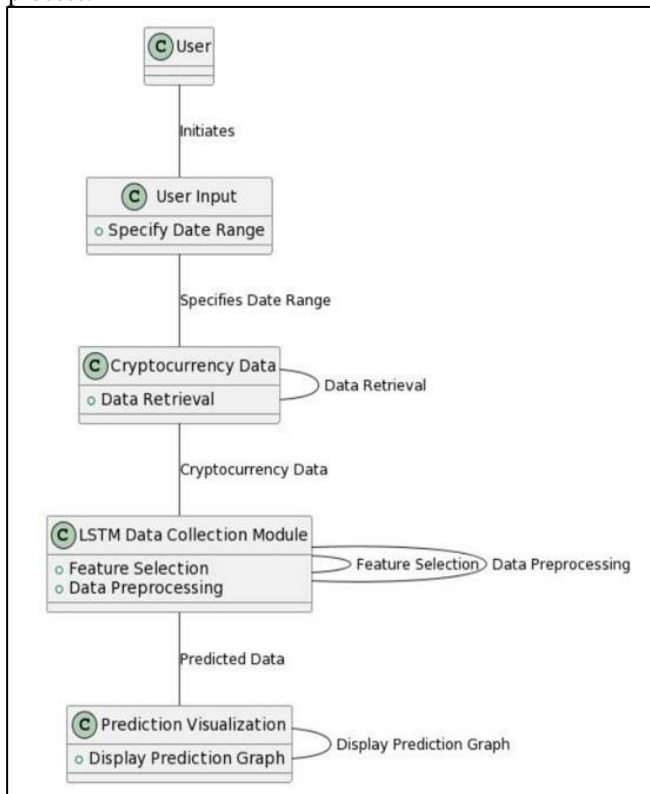


Figure 3. Class Diagram with subprocesses

The Fetch Historical Data subprocess involves a Fetch Data Task that interacts with an external API to retrieve data and store it. Preprocess Data subprocess includes a Preprocess Task responsible for cleaning, transforming, and scaling fetched data. Train LSTM Model subprocess comprises a Train LSTM Model Task that trains the LSTM model using preprocessed data and stores model parameters. Make Predictions subprocess employs a Make Predictions Task to generate future price predictions using the trained model. Finally, Visualize Data subprocess involves a Visualize Data Task to create visual representations of historical and predicted data. Each task interacts with appropriate data stores to access and store relevant information, providing a detailed understanding of the system's inner workings.

A. Data

To conduct our study, we utilized data from two primary sources : Yahoo Finance API and CoinGecko API. These platforms provide extensive and up-to-date data on various financial markets, including historical prices and market indicators for several major cryptocurrencies.

Yahoo Finance API : Through Yahoo Finance, we accessed historical price data and other market metrics for the cryptocurrencies under investigation. The API provides data at different time intervals, including daily, hourly, and minute-level data, allowing us to analyse trends over different horizons.

CoinGecko API : To supplement our dataset, we used the CoinGecko API to obtain additional market data such as trading volume, market capitalization, and price data for the cryptocurrencies. CoinGecko's data offers a comprehensive view of the market dynamics, enabling us to incorporate various factors into our forecasting models.

Date	High	Low	Open	Close	Volume
2023-04-10 00:00:00	29,771.4648	28,189.2715	28,336.0273	29,652.9805	19,282,400,094
2023-04-11 00:00:00	30,509.084	29,609.3008	29,653.6797	30,235.0586	20,121,259,843
2023-04-12 00:00:00	30,462.4805	29,725.5742	30,231.582	30,139.0527	18,651,929,926
2023-04-13 00:00:00	30,539.8457	29,878.623	29,892.7402	30,399.0664	17,487,721,001
2023-04-14 00:00:00	31,005.6074	30,044.498	30,409.5625	30,485.6992	22,659,995,079
2023-04-15 00:00:00	30,601.7402	30,245.8828	30,490.75	30,318.4961	11,940,685,378
2023-04-16 00:00:00	30,555.5371	30,157.832	30,315.9766	30,315.3555	12,854,816,417
2023-04-17 00:00:00	30,319.1973	29,275.3711	30,317.1465	29,445.0449	17,872,186,762
2023-04-18 00:00:00	30,470.3027	29,154.8496	29,449.0918	30,397.5527	19,480,529,496
2023-04-19 00:00:00	30,411.0547	28,669.8984	30,394.1875	28,822.6797	24,571,565,421

Figure 4. Sample Data

B. Preprocessing

Data cleaning and normalization (or scaling) are key parts of the data preprocessing process. The data is retrieved from Yahoo Finance using `pdr.get_data_yahoo`, and we assume that this data is already clean and reliable, for the missing values we can use pandas functions such as `isnull()` and `dropna()`. Normalization is achieved using `MinMaxScaler` from `scikit-learn` to transform the input features (X_{train} and X_{test}) and target variable (y_{train}) in our study it is close price, separately into a range of 0 to 1. This ensures that each feature and the target variable is within a similar range, preventing any one feature from

dominating the learning process. Normalization is crucial for models like Bi-Directional long term short memory LSTM as it helps them train more effectively and converge faster.

After scaling, we convert the normalized data (X_{train} and y_{train}) to PyTorch tensors (`torch.FloatTensor`) for compatibility with the LSTM model. The data is reshaped to ensure it has the correct dimensions for the model's input. This conversion and reshaping process prepares the data for the Bi-LSTM model, enabling it to process the data efficiently and learn more effectively for accurate predictions.

C. Feature Engineering

Enhancing the performance of the model is essential. Iteratively extracting and pruning characteristics was done using a number of different methods. By identifying the most important factors and removing redundant and unnecessary features, feature selection enhances machine learning and boosts the prediction capacity of machine learning algorithms. In our study on improving cryptocurrency price forecasting with bi-directional LSTM, feature selection played a crucial role in enhancing the model's predictive performance. We sourced data from Yahoo Finance and CoinGecko APIs, which included historical prices (open, high, low, close, and adjusted close), trading volume, and market capitalization for several major cryptocurrencies. As feature selection is an important part of the data preprocessing process. Starting by selecting specific columns from the cryptocurrency data as input features (X). These features include 'High', 'Low', 'Open', and 'Volume', which represent various aspects of the cryptocurrency's trading activity. By selecting these particular columns, we focus on key factors that could influence the closing price of the cryptocurrency.

The target variable (\hat{y}) is the 'Close' column, representing the closing price of the cryptocurrency. This is the value the model aims to predict, making it the most important variable for the modelling process. The choice of these features is based on their relevance to the target variable and their potential to provide meaningful insights for the Bi-LSTM model.

By selecting 'High', 'Low', 'Open', and 'Volume' as input features, the code captures key market dynamics, such as the range of price movements and trading volume. This selection ensures that the model has access to diverse and informative data, which can help improve the model's predictive accuracy. Additionally, the choice of these features aligns with common practices in financial modeling, where these variables are known to have an impact on asset prices.

D. Bi-Directional LSTM Model

We predicted the cryptocurrency prices using the Bi-Directional Long Short Term Memory (Bi-LSTM) model. A Bi-Directional LSTM is a type of recurrent neural network (RNN) that can learn long-term dependencies between time steps of sequence data. This deep learning model is very beneficial for time-series data modelling and forecasting. As it combines the power of LSTM with bidirectional processing, allowing the model to capture both past and future context of the input sequence. Because the daily

Bitcoin and Ethereum price and its characteristics are time series data. Unlike traditional RNNs that process input sequences in only one direction (either forward or backward), Bi-LSTM processes the sequence in both directions simultaneously. It consists of two LSTM layers: one processing the sequence in the forward direction and the other in the backward direction. Each layer maintains its own hidden states and memory cells. Our model training involves initializing the model with specific parameters, including input size, hidden size, number of layers, and output size. The model's architecture consists of an LSTM layer followed by a fully connected linear layer that produces the final output as shown in architecture figure 3. The two LSTM networks operate independently, with one receiving the tokens in their original order and the other in reverse. The final output is a combination of the probabilities from both directions. Mathematically, it can be represented in equation(1) :

$$[P_{\text{final}}] = \frac{P_{\text{forward}}}{\text{oplus } P_{\text{backward}}} \quad (1)$$

Where:

(P_{final}) = final probability vector of the network.
 (P_{forward}), (P_{backward}) are probability vector from the forward, backward LSTM network.

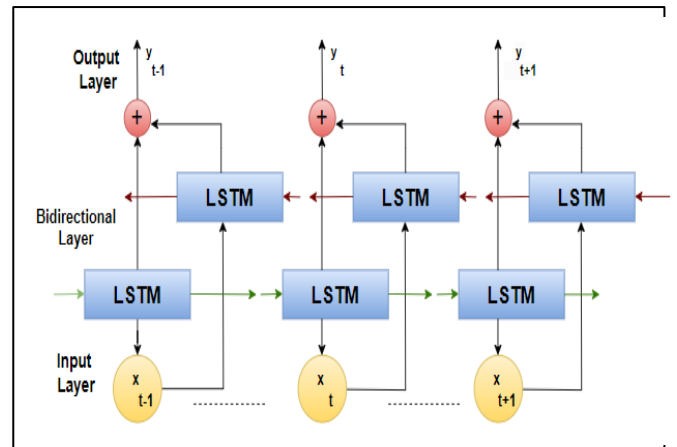


Figure 5. Bi-LSTM Architecture

The loss function used for training is Mean Squared Error (MSELoss), a common choice for regression tasks, which measures the average of the squares of the errors between predicted and actual values.

The optimizer chosen for training the model is Adam, a widely used optimization algorithm that combines the benefits of both AdaGrad and RMSprop, offering adaptive learning rates for each parameter. The learning rate is set to 0.001, a typical starting point for training deep learning models. The model is trained for 50 epochs, during which it goes through the training data multiple times to learn the relationships between input features and the target variable. During each epoch, the model makes predictions based on the input data and calculates the loss using the MSELoss function. The optimizer then adjusts the model's parameters (weights and biases) to minimize the loss. The training process involves iterating through each epoch, updating the model's weights using backpropagation, and monitoring the

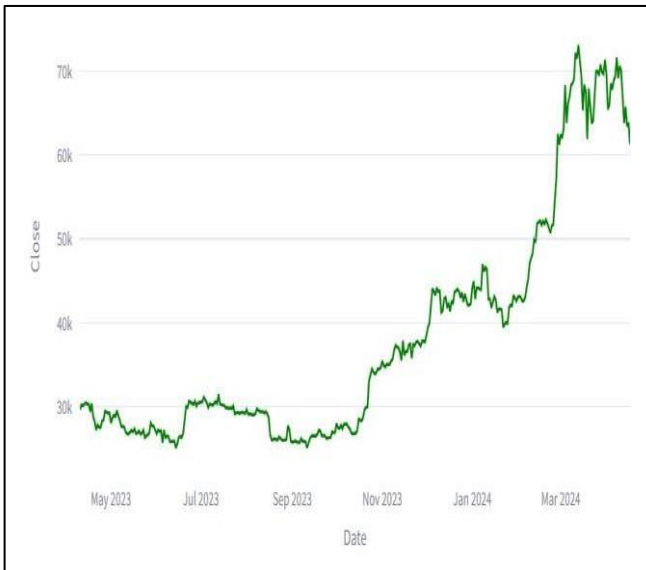
loss to ensure that the model is learning effectively. This iterative process helps the model converge towards an optimal set of parameters, allowing it to make more accurate predictions on the test data and future unseen data.

IV. RESULTS

In this section, the results of the prediction model are summarized.

We successfully created an interactive and user-friendly platform for users to visualize historical and forecasted prices using Streamlit, Pandas, and Altair. This will help investors, traders and finance academics to browse through the interface with ease and get useful insights like forecasting, news related to cryptocurrencies. Created a sidebar menu for navigating the application using option_menu from streamlit_option_menu.

Before we train, test, and predict the outcomes, Figure 6. shows the pre-processing results to load the dataset into machine and algorithm. It also shows the bitcoin closing price data over time. The Yahoo Finance dataset serves for model training, visualization of historical data, moving averages, and change point detection.



Figures 6. Close price History data bitcoin

The CoinGecko dataset is utilized to train a separate LSTM model and forecast cryptocurrency prices for the future, that is the data is used to train another LSTM model and predict cryptocurrency prices for the next year. Hence, traders and investors can have choice for their investments in this leading to profit. Figure 7 shows the predicted value for the next year.

Also generated a line plot that visualizes the 100-day and 200-day moving averages of the closing prices for a specified cryptocurrency over a given time period. This type of visualization helps in analyzing the trends and smoothing out short-term fluctuations in the price data, providing a clearer picture of the underlying long-term trends. Shown as in figure number 8.

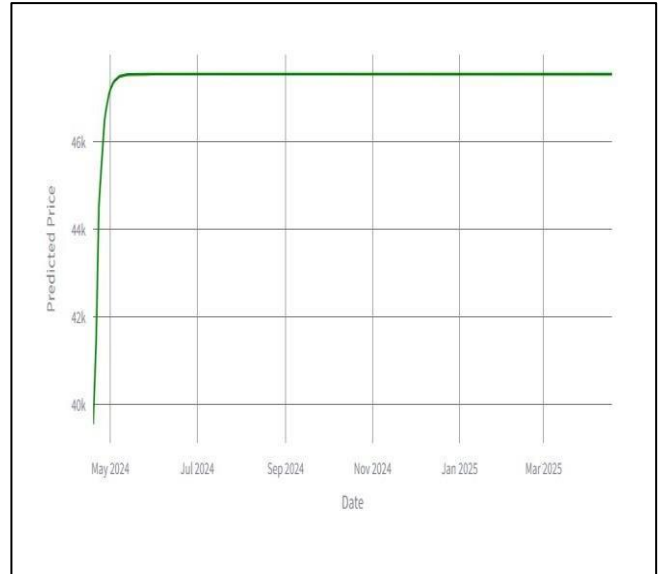
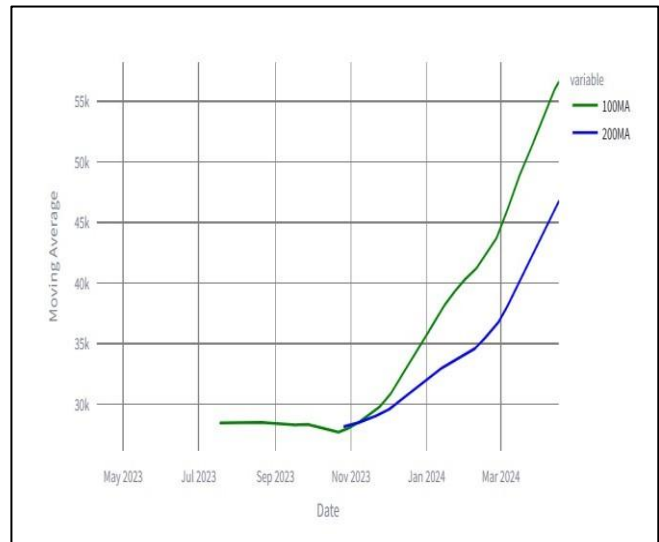


Figure 7. Result of Predicted price for next year



Figures 8. Moving averages of closing price

V. REFERENCES

- [1] Rafi, M., Hameed, S., Sohail, I., Aliasghar, M., Aziz, A., & Mirza, Q. A. K. (2023). Enhancing cryptocurrency price forecasting accuracy: A feature selection and weighting approach with bi-directional LSTM and trend-preserving model bias correction. *IEEE Access*.
- [2] Mudassir, M., Bennbaia, S., Unal, D., & Hammoudeh, M. . Time-series forecasting of Bitcoin prices using high-dimensional features: a machine learning approach , pp.1-15, 2020.
- [3] Rajua Ali Mohammed Tarif, S. M. , Using machine learning techniques and public sentiment analysis, real-time Bitcoin price prediction .pp.13-14,2006.
- [4] Somarouthu Venkat Sai, Tubati Sai Manikanta, and Karunya Rathan , Crypto-Currency price prediction using Decision Tree and Regression techniques, pp.190-194, 2019
- [5] Selvin, Sreelekshmy, R. Vinayakumar, E. A. Gopalakrishnan, Vijay Krishna Menon, and K. P. Soman. "Stock price prediction using LSTM, RNN and CNN-sliding window model." In *2017 international conference on advances in computing, communications and informatics (icacci)*, pp. 1643-1647.
- [6] Seabe, Phumudzo Lloyd, Claude Rodrigue Bambe Moutsinga, and Edson Pindza. "Forecasting cryptocurrency prices using LSTM, GRU, and bi-directional LSTM: A deep learning approach." *Fractal and Fractional* 7, no. 2 (2023),pp.203.

- [7] Hamayel, Mohammad J., and Amani Yousef Owda. "A novel cryptocurrency price prediction model using GRU, LSTM and bi-LSTM machine learning algorithms." *Ai* 2.4 (2021), pp.477-496.
- [8] Livieris, I. E., Kiriakidou, N., Stavroyiannis, S., & Pintelas, P. (2021). An advanced CNN-LSTM model for cryptocurrency forecasting. *Electronics*, 10(3), pp.287.
- [9] Akita, R., Yoshihara, A., Matsubara, T., & Uehara, K. (2016, June). Deep learning for stock prediction using numerical and textual information. In *2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS)* (pp. 1-6). IEEE.
- [10] Wu, S., Liu, Y., Zou, Z., & Weng, T. H. (2022). S_I_LSTM: stock price prediction based on multiple data sources and sentiment analysis. *Connection Science*, 34(1), pp. 44-62.
- [11] Bao, W., Yue, J., & Rao, Y. (2017). A deep learning framework for financial time series using stacked autoencoders and long-short term memory. *PLoS one*, 12(7), e0180944.
- [12] Li, Hao, Yanyan Shen, and Yanmin Zhu. "Stock price prediction using attention-based multi-input LSTM." In *Asian conference on machine learning*, pp. 454-469. PMLR, 2018.
- [13] Guizhu Shen, Qingping Tan, Haoyu Zhang, Ping Zeng, Jianjun Xu, Deep Learning with Gated Recurrent Unit Networks for Financial Sequence Predictions, *Procedia Computer Science*, Volume 131, 2018, pp.895-903, ISSN 1877-0509
- [14] Li, D., Wang, Y., Madden, A., Ding, Y., Tang, J., Sun, G. G., ... & Zhou, E. (2019). Analyzing stock market trends using social media user moods and social influence. *Journal of the Association for Information Science and Technology*, 70(9), pp.1000-1013.
- [15] Sun, L., Xu, W., & Liu, J. (2021). Two-channel attention mechanism fusion model of stock price prediction based on CNN-LSTM. *Transactions on Asian and Low-Resource Language Information Processing*, 20(5), pp.1-12.
- [16] Awoke, T., Rout, M., Mohanty, L., & Satapathy, S. C. Bitcoin price prediction and analysis using deep learning models." *Communication Software and Networks: Proceedings of INDIA 2019*. Singapore: Springer Singapore, 2020, pp.631-640.
- [17] Livieris, I. E., Stavroyiannis, S., Pintelas, E., & Pintelas, P. (2020). A novel validation framework to enhance deep learning models in time-series forecasting. *Neural Computing and Applications*, 32(23), pp.17149-17167.
- [18] Rose, C. (2015). The evolution of digital currencies: Bitcoin, a cryptocurrency causing a monetary revolution. *International Business & Economics Research Journal (IBER)*, 14(4), pp.617-622.
- [19] Murray, K., Rossi, A., Carraro, D., & Visentin, A. (2023). On forecasting cryptocurrency prices: A comparison of machine learning, deep learning, and ensembles. *Forecasting*, 5(1), pp.196-209.
- [20] Derbentsev, V., Matviychuk, A., & Soloviev, V. N. (2020). Forecasting of cryptocurrency prices using machine learning. *Advanced studies of financial technologies and cryptocurrency markets*, pp.211-231.
- [21] Wu, C. H., Lu, C. C., Ma, Y. F., & Lu, R. S. (2018, November). A new forecasting framework for bitcoin price with LSTM. In *2018 IEEE international conference on data mining workshops (ICDMW)* (pp. 168-175). IEEE.
- [22] Nithyakani, P., Tom, R. J., Gupta, P., Shanthini, A., John, V. M., & Sharma, V. (2021, January). Prediction of bitcoin price using Bi-LSTM network. In *2021 International Conference on Computer Communication and Informatics (ICCCI)* (pp. 1-5). IEEE.
- [23] Derbentsev, V., Datsenko, N., Babenko, V., Pushko, O., & Pursky, O. (2020, October). Forecasting cryptocurrency prices using ensembles-based machine learning approach. In *2020 IEEE International Conference on Problems of Infocommunications. Science and Technology (PIC S&T)* (pp. 707-712). IEEE.
- [24] Ferdiansyah F., Othman, S. H., Radzi, R. Z. R. M., Stiawan, D., Sazaki, Y., & Ependi, U. (2019, October). A lstm-method for bitcoin price prediction: A case study yahoo finance stock market. In *2019 international conference on electrical engineering and computer science (ICECOS)* (pp. 206-210). IEEE.
- [25] Jaquart, P., Köpke, S., & Weinhardt, C. (2022). Machine learning for cryptocurrency market prediction and trading. *The Journal of Finance and Data Science*, 8, pp.331-352