Prediction and Analysis of Bitcoin Price using Machine learning and Deep learning models

Vinay Karnati[1], Lakshmi Dathatreya Kanna[2], Trilok Nath Pandey[3, *] and Chinmaya Kumar Nayak[4]

[1,2,3] School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, India
[4] Faculty of Emerging Technologies, Sri Sri University, Odisha, India

Abstract

High Accessibility and Easy Investment makes Cryptocurrency an important income source for many people. Cryptocurrency is a kind of Digital/Virtual currency which is created using blockchain Technology and is protected by Cryptography. Cryptocurrencies enables users to Accept, Transfer and request the capital between the Users without the requirement of intermediaries such as banks. Now a day many Cryptocurrencies are available across the world such as Bitcoin, Litecoin, Monero, Dogecoin etc. This study is more determined over a very famous and demanding Cryptocurrency known as Bitcoin over the past years. Here, firstly we make an effort to predict the price of bitcoin by examining numerous numbers of parameters that affect the cost of bitcoin. Different kinds of Machine learning models will be used to estimate the price of Bitcoin. This study provides the accuracy and precision of each model that are used in this study and determine the suitable method to estimate the price more accurately.

1. Introduction

Cryptocurrency is Digital/virtual currency which is developed using blockchain technology and it is encrypted by cryptography [4]. These are designed in such a way that the two users across the world can accept, transfer, and receive the money that is connected to internet without any third parties such as banks or governments to support it [1].

A blockchain is a distributed ledger that distributes and replicates transactions across a network of computers connected to it. Blockchain technology is a system for keeping public transactional records (also known as "blocks") throughout several databases in a network of peer-to-peer nodes [12]. It is very successful in securing the process and data to get manipulated or hacked. Each and every dealing is stored, and multiple copies of the same Ledger is held on [8].

Since cryptocurrencies, particularly Bitcoin, is dealing with large and enormous amount of data in the past 3 years, the correct prognostication of future predictions is becoming a lot vital and attention-grabbing space for each researcher and investor [3]. Usually, Bitcoin is a financial bonus and is recorded in numerous cryptocurrency exchanges securities market. So, the price fluctuations, patterns and various experiments are conducted by the researchers to know the optimal findings that help in predicting the future price of various cryptocurrencies [8,22]. In this Paper we will study numerous bitcoin postulation models like Autoregressive Integrated Moving Average, Facebook's Prophet, Long Short-Term Memory and Support Vector Machine [3,7,8,9].

The following is the flow of this document. Section II briefly explains the literature survey. The dataset, data preprocessing, and data analytics steps represented in Section III. The following Methodology is given in section IV. Experiments and results are explained in section V. Whereas, the conclusions are given in section VI and finally followed by References.

2. Literature Survey

Temesgen Awoke, Minakhi Rout, Lipika Mohanty, and Suresh Chandra Satapathy has Proposed a research paper on Long short-term memory and gated recurrent unit, are two deep learning models that are used for forecasting and analyzing the price of bitcoin to achieve high accuracy while managing the price volatility of bitcoin [2]. Their study includes comparing the above-mentioned time series deep learning models' result and proved the efficacy in forecasting the price of bitcoin. According to the study, the GRU model is more accurate and quicker to compile than LSTM for time series forecasting.

IsılYenidoğan, AykutCayır, OzanKozan, TugceDag, Cigdem Arslan has proposed a research paper on bitcoin forecasting using statistical analysis models such as ARIMA and PROPHET [3]. Using the same dataset, the performance measures of the PROPHET and ARIMA approaches are compared to determine the best accurate forecast model. The dataset's time series features are taken into consideration for both the models, ARIMA and PROPHET utilizing the threefold splitting approach. After two models were successfully built, they are compared in terms of performance metrics, and have concluded that the PROPHET model is better as it has an R2 value of 0.94 while ARIMA has a value of 0.68.

Suhwan Ji, Jongmin Kim and HyeonseungIm has proposed a research paper based on a comparative study using deep learning models on bitcoin price prediction. A few models of deep neural networks which includes Long short-term memory known as LSTM, Deep neural network known as DNN, Convolutional Neural network known as CNN and their combinations are experimented to predict the bitcoin price. Results have been studied and analyzed to find the accurate model. Despite the fact that LSTM-based prediction models beat the other forecasting model (regression) for Bitcoin price prediction, DNN models are the best when it comes to classifying whether the bitcoin price will increase or decrease. They found that categorization models were more useful for algorithmic trading than regression models [5].

Grace. LK. Joshila, Asha. P, D. Usha Nandini, G. Kalaiarasi has proposed a research paper, price prediction of bitcoin using the machine learning algorithm, SVM (support vector machine). They Followed steps such as taking raw data, feature Selection, data Preprocessing, data splitting, applying model (SVM), and predicting the results. The keras function of SVM model makes it simple to compute and manipulate numerical values. It also takes less time to train data, and the time complexity of SVM model is lower than that of Random Forest and Convolutional Neural Networks. So, after computing the accuracy, they have concluded that support vector machine (SVM) is much more accurate than other algorithms such as Random Forest and Convolutional Neural Networks (CNN) [7].

3. Data Preprocessing

Data preparation refers to the act of converting unstructured data into a structured format. It is an initial and very crucial phase in data mining. It is important to evaluate the quality of data before training the models. The initial phases in data preparation include data collection, data cleaning, data reduction, and data transformation.

1) Data Collection

The data is collected from the Yahoo finance website dated from year 2013 to 2021. This data contains attributes such as date, highest price, lowest Price, opening price, closing price and market cap on those particular days [25].

2) Data Cleaning

The process of restoring missing numbers and eliminating inaccurate, insufficient, and wrong data from datasets is known as data cleaning. In terms of data cleaning procedures, we may either eliminate missing numbers or substitute them using approaches like mean, median, and mode. Imputation is the term for this action [7].

4. Methodology

To compare the performance of the Facebook Prophet, ARIMA, SVM, and LSTM algorithms for time series forecasting and training. Our methods included preprocessing the data, running the algorithms, and evaluating their efficacy with appropriate metrics.

4.1 Facebook's Prophet:

The trends nowadays are nonlinear and change on a day-to-day basis. So, to forecast this change, which is happening in the time series, the Facebook prophet is the best choice. This prophet deals with the outliers well, and it is robust to missing data [3]. This prophet can be used for goal setting of our model and producing reliable forecast and to perform this we just put our model in stan, and we get forecasts within seconds [8, 19]. The prophet Algorithm has many useful applications which are mainly used across Facebook for generating many authentic speculations and for planning of the goals etc.

$$g(t) = \frac{C}{1 + e^{(-k(t-m))}} \qquad (1)$$

Here C is the Carrying Capacity, k the growth rate, and m an offset parameter.

1) **Training Approach:** Prophet models were trained using historical Bitcoin price data as input and any available information on holidays or events that could impact the price. The aim was to capture the trend, seasonality, and holiday components in the data.

2) **Optimization Algorithm:** Prophet models employ an optimization algorithm based on a gradient descent procedure to fit the model to the data and optimize the model parameters.

3) **Hyperparameter Tuning:** Hyperparameters such as the seasonality mode, the changepoint prior scale, and the prior scale for holidays can be tuned to optimize the model's performance.

4.2 Autoregressive Integrated Moving Average (ARIMA):

Arima explains a given time series based on its own past data which is lagged of its forecast errors. Arima can solve that issue and exhibits patterns. Automated integrated flow rate, or Arima is a mathematical model that throws light on time series data which is used to predict the trends in the data [9]. Using past values as a predictive model, this mathematical model is autoregressive.

There are three terminologies that describe the Autoregressive Integrated Moving Average model [3,18]. They are P, Q, and D

1) In this case, P is used to indicate the order in which the AR terms appear.

2) In this case, Q is used to indicate the order in which the MA terms appear.

3) Here D is Minimum differencing steps which is used to make the time series stationary.

If the timeline has recurring patterns, you should add terminals to the season and then 'ARIMA Season' is denoted by SARIMA [9]. Before understanding 'AR name order' we first look at D. The Auto Regressive term's order is P which refers to Y number of lags that are used as predictors. And 'q' is a term system for the term 'Moving Average' (MA), which refers to the number of outdated predictor errors that should fit into the ARIMA model [3,14].

$$Y_t = \propto + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} + \epsilon_1 \qquad (2)$$

1) **Training Approach:** ARIMA models were trained using historical time series data of Bitcoin prices. The aim was to capture the patterns and dependencies in the data for accurate forecasting [26].

2) **Optimization Algorithm:** ARIMA models employ optimization algorithms to estimate the model parameters, including autoregressive (AR) and moving average (MA) coefficients. The estimation is typically done using methods like maximum likelihood estimation (MLE) [26].

3) **Diagnostic Testing:** Before training the ARIMA model, the data's stationarity was checked using the Dickey-Fuller test. If the data was found to be non-stationary, differencing was applied to make it stationary [26].

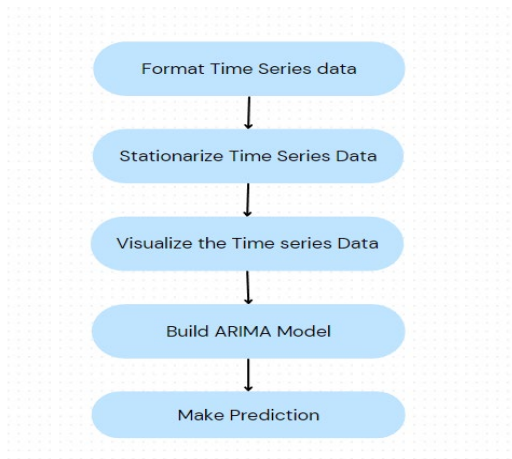Figure 1 below shows the workflow of the ARIMA model.

Figure 1 ARIMA Workflow Model

4.3 Support vector machine (SVM):

It is used for speculative inspection which allocates advanced data points to one of named classes in data we are performing [6].

The SVM is known as Support Vector Machine which is used to solve classification problems. It has many applications in real time like image processing, medical diagnosis, patterns, and text analytics [6,16]. It is a combination of supervised learning methods used to classify the data, regression analysis and detecting anomaly in the data [17,21].

Training and Regularization Technique:

1) **Training Approach:** SVM models were trained using historical data on Bitcoin prices as input features and the corresponding target values. The goal was to find the optimal hyperplane that separates or predicts the Bitcoin prices [15].

2) **Regularization Technique:** SVM models incorporate regularization through the use of a regularization parameter (C) to control the trade-off between maximizing the margin and minimizing the classification error.

4.4 Long Short-Term Memory (LSTM):

Neural networks are useful for bitcoin price prediction because of their capacity to capture complicated patterns and nonlinear correlations in data, allowing them to learn from previous trends and create accurate projections based on many factors impacting bitcoin pricing [23].

Long Short-Term Memory is a type of Recurrent neural network architecture which is effective in handling the issue of Vanishing gradients in traditional RNN's [10,24]. It is considered as an upgraded version of RNN model [13].

LSTM layers are categorized into 3 main layers. They are input layer, LSTM layer and output layer [24].

Input layer basically takes the input data which can be in any format such as text, audio or any other time series data and the input data is preprocessed and encoded. [11] LSTM layer is responsible for long-term dependencies in the input data and LSTM layer consists of cells where each cell consists of 3 gates: input gate, forget gate and output gate. They are:

1) **Input gate:** This gate has the ability to control information. It tells whether to let new information enter the cell or not [20].

2) **Forget gate:** According to the requirement this gate forgets the information in the previous state [20].

3) **Output gate:** It tells which part of the cell state should be output as the LSTM's hidden state.

The output layer takes the final output and provides us with predicted results. The output layer consists of SoftMax activation function which is used to solve classification problems and for regression problems we use linear activation function [20].

1) **Training Approach:** LSTM models were trained using historical Bitcoin price data as input sequences and the corresponding target values. The objective was to capture long-term dependencies and temporal patterns in the data.

2) **Optimization Algorithm:** LSTM models employ optimization algorithms like Stochastic Gradient Descent (SGD) or Adam to update the network weights and minimize the loss function.

3) **Hyperparameter Tuning:** Hyperparameters such as the number of LSTM layers, the number of hidden units, the learning rate, and the batch size were tuned to optimize the model's performance.

4.5 Performance Metrics

In our study, we employed several evaluation metrics to assess the performance of the proposed models for bitcoin price prediction. These metrics allowed us to quantitatively measure the accuracy and predictive capability of our models and make meaningful comparisons with existing approaches [3].

1) **Root Mean Squared Error (RMSE):** RMSE calculates the square root of the average squared differences between the predicted and actual bitcoin prices. It provides a more comprehensive measure of the model's performance by penalizing larger errors more heavily than MAE.

2) **Accuracy**: In addition to regression-based metrics, we also employed accuracy as an evaluation measure for classification tasks, such as predicting price direction (upward or downward). Accuracy measures the percentage of correctly predicted price directions and provides insights into the model's ability to capture market trends.

5. Experimentation and Results

It is vital to note that accuracy is commonly employed as a performance indicator for classification tasks, assessing the proportion of properly categorized cases. The accuracy statistic may not be as useful in time-series forecasting. Instead, the emphasis is frequently placed on assessing the accuracy of anticipated values in comparison to actual values. In this situation, the RMSE measure assesses the average difference between expected and actual values, which aids in quantifying and predicting accuracy.

5.1 Autoregressive Integrated Moving Average (ARIMA):

After scrutinizing the results of actual and predicted prices from the below Fig. 2, to assess the accuracy of the ARIMA model, root mean squared error was calculated. Majorly, the resultant is less than 6%. After that, the RMSE was undaunted, revealing that level of rightness is 87% for ARIMA.

Future time-stamped data that satisfy the Dicky Fuller test, in which the p-value should be below .5, may be speculated using the same ARIMA model of forecast.
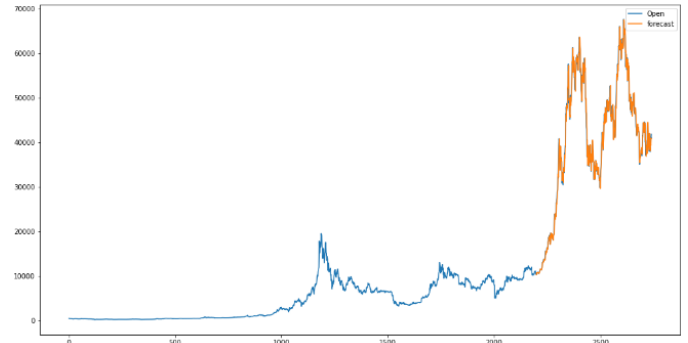

Figure 2 Forecast Data from ARIMA Model

Figure 2 is Plotting the forecasted data from the auto Arima model where x represents the year and y axis represents the price. Here the blue line gives the actual data and orange line gives the forecasted price.

5.2 Support Vector Machine (SVM):

As you can see in Fig. 3, the bitcoin price values are both actual and predicted. For this study, we examined data between 2014 and 2022. To plan this figure, we must take into account the closing price as well as predicted values. According to the graph above, the X-axis represents the number of days, while the Y-axis represents the price of bitcoin. As a result of this graph, we can anticipate the price of bitcoins, which may be extremely beneficial. A few days after the price of bitcoin began to rise exponentially after halfway through, it then started to decline after a few days. By doing so, we are able to observe the price fluctuations of bitcoin over time.
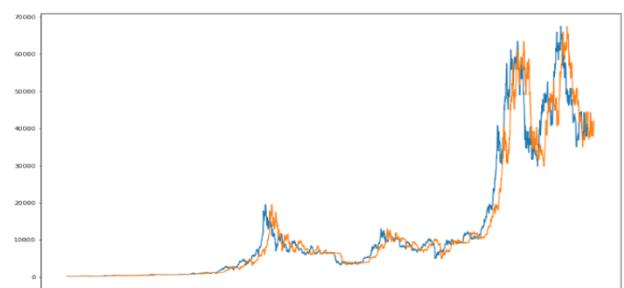

Figure 3 Sample Training and Testing Data

Here x represents the year and y axis represents the price of bitcoin.

5.3 Facebook's Prophet:

Prophet has gained popularity due to its ease of use and its ability to handle noisy and irregularly spaced data. By incorporating historical patterns and leveraging trend and

seasonality components, it can provide accurate predictions for various time series applications.

In a comprehensive analysis of historical data from 2015 to 2022 from Figure 4, utilizing Facebook's Prophet algorithm yielded an impressive accuracy of 88%. This indicates the algorithm's effectiveness in capturing underlying patterns and making reliable predictions over an extended period, providing valuable insights for decision-making, and planning in diverse industries. Here, blue line is predicted value and black dotted lines are actual value also x axis represents the year and y axis represents the price.
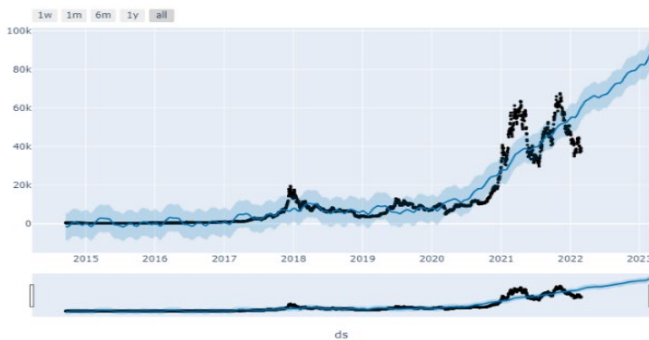

Figure 4 Plotting the model estimates of Prophet Model

5.4 Long Short-Term Memory (LSTM):

Applying the LSTM algorithm to a comprehensive analysis of historical data from 2015 to 2022 yielded remarkable accuracy, achieving a high precision rate of 91%. This highlights the algorithm's capability to understand complex temporal dependencies, capture long-term patterns, and make accurate predictions in a variety of time series applications. Fig.5 shows a total of 2900 data points, or almost 8 years, indicate the price of bitcoin in US dollars.
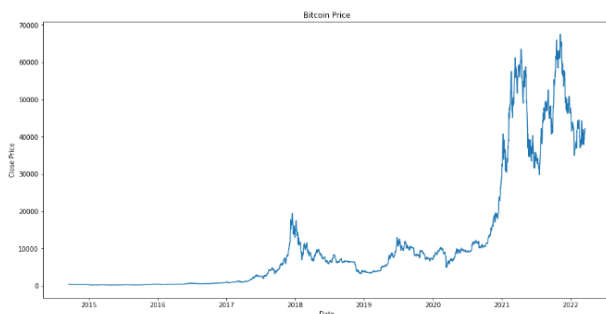

Figure 5 Bitcoin price over the time.

Here the blue line represents the price of bitcoin over the years.

Leveraging the memory cells and recurrent connections, LSTM demonstrates superior performance in forecasting and can be invaluable in predicting future trends, making informed decisions, and optimizing strategies across diverse industries and domains.
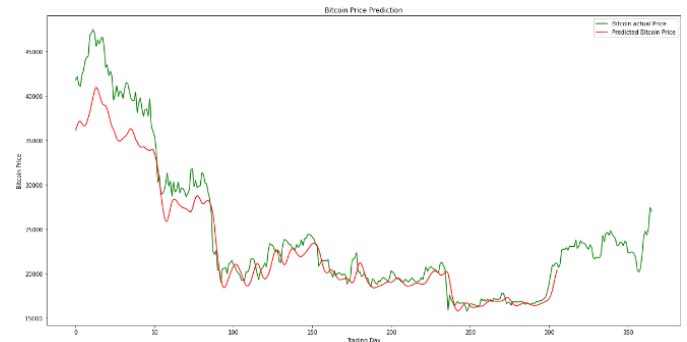

Figure 6 Actual Price vs Predicted price.

The red line represents the predicted price and the green line represents actual data.

The above Fig. 6, depicts the actual price and predicted price of the bitcoin using the LSTM model where x-axis gives the price at the time from y-axis. Table 1. shows the RMSE and accuracy values between actual and predicted values. Comparison of root mean square error (RMSE) and accuracy achieved by different models.

Table.1 Comparison of Model Performance

Model	RMSE	Accuracy
ARIMA	2806.73	87%
SVM	2945.84	86%
Prophet	2792.34	88%
LSTM	2523.26	91%

6. Conclusion

We have implemented four different models (Prophet, SVM, ARIMA, LSTM). This study is intended to differentiate Bitcoin's future price using these algorithms. Till now we have found that LSTM performs the best in

prediction of future values. SVM model shows around 86% accuracy and talking about Prophet, it performs very good over small changes in values but performs slightly low when the values difference is large. There are lot more models that are built for predicting values using previous data, we have used the most famous among them. After comparing the results of the above 3 prediction algorithms the Facebook prophet algorithm is giving results with almost 88% accuracy Overall, predicting price-related changing is difficult given the relative strength of the market. However, using deep neural networks has given us a better understanding of Bitcoin, as well as LSTM structures. Present work involves using the hyper parameter, to obtain a more accurate network configuration. Also, other factors can be considered, although from our experiment with Bitcoin, many factors did not lead to better results. Smaller economic factors may be included in the model for a better understanding.

References

[1] S. Velankar, S. Valecha and S. Maji, "Bitcoin price prediction using machine learning," 2018 20th International Conference on Advanced Communication Technology (ICACT), 2018, pp. 144-147, doi: 10.23919/ICACT.2018.8323676.

[2] Muniye, Temesgen& Rout, Minakhi& Mohanty, Lipika&Satapathy, Suresh. (2020). Bitcoin Price Prediction and Analysis Using Deep Learning Models. 10.1007/978-981-15-5397-4_63.

[3] Cayir, Aykut&Kozan, Ozan &Dağ, Tuğçe&Yenidoğan, Işıl& Arslan, Çiğdem. (2018). Bitcoin Forecasting Using ARIMA and PROPHET. 10.1109/UBMK.2018.8566476.

[4] Ankalkoti, Prashant &Santhosh,. (2017). A Relative Study on Bitcoin Mining. "Imperial Journal of Interdisciplinary Research (IJIR).

[5] Ji, Suhwan& Kim, Jongmin&Im, Hyeonseung. (2019). A Comparative Study of Bitcoin Price Prediction Using Deep Learning. Mathematics. 7. 898. 10.3390/math7100898.

[6] D. Shah and K. Zhang, "Bayesian regression and Bitcoin," 2014 52nd Annual Allerton Conference on Communication, Control, and Computing (Allerton), 2014, pp. 409-414, doi: 10.1109/ALLERTON.2014.7028484.

[7] G. L. Joshila, A. P, D. U. Nandini and G. Kalaiarasi, "Price Prediction of Bitcoin," 2021 5th International Conference on Trends in Electronics and Informatics (ICOEI), 2021, pp. 113-116, doi: 10.1109/ICOEI51242.2021.9452976.

[8] R. N, S. R. R, V. S. R and K. P. D, "Crypto-Currency Price Prediction using Machine Learning," 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI), 2022, pp. 1455-1458, doi: 10.1109/ICOEI53556.2022.9776665.

[9] G Peter Zhang. Time series forecasting using a hybrid arima and neuralnetwork model. Neurocomputing, 50:159–175, 2003

[10] M. Fernandes, S. Khanna, L. Monteiro, A. Thomas and G. Tripathi, "Bitcoin Price Prediction," 2021 International Conference on Advances in Computing, Communication, and Control (ICAC3), Mumbai, India, 2021, pp. 1-4, doi: 10.1109/ICAC353642.2021.9697202.

[11] Samiksha Marne, Shweta Churi, Delisa Correia, Joanne Gomes, 2021, Predicting Price of Cryptocurrency – A Deep Learning Approach, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) NTASU – 2020 (Volume 09 – Issue 03),

[12] Awoke, T., Rout, M., Mohanty, L., Satapathy, S.C. (2021). Bitcoin Price Prediction and Analysis Using Deep Learning Models. In: Satapathy, S.C., Bhateja, V., Ramakrishna Murty, M., Gia Nhu, N., Jayasri Kotti (eds) Communication Software and Networks. Lecture Notes in Networks and Systems, vol 134. Springer, Singapore. https://doi.org/10.1007/978-981-15-5397-4_63

[13] Yiqing Hua "Bitcoin price prediction using ARIMA and LSTM". E3S Web Conf. 218 01050 (2020)DOI: 10.1051/e3sconf/202021801050

[14] B. Malhotra, C. Chandwani, P. Agarwala and S. Mann, "Bitcoin Price Prediction Using Machine Learning and Deep Learning Algorithms," 2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 2022, pp. 1-6, doi: 10.1109/ICRITO56286.2022.9964677.

[15] Huang, W., Nakamori, Y., & Wang, S. Y. (2005). Forecasting stock market movement direction with support vector machine. Computers & operations research, 32(10), 2513-2522.

[16] Tay, F. E., & Cao, L. (2001). Application of support vector machines in financial time series forecasting. omega, 29(4), 309-317.

[17] Cao, L., & Tay, F. E. (2001). Financial forecasting using support vector machines. Neural Computing & Applications, 10, 184-192.

[18] Anusha, Garlapati & Krishna, Doredla & Garlapati, Kavya & Yaswanth, Nandigama & Rahul, Udayagiri & Mannazhath, Gayathri. (2021). Stock Price Prediction Using Facebook Prophet and Arima Models. 1-7. 10.1109/I2CT51068.2021.9418057.

[19] R. G. Tiwari, A. K. Agarwal, R. K. Kaushal and N. Kumar, "Prophetic Analysis of Bitcoin price using Machine Learning Approaches," 2021 6th International Conference on Signal Processing, Computing and Control (ISPCC), Solan, India, 2021, pp. 428-432, doi: 10.1109/ISPCC53510.2021.9609419.

[20] T. Li, "Prediction of Bitcoin Price Based on LSTM," 2022 International Conference on Machine Learning and Intelligent Systems Engineering (MLISE), Guangzhou, China, 2022, pp. 19-23, doi: 10.1109/MLISE57402.2022.00012.

[21] T. N. Pandey, A. K. Jagadev, S. K. Mohapatra and S. Dehuri, "Credit risk analysis using machine learning classifiers," *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing*

(ICECDS), Chennai, India, 2017, pp. 1850-1854, doi: 10.1109/ICECDS.2017.8389769.

[22] Trilok Nath Pandey, Alok Kumar Jagadev, Satchidananda Dehuri and Sung-Bae Cho,"Prediction of exchange rate using improved particle swarm optimised radial basis function networks", International Journal of Advanced Intelligence Paradigms,Vol. 23, No. 3-4,pp 332-356https://doi.org/10.1504/IJAIP.2022.126695.

[23] Pandey, T.N., Priya, T., Jena, K.: Prediction of Exchange rate in a cloud computing environment using machine learning tools. Intell. Cloud Comput. 137–146 (2021).

[24] Pandey, T.N., Mahakud, R.R., Patra, B., Giri, P.K., Dehuri, S. (2022). Performance of Machine Learning Techniques Before and After COVID-19 on Indian Foreign Exchange Rate. In: Dehuri, S., Prasad Mishra, B.S., Mallick, P.K., Cho, SB. (eds) Biologically Inspired Techniques in Many Criteria Decision Making. Smart Innovation, Systems and Technologies, vol 271. Springer, Singapore. https://doi.org/10.1007/978-981-16-8739-6_41.

[25] Ferdiansyah, Ferdiansyah & Othman, Siti & Raja Mohd Radzi, Raja Zahilah & Stiawan, Deris & Sazaki, Yoppy & Ependi, Usman. (2019). A LSTM-Method for Bitcoin Price Prediction: A Case Study Yahoo Finance Stock Market. 206-210. 10.1109/ICECOS47637.2019.8984499.

[26] K. Ramani, M. Jahnavi, P. J. Reddy, P. VenkataChakravarthi, P. Meghanath and S. K. Imran, "Prediction of Bitcoin Price through LSTM, ARIMA, XGBoost, Prophet and Sentiment Analysis on Dynamic Streaming Data," 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2023, pp. 1514-1518, doi: 10.1109/ICACCS57279.2023.10113014.