IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 1 September 2023, accepted 16 October 2023, date of publication 25 October 2023, date of current version 6 November 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3327440

RESEARCH ARTICLE

Forecasting and Trading of the Stable Cryptocurrencies With Machine Learning and Deep Learning Algorithms for Market Conditions

HASIB SHAMSHAD[1], FASEE ULLAH[2], ASAD ULLAH[1], VICTOR R. KEBANDE[3], SIBGHAT ULLAH[1], AND ARAFAT AL-DHAQM[2]

[1]Department of Computer Science and IT, Sarhad University of Science and Information Technology, Peshawar 25000, Pakistan
[2]Department of Computer and Information Sciences, Universiti Teknologi PETRONAS, Bandar Seri Iskandar, Perak 32610, Malaysia
[3]Department of Computer Science (DIDA), Blekinge Institute of Technology, 371 79 Karlskrona, Sweden

Corresponding authors: Fasee Ullah (fasee.csit@suit.edu.pk) and Victor R. Kebande (victor.kebande@bth.se)

This work was supported in part by the Blekinge Institute of Technology (BTH), Sweden; and in part by the Swedish Knowledge Foundation through the Project Symphony—Supply-and-Demand-Based Service Exposure Using Robust Distributed Concepts under Grant Dnr 20190111.

ABSTRACT The digital market trend is rapidly expanding due to key characteristics like decentralization, accessibility, and market diversity enabled by blockchain technology. This study proposes a Predictive Analytics System to provide simplified reporting for the three most popular cryptocurrencies with varying digits, namely ADA Cardano, Ethereum, and Binance coin, for ten days to contribute to this emerging technology. Thus, this proposed system employs a data science-based framework and six highly advanced data-driven Machine learning and Deep learning algorithms: Support Vector Regressor, Auto-Regressive Integrated Moving Average (ARIMA), Facebook Prophet, Unidirectional LSTM, Bidirectional LSTM, Stacked LSTM. Moreover, the research experiments are repeated several times to achieve the best results by employing hyperparameter tuning of each algorithm. This involves selecting an appropriate kernel and suitable data normalization technique for SVR, determining ARIMA's (p, d, q) values, and optimizing the loss function values, number of neurons, hidden layers, and epochs in LSTM models. For the model validation, we utilize widely used evaluation techniques: Mean Absolute Error, Root Mean Squared Error, Mean Absolute Percentage Error, and R-squared. Results demonstrate that ARIMA outperforms the other models in all cases, accurately projecting the price variability within the actual price range. Conversely, Facebook Prophet exhibits good performance to some extent. The paper suggests that the ARIMA technique offers practical implications for market analysts, enabling them to make well-informed decisions based on accurate price projections.

INDEX TERMS Regression analysis, predictive analytics, time series forecasting, ARIMA, Ethereum, ADA Cardano, Binance, cryptocurrency forecasting, machine learning, deep learning, support vector regressor, FB prophet, bidirectional LSTM, unidirectional LSTM.

I. INTRODUCTION

The prominent cryptocurrencies have entirely transformed the financial landscape with their secure and decentralized digital alternatives to physical cash [1]. Stablecoins received much attention among the wide range of cryptocurrencies because of their unique ability to maintain a steady value by being connected to outside assets like cash or commodities [2]. These cryptocurrencies have the potential for various use cases due to their stability, such as international trade,

The associate editor coordinating the review of this manuscript and approving it for publication was Junhua Li.

remittances, and market volatility hedging. Although stable, the cryptocurrency market is nevertheless quite volatile and prone to sharp volatility. Therefore, for investors, traders, and market players to make wise decisions and reduce risks, accurate stablecoin price predictions are crucial [3], [4]. Finding a good algorithm that accurately captures the underlying market trends, patterns, and dynamics is essential for achieving reliable and responsive stable Bitcoin forecasts. Three well-known cryptocurrencies with a significant impact on blockchain and digital assets are ADA Cardano (or ADA Coin), Binance (BNB), and Ethereum (ETH) [5], [6]. Moreover, the Cardano blockchain platform's native cryptocurrency is designed to provide a scalable and secure framework for decentralized applications and smart contracts. The ADA coin emphasizes scientific diligence and is essential in staking, transactions, and governance within the Cardano ecosystem [7]. BNB Coin, a native cryptocurrency of the Binance exchange, has attained immense popularity in the crypto market. The Binance network runs on BNB coin, which offers users advantages, including lower transaction fees, access to token sales, and numerous services like staking and lending. ETH Coin, the native coin of the Ethereum blockchain, is widely recognized as a predecessor in programmable blockchain technology. ETH coin, the second-largest cryptocurrency by market capitalization, is employed for payments, decentralized financial applications, and as compensation for Ethereum network miners [6], [7], [8], [9]. This paper aims to find and assess appropriate algorithm(s) for highly volatile cryptocurrency forecasting. We want to create a model that can predict digital coin prices accurately and quickly using historical market data. Additionally, this paper will contribute to market analytics and offer helpful information to market participants looking to improve their risk-taking and investing plans.

A. CONTRIBUTIONS

The main objectives of this paper are listed below:

1) This paper comprehensively evaluates the most advanced data-driven ML and DL models, focusing on three of the top five cryptocurrencies: ADA Cardano, Binance, and Ethereum. This paper presents for the first time three of the top five cryptocurrencies studied in a single manuscript. Remarkably distinct in their values, spanning lengths of 10, 100, and 1000 digits, these coins are the foundation for our comprehensive assessment of the cutting-edge ML and DL models.
2) This paper offers an in-depth and inclusive analysis compared to other existing research and visualizes cryptocurrency forecasts date-wise with a 10-day window size. The results of this research not only contribute to the existing literature on cryptocurrency and also offer practical commercial value as an indicator of coin trends for investors.
3) This paper introduces an innovative approach to predicting cryptocurrency prices with ML and DL models within the realm of Data Science. Our primary goal is to improve decision-making while providing a comprehensive view of our study without obscuring potential shortcomings. This distinctive approach remains to be explored in existing works.

B. IMPLEMENTATION PHASES

This research will use a practical Data Science approach that includes the following steps to find the appropriate algorithm(s) for highly non-linear cryptocurrency forecasting:

a. **Data collection:** We will collect historical price information for stablecoins from reliable sources. Information like trading volume, market sentiment, and pertinent news events will also be considered.
b. **Preprocessing and Feature Engineering:** The preprocessing techniques will be applied to the obtained data to deal with missing values, outliers, and normalization. Some Feature engineering techniques will be used to extract critical characteristics that accurately reflect the underlying market dynamics.
c. **Algorithm Selection:** The performance of various forecasting supervised Machine learning algorithms(such as ARIMA, Support Vector Regressor, and Facebook Prophet) and Deep learning models (such as Unidirectional LSTM, Bidirectional LSTM, and Stacked LSTM), will be compared.
d. **Model Development and Evaluation:** The forecasting performance of the chosen algorithms will be validated using relevant evaluation measures, such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), R- Squared (R2), and Mean Squared Error (MSE) after they have been built and trained on historical data.
e. **Model Refinement and Validation:** More data will be used for further refinement and validation to evaluate the selected approach's robustness, dependability, and scalability.

This paper offers insightful information and valuable tools to aid decision-making in the dynamic cryptocurrency market by identifying suitable algorithm(s) for highly non-linear cryptocurrency forecasting. The results will advance the field of cryptocurrency analytics, increase understanding of stablecoin price dynamics, and facilitate the development of more accurate forecasting models. Additionally, the research will help traders, investors, and market participants by empowering them to make more knowledgeable choices and successfully navigate the turbulent cryptocurrency ecosystem.

Several practical implications will follow from successfully identifying the algorithms that work well for stable currency forecasting. It will give stablecoin issuers a valuable tool to stabilize things and boost trust in their coins. Additionally, it will help investors manage their risk exposure in the volatile digital market and make informed decisions. The results of this study help advance our knowledge of algorithmic stability mechanisms and forecasting methods as they apply to cryptocurrencies.

C. THE PAPER ORGANIZATION

This paper is constructed as follows: A thorough review of the literature conducted in Section II discusses the variables affecting the stability of cryptocurrencies. Section III presents the Materials and Methods, including various ML and DL algorithms that forecast the prices of the ADA, BNB, and ETH coins. Section IV explains that simulation Requirements and Environment setup have established various fundamental results of the selected ML and DL models. The presentation of Results and Analysis can be found in Section V, followed by a Discussion in Section VI. Finally, the Conclusion of our study is presented in Section VII.

II. RELATED WORK

There need to be more studies and research work to predicate the price fluctuation of Cryptocurrency (Bitcoin) trading.

Nakamoto introduced the first cryptocurrency, Bitcoin, in October 2008 [10], [11], and it captured 35% shares of the total world market capitalization [12]. Early research on Bitcoin debated if it was, in fact, another type of currency or a purely speculative asset, with the majority of the authors supporting this last view on the grounds of its high volatility, extreme short-run returns, and bubble-like price behaviour. This study [13] has suggested first identifying the changes in prices in daily routine, and then this trend represents knowledge to use for Bitcoin digital currency. Additionally, this scheme has employed linear regression and decision trees to measure currency accuracy using five-day prediction values. However, the accuracy level must be increased to efficiently compare the predicted values using other machine learning algorithms with real-time trading. Moreover, the Bayesian neural networks (BNN) have been used with blockchain and compared with the linear regression model to analyze the predication values according to the real-time currency accuracy [14], [15]. The limitation of this scheme is not to predicate the exact real-time currency accuracy according to the experimental values. Subsequently, this scheme has also not shown how many days of data are included for simulation and results analysis. The sentiment analysis performance is measured based on three machine learning approaches that have been considered for investors versus the volatility of the cryptocurrency prices [16]. Thus, the limitation of this scheme is not mentioning the window lengths of days and is also not satisfactory accuracy value achieved, which may exactly predicate the importance of currency by ensuring the trust level of people to invest. The authors of this suggested study [17], [18] have considered the Recurrent Neural Network (RNN) based on the deep learning model for cryptocurrency predictions exchange rates. Moreover, the Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Bi-Directional LSTM (Bi-LSTM) are the sub-models of the RNN used for cryptocurrencies of Bitcoin (BTC), Ethereum (ETH), and Litecoin (LTC) to predicate the exchange rates and stability of the currency. The Bi-LSTM outperformed the other two cryptocurrencies, considering performance parameters RMSE and MAPE. However, this suggested scheme considered only five days' prediction values, which is insufficient for the unstable financial situation by not considering other vital cryptocurrencies. Additionally, this scheme has not specified the exact initial cost of the cryptocurrency, which may assist the client in prediction. Furthermore, the Ethereum (ether) blockchain cryptocurrency is presented for change rates by [19], [20] and evaluated by using linear regression (LR) and support vector machine (SVM) machine learning algorithms. This scheme also claims different window lengths of days and suggests using SVM compared to LR by producing a high accuracy prediction rate. This scheme has not shown different window lengths and considered only one cryptocurrency. These are limitations of this scheme by not considering other cryptocurrencies nor showing different window lengths. Thus, such evaluation may not fit the exact values of cryptocurrency prediction. Another study [21], [22] has used the long short-term memory (LSTM), the bidirectional LSTM (Bi-LSTM), and the gated recurrent unit (GRU) as RNNs deep learning algorithms for prediction of the cryptocurrency change rates. Additionally, these RNN-based deep learning models have tested on five cryptocurrencies that included Bitcoin (BTC), Ethereum (ETH), Cardano (ADA), Tether (USDT), and Binance Coin (BNB). LSTM and Bi-LSTM have outperformed in the accurate prediction of changes. However, this suggested scheme has not considered the window lengths of days or included other vital cryptocurrencies. While searching for the accuracy of the Box-Jenkins statistical method to forecast and share price performance for the oil and gas industry in Malaysia, this study [23] discovered that the performance of Gas Malaysia Berhad could be accurately predicted using the autoregressive integrated moving average (ARIMA) model. Similarly, this study has not shown the window lengths of days nor the significance of the ARIMA Model. Like Malaysia, the authors of [24] compared the linear techniques and found that the seasonal ARIMA model yields better estimation for short-term projections in Qatar. These projections of the short-term currency in circulation (CIC), the seasonal ARIMA model's range of forecasting errors is less than 100 million Qatari riyals. The authors of [25] discussed the importance of stock market forecasting techniques and explored the problems associated with their use. Additionally, the findings demonstrated the drawbacks of utilizing traditional methods for cryptocurrencies and that the ARIMA models can help predict broad market movements. However, this scheme has the same limitations as aforementioned, such as there is no information on the window lengths of days and considered only the ARIMA model. Nonetheless, Jadevicius and Huston [26] reveals that ARIMA is a good tool for evaluating broad market price movements. The government and central bank can forecast the national house price inflation using the ARIMA model approach. For timing purposes, the investors can incorporate forecasts from ARIMA models into their trading strategies. Strategic planning can be changed if this player predicates the future investment changes [27], [28], [29], [30], [31]. The ensemble learning method, gradient boosted trees model, neural net model, and K-Nearest Neighbor (K-NN) model have been employed for cryptocurrencies such as Bitcoin, Dash, Ethereum, IOTA, Litecoin, NEM and NEO to predicate the volatility [32]. The considered window lengths are 30 days, and the authors accomplished a better accuracy prediction rate without mentioning cryptocurrency using the ensemble learning method. However, this scheme has not considered other important cryptocurrencies like Binance and ADA Cardano, which will be beneficial for people to understand the market situation. Similarly, [33], [34] has the same issues as described in [32]. Similarly, this study [35] has considered AMP, Ethereum, Electro-Optical System, and XRP cryptocurrencies and applied LSTM to predicate the change rates. Furthermore, the authors of this paper claim that the suggested model can predicate for 180 days. However, this scheme also does not consider other cryptocurrencies, which could benefit people investing in digital amounts at the right time. Moreover, the three cryptocurrencies, e.g. Bitcoin, Ethereum, and Litecoin, are employed to predicate future investment in cryptocurrency [36]. These currencies were tested using different machine-learning algorithms of linear models, random forests, and support vector machines. In conclusion, Ensemble 5 was the better prediction model based on the actual values and tested values for Ethereum and Litecoin cryptocurrencies. However, the drawbacks of this scheme should be considering other important cryptocurrencies for better investment and earning. The Ethereum cryptocurrency is used to predicate the future and investment without giving knowledge of which machine/deep learning model is used [37]. Further, this study has several limitations, as previously mentioned in this literature.

The digital market trend is globally increasing because of the salient features of cryptocurrencies, such as decentralization, market diversity, and accessibility. Therefore, this requires a predictive analytics system to monitor digital coins' sale/purchase volume frequently. In such cases, all the communities, especially the stock markets and traders, will be streamlined with simplified reporting.

III. MATERIALS AND METHODS

A. OVERVIEW

This paper forecasts three popular cryptocurrencies: Ethereum, Binance, and ADA Cardano. Cryptocurrency has proved a present potential to people across the globe. The investors buy the USDT (USD Tether) over the websites or mobile Applications on Android and iPhone platforms for which they can buy the crypto coin for its value. There exists high volatility in this market of every cryptocurrency. Investors often monitor the market buying and selling trend in run-time using popular platforms like CoinMarketCap and Binance with other websites. In contribution to this evolving technology, this research aims to perform future forecasting of the daily price of the proposed coins in the upcoming ten days using Machine learning and Deep Learning models appropriate for such time series forecasting. Additionally, the six most advanced data-driven regression algorithms of both Machine Learning and Deep Learning have been used on the proposed cryptocurrency coins for forecasting:

- SVR (Support Vector Regressor)
- ARIMA (Auto Regressive Integrated Moving Average)
- Facebook Prophet
- Unidirectional (Uni) LSTM
- Bidirectional LSTM (BLSTM)
- Stacked LSTM

A typical Data Science process is performed to gain better insights into our research, described in Figure 1. The following subsections explain valuable insights and knowledge about datasets, data preparation, analysis, model training, and testing, as well as the prediction of results.

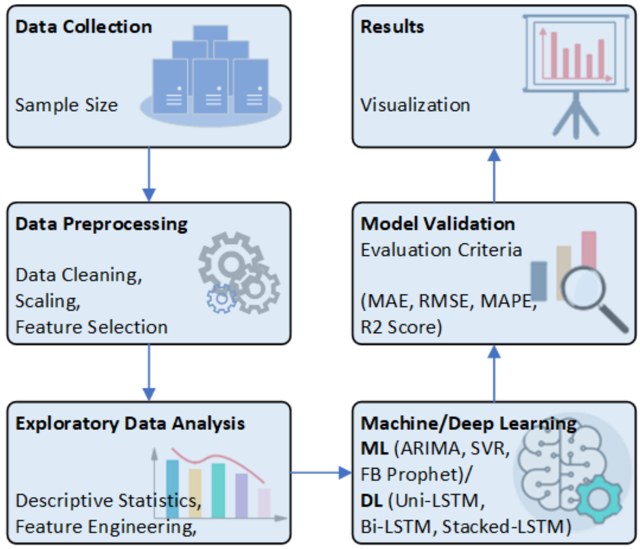

FIGURE 1. Typical overview of working steps of the ML and DL models.

B. DATA COLLECTION

The original data of the cryptocurrency has been collected from the Kaggle Repository (see Dataset Obtained section) containing 425 coins produced by the author from Yahoo (see Dataset Obtained section). In this paper, we have taken Ethereum (ETH), Cardano(ADA), and Binance(BNB) coins datasets exhibiting currencies in the USDT price. Furthermore, we have specifically selected these coins with varying values, such as ADA with ten digits, BNB with a hundred, and ETH with a thousand-digit price, for having a comprehensive evaluation of our proposed machine learning and deep learning algorithms based on diverse sets of values. By doing so, we aim to enhance the accuracy and reliability of our analysis and obtain more precise insights into the performance of our algorithms. Additionally, they are

TABLE 1. ADA cardano dataset information.

S.No	timestamp	adjclose	open	high	volume	low	close
1	11/9/2017	0.0320	0.0251	0.0350	18716200	0.0250	0.0320
2	11/10/2017	0.0271	0.0322	0.0333	6766780	0.0264	0.0271
3	11/11/2017	0.0274	0.0268	0.0296	5532220	0.0256	0.0274
4	11/12/2017	0.0239	0.0274	0.0279	7280250	0.0225	0.0239
5	11/13/2017	0.0258	0.0243	0.0263	4419440	0.0234	0.0258

TABLE 2. Binance dataset information.

S.No	timestamp	adjclose	open	low	high	volume	close
1	11/9/2017	1.9907	2.0531	1.8939	2.1742	19192200	1.9907
2	11/10/2017	1.7968	2.0077	1.6447	2.0694	11155000	1.7968
3	11/11/2017	1.6704	1.7862	1.6142	1.9177	8178150	1.6704
4	11/12/2017	1.5196	1.6688	1.4625	1.6727	15298700	1.5196
5	11/13/2017	1.6866	1.5260	1.5176	1.7350	12238800	1.6866

TABLE 3. Ethereum dataset information.

S.No	timestamp	adjclose	low	open	volume	close	high
1	11/9/2017	320.88	307.05	308.64	893249984	320.88	329.45
2	11/10/2017	299.25	294.54	320.67	885985984	299.25	324.71
3	11/11/2017	314.68	298.19	298.58	842300992	314.68	319.45
4	11/12/2017	307.90	298.51	314.69	1613479936	307.90	319.15
5	11/13/2017	316.71	307.02	307.02	1041889984	316.71	328.41

comparatively very stable coins and ranked in the top five list with the highest market capitalization exceeding one billion US dollars in investment after Bitcoin. The dataset files with the Comma Separated Values (.csv) extension contain eight features with the daily times series summary of five years old threshold timestamp starting 11-9-2017 - 9-28-2022 with weekly update frequency. Tables 1, 2 and 3 depict the initial five days' dataset information of ADA, BNB, and ETH, respectively.

C. DATA PRE-PROCESSING

To ensure the effectiveness of the proposed models, we must undertake data preparation. In this case, the regression analysis involves a qualitative dependent variable, denoted as 'Y' (output), necessitating the proper preparation of the CSV file to align with the requirements of the Regression algorithm.

1) DATA CLEANING

In the CSV file of each coin, there are seven features in total, indicating each coin timestamp (in 24 hrs.), adjusted close, open(coin), high, volume, low and close. As we solve a univariate time series problem, the features except for the 'timestamp' and 'adjusted close' were cleaned.

2) FEATURE SELECTION

In the coin datasets, the "timestamp" and "adjusted close" features have been employed for time series forecasting. The "timestamp" feature represents dates, considered as the independent variable x, while the "adjusted close" represents corresponding prices, treated as the target variable Y.

3) FEATURE SCALING

Target features have been normalized between zero and one using the *Min-Max Scalar* technique to improve performance and stabilize the SVR and Prophet models.

D. EXPLORATORY DATA ANALYSIS

The descriptive statistics of the proposed coin's target variable were thoroughly examined using *Python's describe()* function, as calculated and described in Table 4. Our observations, encompassing numerical and visual aspects, reveal substantial variability over time. Therefore, we have chosen the most recent data, exhibiting the smooth exponential growth, to ensure the stability and robustness of each coin's sample. The ADA sample size spans from January 1, 2021, to the final date, September 28, 2022, encompassing 636 days. For BNB, the sample size spans from November 1, 2020, to September 28, 2022, totalling 697 days. Similarly, the ETH sample size covers September 1, 2020, to September 28, 2022, spanning 697 days, represented in Figure [2,3,4], respectively.

TABLE 4. Descriptive statistics of the proposed coins.

Statistic	ADA	BNB	ETH
count	1785	1785	1785
mean	0.497	139.749	1120.830
std	0.639	184.834	1217.813
min	0.023	1.510	84.308
25%	0.059	13.835	207.082
50%	0.142	22.884	462.436
75%	0.805	293.407	1817.624
max	2.968	675.684	4812.087

1) DATA SPLITTING

Samples were extracted within specific periods. For ADA, the sample covered 636 days from January 1st, 2021, to September 28th, 2022*(636 days)*.For ETH,September 1st, 2020, to September 28th,2022 *(758 days)*. Finally, for BNB, sample taken between November 1st, 2020 and September 28th, 2022 *(697 days)*, as depicted in Table 5.

TABLE 5. Data splitting of the proposed samples.

Currency Sample	Total Days	Training Set Size	Test Set size
ADA	636	74%(468 days)	26% (168 days)
ETH	758	74% (560 days)	26% (198 days)
BNB	697	70% (487 days)	30% (210 days)

Subsequently, the samples of the proposed coins were sequentially split into training and test sets using the *"train_test_split"* function to ensure the quality robustness and to avoid overfitting of the proposed algorithms. For ADA and ETH, 74% of the data was allocated to the training set, while the remaining 26% was designated as the test set. In the case of BNB, 70% of the data were utilized for the training set, and the remaining 30% were assigned to the test set to evaluate the models' performance presented in Figures 5, 6 and 7, respectively. The training dataset of

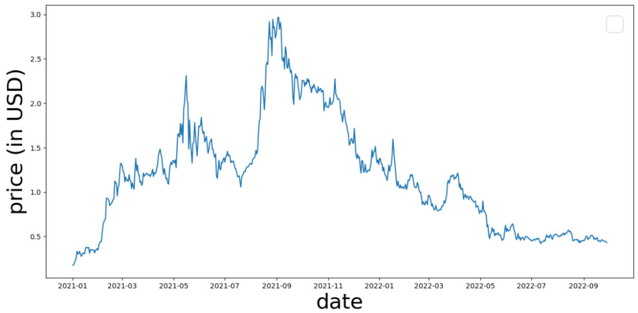

FIGURE 2. ADA cardano sample for algorithms training.

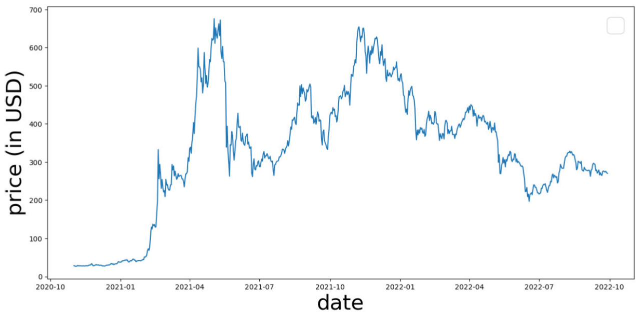

FIGURE 3. Binance sample for algorithms training.

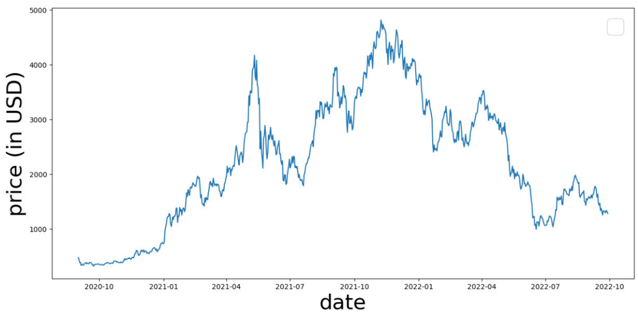

FIGURE 4. Ethereum sample for algorithms training.

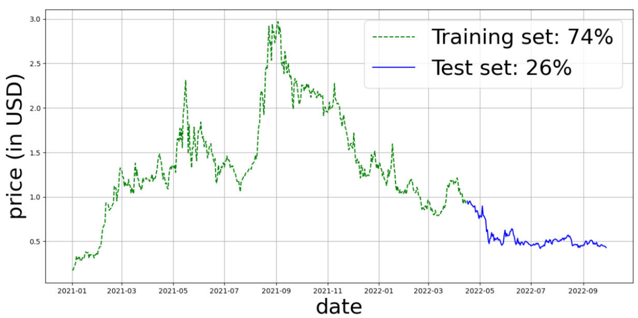

FIGURE 5. Splitting of the ADA sample into training and test set.

each cryptocurrency is used to train the ML and DL models, and then the final performance of each model is evaluated using the test set. For a comprehensive evaluation, we have kept static the above training and the test set for all the models.

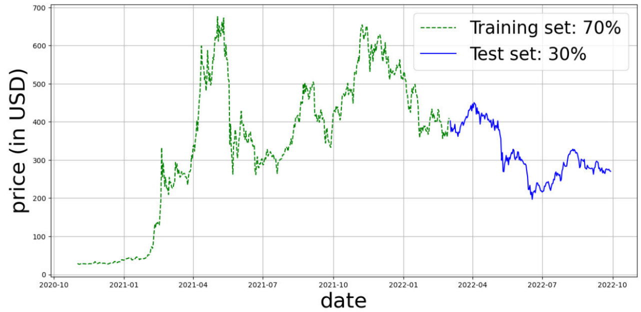

FIGURE 6. Splitting of the BNB sample into training and test set.

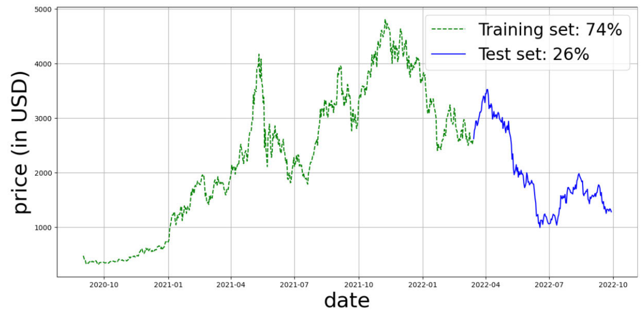

FIGURE 7. Splitting of the ETH sample into training and test set.

E. PROPOSED MODELS

1) SUPPORT VECTOR REGRESSOR

Support Vector Regression is a supervised ML technique used for regression analysis. This technique computes the hyperplane that effectively fits the non-linear data points. The hyperplane is responsible for maximizing the boundary around the data points. The margin boundary is the distance between the nearest points and the hyperplane. A typical kernel works as a hyperparameter, reading the given training data. SVR requires picking the correct kernel to be trained effectively and to minimize outliers; the three kernels were checked over the model training, i.e. Linear, Polynomial and the Radial Basis (RBF) Function. Among these, the *rbf* kernel has been selected after analyzing the input vectors of each cryptocurrency graphically to be in a Gaussian pattern. The objective function minimizes the differences between the predicted and the actual values of the dependent variable while keeping the margin around the hyperplane to be maximized, as described in Figures 8 (ADA Coin), 9 (BNB) and 10 (ETH).

2) AUTO-REGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA)

The working mechanism of the ARIMA algorithm is shown in Figure 11. In the first step, it is required to collect and pre-process the time series data. It is then tested using the Augmented Dicky Fuller test to determine whether the time series data is stationary. If the data is not stationary, differencing is performed to make the data stationary and then determine the appropriate values of p, d, and q parameters for

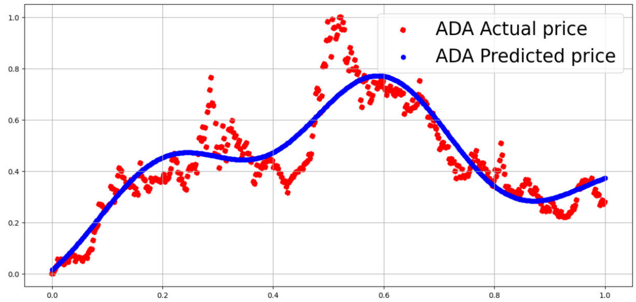

FIGURE 8. **ADA: SVR model fitting performance over training set.**

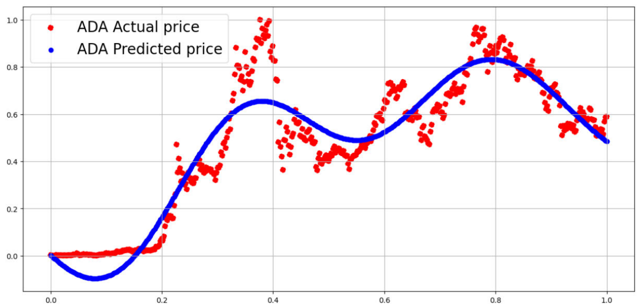

FIGURE 9. **BNB: SVR model fitting performance over training set.**

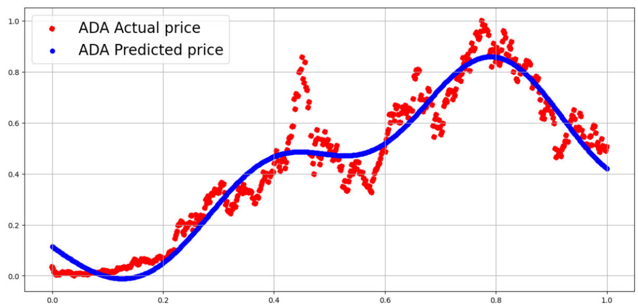

FIGURE 10. **ETH: SVR model fitting performance over training set.**

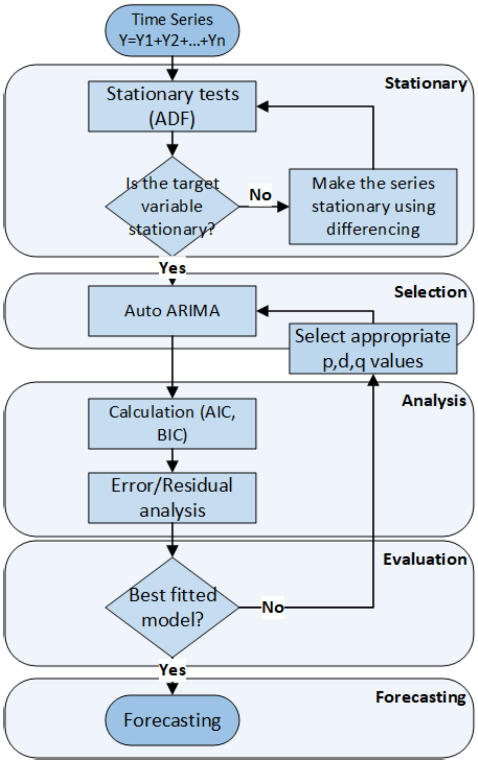

FIGURE 11. **The working flow of the ARIMA algorithm.**

the ARIMA model. In order to do this, the autocorrelation function (ACF) and partial autocorrelation function (PACF) of the differenced time series data are plotted by using *plot_acf* function. The ACF plot is used to identify the value of q, and the PACF plot to identify the value of p. The value of d can be determined by the number of times differencing was applied to make the time series stationary. Afterwards, the ARIMA model is fitted to the time series data using maximum likelihood estimation. The analysis is performed on the residuals of the fitted model to ensure that they are white noise (i.e., have zero mean, constant variance, and no autocorrelation). Finally, the ARIMA model is validated using statistical measures such as the MAE, MAPE, RMSE and R-squared formulas to evaluate the model's performance. Once the ARIMA model has been validated, it forecasts future time series values.

a. **Testing for Stationarity of training set:** For checking the stationarity of the training set, we have used *Augmented Dicky Fuller*, also called the Adfuller test. Using Dicky fuller test, the hypothesis testing with p-value= 0.05 have been computed by considering the Null hypothesis H_0 as non-stationary and the Alternate hypothesis H_1, representing the training data as stationary. The results showed our proposed training sets as non-stationary. To make the data stationary, we differentiated data with different orders, i.e., the ARIMA model's value of 'q', until it became stationary.

b. **Optimization of the hyperparameters (p,d,q) using Auto-ARIMA Technique:** The *"Auto-ARIMA"* is an optimization method we have performed to determine the p,d, and q parameters for the ARIMA model. This technique searches for the lower *Akaike's Information Criterion* (AIC). The lower AIC of the ARIMA's p,d, and q parameters is considered an optimized model for forecasting. The results recommended by the Auto-ARIMA on our coins training sets showing the lowest AIC score for the order (p,d,q) are (1,1,1), (5,1,5), and (2,1,3) for ADA, BNB, and ETH training sets, respectively.

With the *Auto-ARIMA* recommendation of the (p,d,q) values, we have also analyzed the training set of each coin graphically using *Auto-Correlation Function(ACF)* and *Partial Auto-Correlation Function(PACF)* methods by taking different differentiation orders. Finally, we concluded and used the ARIMA values of the (p,d,q) as (1,1,1), (5,2,5), and (2,1,3) for ADA, BNB, and ETH, respectively.

3) FACEBOOK PROPHET

The Facebook Prophet (FB) is becoming a popular ML tool created by Facebook's Core Data Science team. The algorithm became popular due to its accurate forecasting by automatically modelling the underlying pattern of the given time series dataset. In this way, the algorithm automatically handles the multiple levels of seasonality modelling using the Fourier series, including the weekly, monthly, and yearly patterns. It thus becomes flexible for the developer to visualize data with any dates. The FB model uses the Bayesian approach to decompose the time series data into its underlying components: trend, seasonality, and holidays. Thus, it estimates these components and their uncertainty and combines them for forecasting. It also provides built-in diagnostics with graphical representations to cope with model interpretation and performance. Some salient features of the model include short forecasting with large datasets and easy-to-use, flexible, and accurate forecasting of complex time series data. Many researchers and industries have used this model for various applications. The workflow of the model is explained in Figure 12.

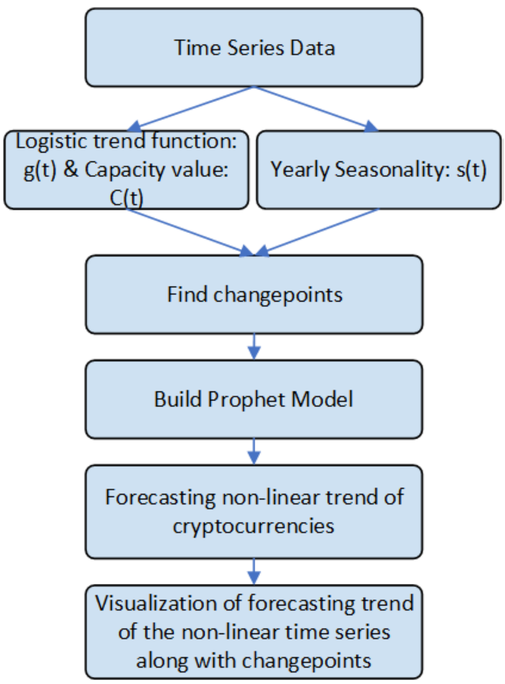

FIGURE 12. The working flow of the FB prophet algorithm.

Initially, the proposed cryptocurrency coins with the expected date-time format (already done in data preparation) were selected, and their time series data was processed. Next, two features were selected from the data - the 'Date' and 'Adjusted close' columns. The 'Date' column was renamed as *'ds'*, and the 'Adjusted close' column was renamed as *'y'*, as the algorithm only expects these names during the training phase. The FB Prophet was built on the training set, and the trained model was used to forecast future values. The model provided the model training details with the 'tail' method. Finally, the model was evaluated by comparing the test set and the predicted values. The build models of the corresponding three coins were displayed using the *plot function*. The following visualization provides a clear view of the model with the training set of each coin, as shown in Figures 13 (ADA), 15 (BNB) and 17 (ETH), respectively. The dotted points in each graph show the actual value, while the blue line shows the predicted values on the corresponding dates. The trained model fitting over the last two weeks' instances of each coin for the forecasting trend was shown in Figures 14 (ADA), 16 (BNB) and 18 (ETH), respectively.

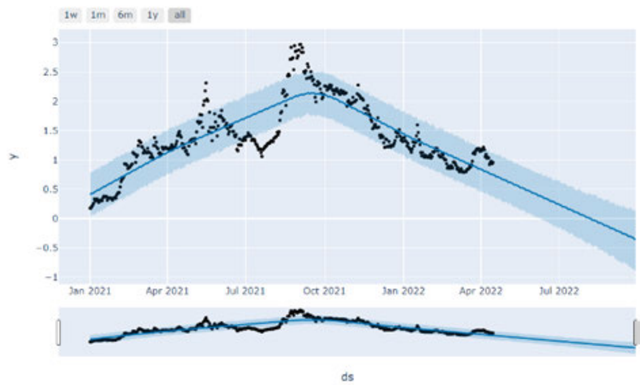

FIGURE 13. ADA model fitting of FB prophet.

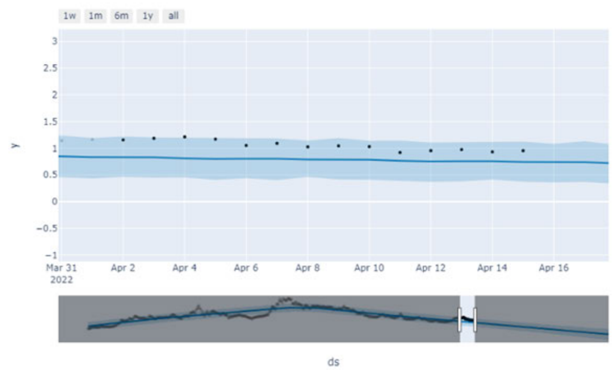

FIGURE 14. ADA - Week-wise representation of model fitting performance.

4) UNIDIRECTIONAL LSTM

A Unidirectional *(Uni)* LSTM network consists of a sequence of LSTM cells, each of which has an input gate i, a forget gate f, and an output gate o, as well as a memory cell c and a hidden state h. At each time step t, the UniLSTM takes an input vector x_t and the previous hidden state h_{t-1} as inputs. The input gate i_t determines how much of the input should be added to the memory cell, while the forget gate f_t determines how much of the previous memory cell value

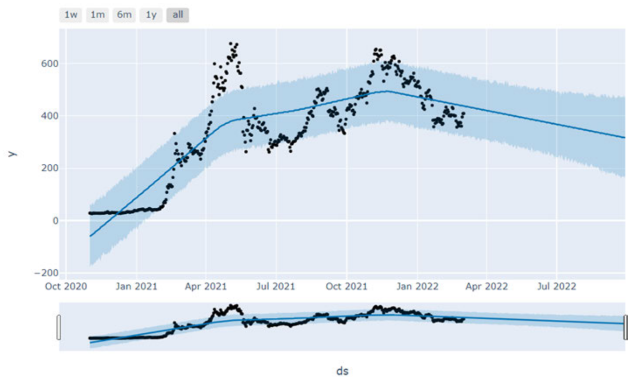

FIGURE 15. BNB model fitting of FB Prophet.

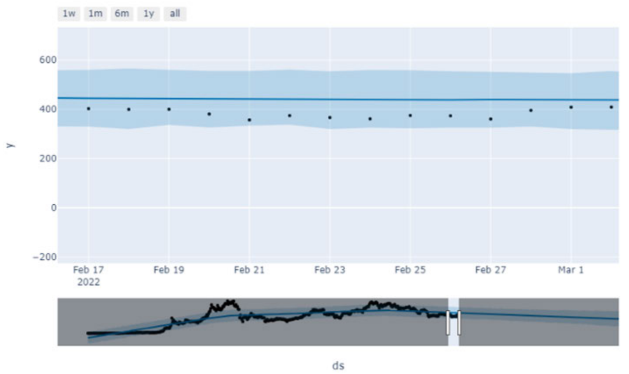

FIGURE 16. BNB - Week-wise representation of model fitting performance.

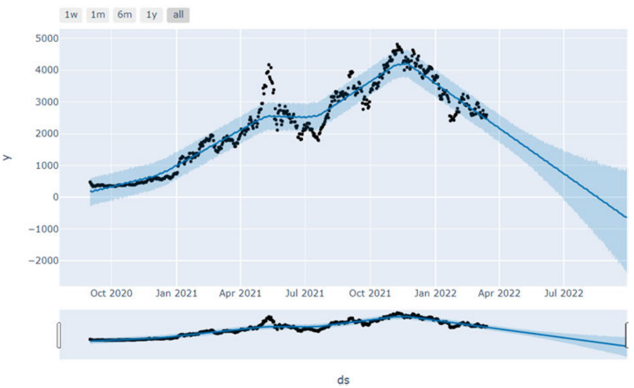

FIGURE 17. ETH model fitting of FB Prophet.

c_{t-1} should be retained. The candidate activation vector g_t is computed based on the input x_t and the hidden state h_{t-1}, and is scaled by i_t to determine how much of it should be added to c_{t-1}. The memory cell c_t is then updated as a weighted combination of f_t times c_{t-1} and i_t times g_t. The output gate o_t determines how much of the current memory cell value c_t should be output as the hidden state h_t. Finally, the output y_t is computed as a function of h_t and can be used for prediction.

Initially, the sequential layer and a single hidden layer of LSTM with 50 neurons were declared. The *ReLU* activation

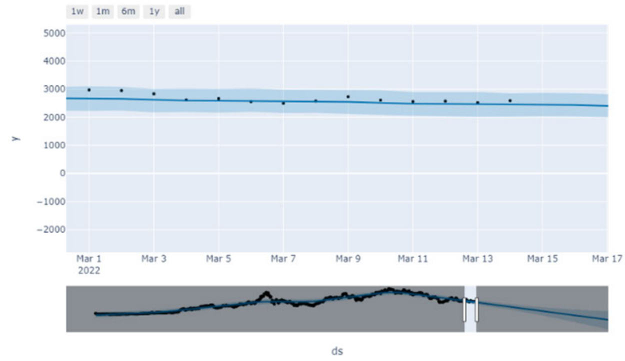

FIGURE 18. ETH - Week-wise representation of model fitting performance.

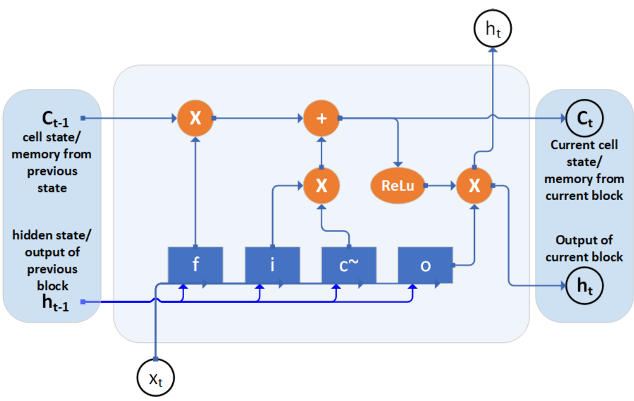

FIGURE 19. The Working flow of undirectional lstm algorithm.

function has been used to learn the Non-linear dependencies effectively by the neural network. The training set in a 3-dimensional input shape is passed through the model, as shown in Figure 19. The trained model computation is obtained with 200 epochs of the Uni-LSTM algorithm. After picking the optimal hyper-parameters for the model, the model computed loss function with the *Mean Squared Error (MSE)* of 0.0109 (ADA), 493.4 (BNB), and 21570.22 (ETH) as depicted in Table 6. We have used *Adam Optimizer* for efficient neural network weight adjustment for optimization purposes. Furthermore, to utilize the trained model effectively in all three coins, Initially, we provided the last three days' target variables of the training set as the initial three inputs to become the input gate, forget gate, and the memory cells for the next day forecasting. This way, a while-loop was operated for the ten forecasting days, and the desired outputs were obtained. We tuned the hyperparameters for optimization, i.e., increasing and decreasing the number of neurons, and tried the model with *Tanh* activation function, thus resulting in poor predictions than *ReLU*.

5) BIDIRECTIONAL LSTM

A Bidirectional LSTM (BLSTM), also called bidirectional GRU, as shown in Figure 20, is a recurrent neural network that processes the input sequence in both the forward and backward directions, using two LSTM layers. When making

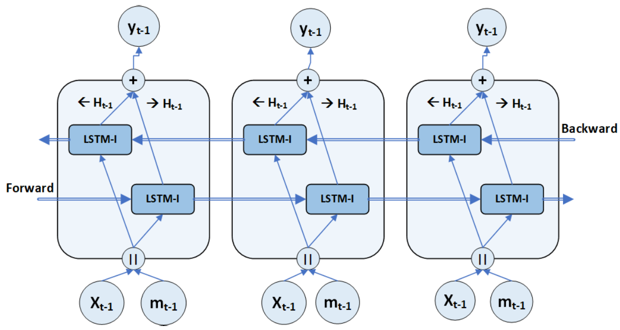

FIGURE 20. The working flow of blstm algorithm.

predictions, the neural network takes the values from the past training set and the future. It is beneficial for our currency forecast, where the context of the current time depends on the previous two values and future events. The forward LSTM processes the input sequence from the beginning to the end, while the backward LSTM processes it from the end to the beginning. At each timestep, the outputs from both LSTMs are concatenated to obtain the final bidirectional output sequence. The input gate (i), forget gate (f), cell state (c), and output gate (o) are used in both the forward and backward LSTMs to control the flow of information, as shown in Figure 20. The model training performance computed by the *MSE* as a loss function is 0.0083 (ADA), BNB (512.2660), and 17782.30 (ETH), as shown in Table 6, respectively.

6) STACKED LSTM

It begins by declaring a sequential layer consisting of two hidden layers of LSTM cells, described in Figure 21. Each layer contains 50 neurons, and the ReLU activation function is used effectively to learn non-linear dependencies in the neural network. The model is trained on a 3-dimensional input shape, where the input vector x_t and the previous hidden state c_{t-1} are passed through the model. The model is trained using 200 epochs of the Stacked LSTM algorithm, as shown in Figure 21. During the algorithm training, the *MSE* as a loss function is computed as 0.0086 (ADA), 440.6084 (BNB), and 21748.2773 (ETH), as shown in Tables 6. For efficient neural network weight adjustment, *Adam Optimizer* is used for optimization.

To effectively utilize the trained model in all three coins, the last three days' target variables of the training set are initially provided as the input gate, forget gate, and memory cells for the next day's forecasting. This process is done for ten days using a *while-loop* to obtain the desired outputs. Hyperparameters were tuned for optimization, including adjustments to the number of neurons. Additionally, the

TABLE 6. Loss function computation of LSTM models.

Coin	Uni-LSTM	BLSTM	Stacked LSTM
ADA	0.01	0.0083	0.0086
BNB	493.41	512.26	440.6
ETH	21570.22	17782.306	21748.27

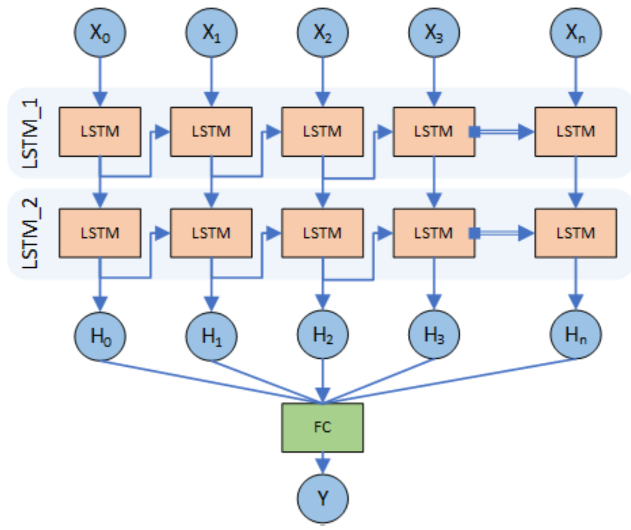

FIGURE 21. The working flow of the stacked LSTM algorithm.

Tanh activation function was experimented with, resulting in inferior predictions compared to *ReLU*.

F. MODEL VALIDATIONS

We have trained and tested the compared models multiple times to eliminate the outliers, and their results are averaged to reduce the biases. The trained model validation w.r.t the test set of the proposed coins was performed with the evaluation criteria to be R-Squared, MAPE, RMSE and MAE, which is explained in the subsequent sections.

1) MEAN ABSOLUTE ERROR (MAE)

Mean Absolute Error (MAE) is the most popular criterion for evaluating the ML model's performance, particularly in regression cases. It is less sensitive to outliers, making it a better metric for evaluating the best model for a given problem. MAE computes the average absolute values difference between the model predicted values and the actual values where each set of differences has equal weight. The MAE evaluates the model performance with a single value ranging from 0 to infinity. This single value of the MAE portrays the overall performance of the algorithm. When the output shows fewer values, the model has better performance or goodness of learning, indicating that the model's forecast values are closer to the actual values. This behaviour can be measured mathematically [38], expressed below:

$$\mathrm{MAE} = \frac{\sum_{j=1}^{n} |y_j - \hat{Y}_j|}{N} \tag{1}$$

where N is the number of coins, y_j is the cryptocurrency Actual value and $\hat{Y}_j$ is the Forecast value.

2) ROOT MEAN SQUARED ERROR (RMSE)

RMSE is a commonly used evaluation technique to measure forecasting quality, particularly in regression cases. The formula calculates the average deviation between the forecast and actual values by taking the square root of the mean of the squared errors. RMSE outputs a single value that exhibits the total error of the ML model. Due to its scale-dependent feature, it is mainly used to assess the quality of different ML algorithms or to check the quality of a single algorithm over different datasets using the derived equation [38], as expressed below:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{j=1}^{n}(y_j - \hat{Y}_j)^2}{N}} \tag{2}$$

where N is the number of coins, y_j is the cryptocurrency's Actual value and $\hat{Y}_j$ is the Forecast value. A model with a lower RMSE value represents better performance, indicating that the predicted values of the cryptocurrency are closer to the actual values.

3) MEAN ABSOLUTE PERCENTAGE ERROR (MAPE)

MAPE is a valuable criterion to quantify the degree of error between the cryptocurrency coin's actual value and the forecast value, expressed as a percentage. It is always positive; lower MAPE means better predictive model performance. The average percentage difference is computed [39] as expressed below:

$$\mathrm{MAPE} = \frac{100\%}{N}\sum_{j=1}^{n}\frac{|y_j - \hat{Y}_j|}{|y_j|} \tag{3}$$

where N is the number of coins, y_j is the cryptocurrency Actual value and $\hat{Y}_j$ is the Forecast value.

4) R-SQUARED (R2)

R-squared, or the coefficient of determination is a statistical technique to evaluate a regression algorithm's quality. R2 measures the goodness of fit that determines how the data fits the regression model. It has a scale between 0 and 1, representing none or all of the variances in the target variable is explained by the input variable(s). The coefficient of determination [40] can be mathematically expressed as given below:

$$\mathrm{R2} = \frac{\bar{\sigma}^2}{\sigma^2} \tag{4}$$

where $\bar{\sigma}^2$ represents the Explained variation, which is the sum of squared differences between the forecast values and the response variable values, while σ^2 denotes the Total variation and is the sum of squared differences between the crypto original values and its corresponding mean value.

5) MEAN SQUARED ERROR (MSE)

MSE is a popular metric for evaluating the forecasting of ML and DL models. It computes the average squared difference between the cryptocurrency forecast values and the original values [41] as expressed below:

$$\mathrm{MSE} = \frac{\sum_{j=1}^{n}(y_j - \hat{Y}_j)^2}{N} \tag{5}$$

where N is the number of coins, y_j is the cryptocurrency's Actual value and $\hat{Y}_j$ is the Forecast value.

Following are the results of the performance evaluation of each ML model using the above validation techniques.

TABLE 7. ADA: ML models performance evaluation on test Set.

Evaluation Criteria	SVR	ARIMA	FB PROPHET
MAE	0.205	0.021	0.35
RMSE	0.233	0.03	0.429
MAPE	0.38	0.218	0.71
R2 Score	-0.63	0.948	-9.52

TABLE 8. BNB: ML models performance evaluation on test Set.

Evaluation Criteria	SVR	ARIMA	FB PROPHET
MAE	113.355	8.381	63.2
RMSE	128.061	11.357	77.63
MAPE	0.354	0.247	0.23
R2 Score	-1910.89	0.971	-0.35

TABLE 9. ETH: ML models performance evaluation on test Set.

Evaluation Criteria	SVR	ARIMA	FB PROPHET
MAE	1208.49	66.967	1085.42
RMSE	1342.11	90	1223.86
MAPE	0.641	0.4324	0.6
R2 Score	-2.49	0.985	-1.9

The evaluation criteria mentioned above are computed to assess the training performance of ML models on the test sets corresponding to ADA, BNB, and ETH (Tables 7, 8, and 9). Interestingly, the ARIMA model outperforms others, demonstrating exceptional performance with the lowest values for MAE, RMSE, and MAPE. The R2-score also displayed predictive solid power for all three cryptocurrencies, with calculated values of 0.948, 0.971, and 0.985 for ADA, BNB, and ETH, respectively. Despite conducting further analysis to identify potential confounding effects, the remaining models exhibit poor training performance, except for the FB prophet model, which shows some good fitting.

IV. SIMULATION REQUIREMENTS AND ENVIRONMENT SETUP

This paper presents to forecast the Ethereum, ADA Cardano, and Binance coins using the six most advanced data-driven regression techniques, which are Support Vector Regression (SVR), Auto Regressive Integrated Moving Average (ARIMA), Unidirectional LSTM, Bidirectional LSTM, Stacked LSTM, and Facebook Prophet. The original dataset of 425 cryptocurrency coins was obtained from Kaggle (see

Appendix I). To better compare the models, this research focuses on three popular coins with different value digits. We have used Python language(latest version:3.9.13) with the IDE, Google Colaboratory, and Jupyter Notebook to conduct the research experiments on the Windows 11 Pro OS with 20GB RAM. Table 5 shows the details of the simulation performance parameters.

TABLE 10. Simulation parameters setup.

Parameter	Description
OS	Windows 11 Pro
RAM	20GB
IDE Used	Google Colaboratory,Jupyter Notebook
Implementation	Python(latest version:3.9.13)
Dataset files	425 cryptocurrencies, each file size : 200 - 230 KB
Dataset timestamp	11/9/2017 to 9/28/2022
Proposed Coins	Cardano(ADA), Ethereum(ETH), Binance(BNB)
Predict duration	24 Hrs

The data normalization during training of the SVR model was done using a Min-Max Scaler. At the same time, the Radial Basis Function 'rbf' kernel was set as a hyperparameter for the SVR model after analysis. The ARIMA model was trained with p,d, and q values of (1,1,1), (5,2,5), and (2,1,3) for ADA, BNB and ETH, respectively. The LSTM models with 50 neurons, Mean Squared Error as loss function, and 50 epochs were computed for ADA, BNB, and Eth. In the case of Stacked-LSTM, two hidden layers were used. Numerical experiments were performed on the dataset, and four widely used evaluation metrics, MAE, MAPE, RMSE, and R-squared, were computed.

V. RESULTS AND ANALYSIS

This study strives to create a forecasting system for cryptocurrency using Machine learning & deep learning algorithms. Due to the high volatility trend in the market, we have utilized the real datasets of the three cryptocurrency coins, namely ADA Cardano, Ethereum, and Binance comprising varying digits *(performance measures)* used to gain better insights into each algorithm performance for predictive analytics. Each coin dataset contains the 24 Hr. features like timestamp between Nov 9th, 2017 to Sept 28, 2022(1785 days), volume, high-low rate, and open−close rates. As the cryptocurrency investment rate is progressively growing globally, each coin in the market is gaining immense popularity. In such a situation, we may commercially require simplified reporting for all the communities, especially the traders and stock markets. Therefore, this research has focused on identifying suitable algorithm(s) in forecasting the 24-hour price of each coin using the six most advanced models: Support Vector Regressor, ARIMA, Facebook Prophet, Unidirectional LSTM, Bidirectional LSTM, and Stacked LSTM. In addition, all the models have been implemented using the data science framework to demonstrate the potential for better forecasting. The model will project the returns upon the investor's selection of the coins and the risks they are willing to take. For better evaluation, we have kept constant the training and testing set of each coin for all the models.

After preparing and analyzing each dataset, we extracted samples within the specific time frame. For ADA, the sample selected covered 636 days from January 1st, 2021, to September 28th, 2022. For ETH, the sample spanned 758 days from September 1st, 2020, to September 28th, 2022. Lastly, for BNB, the sample encompassed 697 days from November 1st, 2020, to September 28th, 2022. Next, the samples are sequentially split into training and test sets using the "train_test_split" function discussed in the "Data Splitting" section. For ADA and ETH, 74% of the data is allocated to the training set, while the remaining 26% has been designated as the test set. In the case of BNB, 70% of the data is utilized for the training set, and the remaining 30% have been assigned to the test set to evaluate the models' performance. Various performance parameters discussed in the previous section have been computed for this evaluation. The subsequent part provides a detailed discussion of the results.

A. MODELS PERFORMANCE EVALUATION WITH 10 DAYS PREDICTION INTERVALS

We considered five different validation methods to evaluate date-wise forecasts of the ten days, as discussed in the previous section. The performance evaluation showed best performance in the prediction of ARIMA prices with comparatively minimum values for ADA (MAE = 0.016, RMSE = 0.019, MAPE = 0.03, R2 Score = 0.344), BNB (MAE =10.47, RMSE =13.96, MAPE =0.03, R2 Score = −1.06) and ETH (MAE =72.2, RMSE =88.54, MAPE =0.054, R2 Score =0.547) as depicted in Table 11, 12 and 13, respectively.

TABLE 11. ADA: models performance evaluation with 10 days prediction intervals.

Criteria	SVR	ARIMA	FB Prophet	Uni-LSTM	BLSTM	Stacked
MAE	0.462	0.016	0.21	0.032	0.160	0.1
RMSE	0.464	0.019	0.21	0.04	0.179	0.113
MAPE	0.5	0.03	0.228	0.036	0.176	0.11
R2 Score	-5475.08	0.344	-78.85	-1.86	-56.5	-22.1

TABLE 12. BNB: models performance evaluation with 10 days prediction intervals.

Criteria	SVR	ARIMA	FB Prophet	Uni-LSTM	BLSTM	Stacked
MAE	193.77	10.47	53.12	42.0328	14.92	66.78
RMSE	193.95	13.96	53.87	44.54	17	72.092
MAPE	0.5	0.034	0.14	0.11	0.039	0.176
R2 Score	-92475.8	-1.06	-29.67	-19.97	-2	-53.926

TABLE 13. ETH: models performance evaluation with 10 days prediction intervals.

Criteria	SVR	ARIMA	FB Prophet	Uni-LSTM	BLSTM	Stacked
MAE	1699.9	72.28	530.744	294.03	375.61	288.7
RMSE	1709.5	88.54	558.635	316.42	406.69	310.06
MAPE	0.585	0.054	0.18	0.099	0.1274	0.098
R2 Score	-50427.37	0.547	-16.99	-4.772	-8.536	-4.54

TABLE 14. ADA: models performance evaluation with 10 days prediction intervals.

		Uni-LSTM	Stacked	BLSTM	SVR	ARIMA	FB Prophet
Date	Actual Values	Predicted Price	Predicted Price	Predicted Price	Predicted Price	Predicted Price	Predicted Price
4/16/2022	0.95264	0.9542	0.9638	0.979	0.42598314	0.9517	0.741657
4/17/2022	0.917466	0.9476	0.9799	1.004	0.4318842	0.9548	0.739054
4/18/2022	0.93673	0.9521	0.9922	1.027	0.43816769	0.9189	0.717498
4/19/2022	0.953333	0.9515	1.005	1.05	0.44481402	0.9343	0.706928
4/20/2022	0.937341	0.9506	1.0165	1.072	0.45180133	0.9536	0.711958
4/21/2022	0.910474	0.9515	1.0276	1.093	0.45910556	0.9384	0.711104
4/22/2022	0.907154	0.9514	1.038	1.113	0.46670056	0.9119	0.697527
4/23/2022	0.888503	0.9514	1.0479	1.132	0.47455818	0.9065	0.695486
4/24/2022	0.88635	0.9516	1.0572	1.15	0.48264843	0.8904	0.692883
4/25/2022	0.898695	0.9516	1.066	1.167	0.49093964	0.8853	0.671326

TABLE 15. BNB: models performance evaluation with 10 days prediction intervals.

		Uni-LSTM	Stacked	BLSTM	SVR	ARIMA	FB Prophet
Date	Actual Values	Predicted Price	Predicted Price	Predicted Price	Predicted Price	Predicted Price	Predicted Price
3/3/2022	402.6	411.793	416.104	404.12	197.639	412.8	436.81
3/4/2022	374.3	414.374	422.411	402.502	195.0292	408.1	436.20
3/5/2022	384.9	416.933	429.517	400.8	192.521	372.9	434.36
3/6/2022	375	419.466	436.618	398.678	190.122	381.4	436.01
3/7/2022	381.3	421.972	443.881	396.739	187.841	379.5	434.86
3/8/2022	382	424.453	451.197	394.865	185.687688	381.8	434.69
3/9/2022	393.6	426.909	458.577	392.997	183.6	382.6	433.86
3/10/2022	372.2	429.34	465.997	391.168	181.797	389.9	432.77
3/11/2022	372.2	431.749	473.45	389.378	180.076	380.7	432.16
3/12/2022	372.7	434.134	480.921	387.621	178.516	369.6	430.31

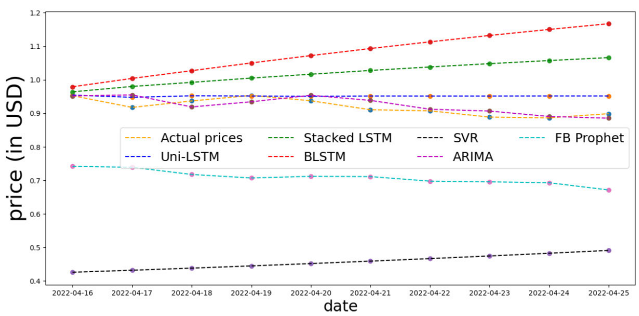

FIGURE 22. ADA: graphical illustration of the 10-day prices forecasting by all models vs. actual trend.

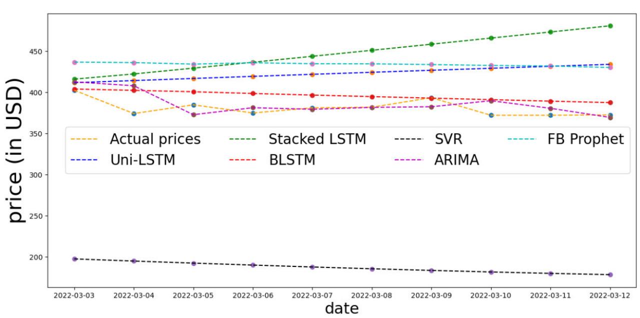

FIGURE 23. BNB: graphical illustration of the 10-day prices forecasting by all models vs. actual trend.

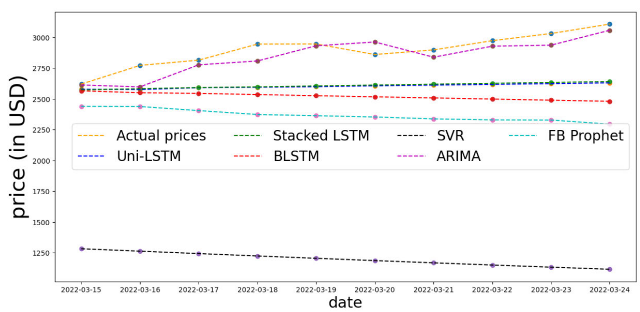

FIGURE 24. ETH: graphical illustration of the 10-day prices forecasting by all models vs. actual trend.

VI. DISCUSSION

When assessing the model performance over ten days prices, as presented in Tables 12 (ADA), 13 (ETH), and 14 (BNB), each model focuses on forecasting specific decimal values for ADA, 100's place values for BNB, and 1000's place digits for ETH. These forecasts were compared to the prices recorded during different time frames for each cryptocurrency. Evaluating the date-wise forecasted prices showed that ARIMA outperformed other models by exhibiting the lowest error, indicating its accurate fitting over the training set. Notably, the ARIMA model accurately predicted a downward trend for ADA and BNB and an upward trend for ETH, as depicted in the scatter plot Figure 22 (ADA), 23 (BNB), 24 (ETH). These showcase the variability of predicted prices within the actual price range. The ability of the ARIMA algorithm to forecast the 10-day trend proves highly advantageous for market analysts, enabling them to make well-informed decisions.

The LSTM variants yielded inaccurate predictions that contradicted the actual values, displaying an opposite trend as evidenced by scatter plots Figure 22 (ADA), 23 (BNB), 24 (ETH). This discrepancy suggests that these predictions

TABLE 16. ETH: models performance evaluation with 10 days prediction intervals.

		Uni-LSTM	Stacked	BLSTM	SVR	ARIMA	FB Prophet
Date	Actual Values	Predicted Price	Predicted Price	Predicted Price	Predicted Price	Predicted Price	Predicted Price
3/15/2022	2620.149658	2578.801	2573.0315	2565.606	1282.30428732	2613.0166	2439.411766
3/16/2022	2772.055664	2575.724	2582.6973	2550.0344	1262.67595617	2597.5062	2438.168846
3/17/2022	2814.854492	2591.257	2592.4207	2544.5293	1243.20959654	2776.7801	2404.554646
3/18/2022	2945.343018	2594.513	2597.568	2535.1416	1223.95480444	2808.0972	2373.021828
3/19/2022	2946.25708	2599.114	2605.1973	2525.547	1204.96284803	2931.4056	2363.475904
3/20/2022	2860.459229	2606.434	2612.2463	2516.5679	1186.28653673	2962.4337	2353.056995
3/21/2022	2897.976563	2611.962	2619.2034	2507.506	1167.98007557	2838.7963	2337.73908
3/22/2022	2973.131104	2617.703	2626.313	2498.442	1150.09890491	2928.1015	2329.528152
3/23/2022	3031.067139	2623.849	2633.361	2489.4446	1132.69952563	2936.6693	2328.285232
3/24/2022	3108.062012	2629.742	2640.4226	2480.4844	1115.83931008	3057.6463	2294.671033

may not substantially assist investors seeking insightful forecasts. Conversely, FB Prophet demonstrated inadequate and inconsistent results with minimal variation. It accurately estimated the trend for ADA and BNB; however, it projected the opposite outcome for ETH, as shown in the corresponding scatter plots. Furthermore, SVR exhibited poor results in all scenarios, indicating the limited capability of its hyperplane to effectively fit the given data, as illustrated in the scatter plots.

VII. CONCLUSION

This paper proposed a predictive analytics system to provide communities with simplified reporting. The system aims to project the returns based on the investor's choice of coins and the risks they are willing to take in this highly volatile environment. A typical data science framework provides deeper insights into our research, incorporating six of the most advanced data-driven ML and DL techniques: SVR, ARIMA, FB Prophet, Uni LSTM, BLSTM, and Stacked LSTM. To thoroughly evaluate each algorithm, we used the three digital currency datasets, ADA, BNB, and ETH, with varying digits as performance measures. Each model was trained on each dataset to forecast the next ten days' 24-hour price. To further improve evaluation, each coin's training and testing samples are kept constant for all models, and their performance is assessed using various validation techniques such as MAE, RMSE, MAPE, R-Square, and MSE. The results of our study show that the ARIMA model outperforms the other models in forecasting these 10s, 100s, and 1000 place digit coins for the next ten days, accurately capturing upward and downward trends. Notably, ARIMA demonstrated the variability of predicted prices within the actual price range. On the other hand, FB Prophet demonstrated good performance to some extent with minimal price variation; it accurately forecasts ADA and BNB trends but incorrectly forecasts ETH coins. Except for the BNB coin, which the SVR model forecast correctly, the LSTM variants and SVR model produce inaccurate forecasts with minimal price variation and opposite trends. Overall, the ARIMA technique's ability to forecast the 10-day trend and prices benefits market analysts, allowing them to make well-informed decisions.

In conclusion, this study shows valuable insights into forecasting stable digital coins ADA, BNB, and Ethereum using advanced ML and DL techniques. However, it is imperative to acknowledge certain limitations that inherently shape our research. These include market volatility and external factors, given that digital coin trading is closely linked with events such as regulatory announcements and security breaches. These events can prompt abrupt and unpredictable trends, potentially challenging the reliability of our optimal models. Despite these limitations, our study highlights the proposed algorithms' performance that helps provide simplified reporting to the traders, laying the robust foundation for future research endeavours. Our next goal is to create a reactive machine for cryptocurrency forecasting, augmented by federated learning techniques focusing on hourly and minute-by-minute price predictions for different coins. This machine will predict trends accurately, allowing us to investigate usage and potential future investments in the global cryptocurrency market.

DATASET OBTAINED

1) https://www.kaggle.com/datasets/tr1gg3rtrash/time-series-top-100-crypto-currency-dataset
2) https://finance.yahoo.com/cryptocurrencies/

REFERENCES

[1] Z. Zhang and R. Zhao, "Good volatility, bad volatility, and the cross section of cryptocurrency returns," *Int. Rev. Financial Anal.*, vol. 89, Oct. 2023, Art. no. 102712.
[2] C. Zhong, W. Du, W. Xu, Q. Huang, Y. Zhao, and M. Wang, "LSTM-ReGAT: A network-centric approach for cryptocurrency price trend prediction," *Decis. Support Syst.*, vol. 169, Jun. 2023, Art. no. 113955.
[3] L. Gambarelli, G. Marchi, and S. Muzzioli, "Hedging effectiveness of cryptocurrencies in the European stock market," *J. Int. Financial Markets, Institutions Money*, vol. 84, Apr. 2023, Art. no. 101757.
[4] S. F. Stefenon, L. O. Seman, V. C. Mariani, and L. D. S. Coelho, "Aggregating prophet and seasonal trend decomposition for time series forecasting of Italian electricity spot prices," *Energies*, vol. 16, no. 3, p. 1371, Jan. 2023. [Online]. Available: https://www.mdpi.com/1996-1073/16/3/1371
[5] E. Bouri, E. Kamal, and H. Kinateder, "FTX collapse and systemic risk spillovers from FTX token to major cryptocurrencies," *Finance Res. Lett.*, vol. 56, Sep. 2023, Art. no. 104099.
[6] S. Lee, J. Lee, and Y. Lee, "Dissecting the Terra-LUNA crash: Evidence from the spillover effect and information flow," *Finance Res. Lett.*, vol. 53, May 2023, Art. no. 103590.
[7] Z. J. Zhang, "Cryptopricing: Whence comes the value for cryptocurrencies and NFTs?" *Int. J. Res. Marketing*, vol. 40, no. 1, pp. 22–29, Mar. 2023.

[8] T. Cui, S. Ding, H. Jin, and Y. Zhang, "Portfolio constructions in cryptocurrency market: A CVaR-based deep reinforcement learning approach," *Econ. Model.*, vol. 119, Feb. 2023, Art. no. 106078.

[9] I. Ghosh and R. K. Jana, "A granular machine learning framework for forecasting high-frequency financial market variables during the recent black swan event," *Technol. Forecasting Social Change*, vol. 194, Sep. 2023, Art. no. 122719. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0040162523004043

[10] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," *Decentralized Bus. Rev.*, vol. 15, p. 21260, Oct. 2008.

[11] R. G. da Silva, S. R. Moreno, M. H. D. M. Ribeiro, J. H. K. Larcher, V. C. Mariani, and L. D. S. Coelho, "Multi-step short-term wind speed forecasting based on multi-stage decomposition coupled with stacking-ensemble learning approach," *Int. J. Electr. Power Energy Syst.*, vol. 143, Dec. 2022, Art. no. 108504. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0142061522005051

[12] W. Zhang, P. Wang, X. Li, and D. Shen, "Some stylized facts of the cryptocurrency market," *Appl. Econ.*, vol. 50, no. 55, pp. 5950–5965, Nov. 2018.

[13] K. Rathan, S. V. Sai, and T. S. Manikanta, "Crypto-currency price prediction using decision tree and regression techniques," in *Proc. 3rd Int. Conf. Trends Electron. Informat. (ICOEI)*, Apr. 2019, pp. 190–194.

[14] H. Jang and J. Lee, "An empirical study on modeling and prediction of Bitcoin prices with Bayesian neural networks based on blockchain information," *IEEE Access*, vol. 6, pp. 5427–5437, 2018.

[15] X. Xu, Y. Zhang, C. A. McGrory, J. Wu, and Y.-G. Wang, "Forecasting stock closing prices with an application to airline company data," *Data Sci. Manag.*, vol. 2023, Oct. 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666764923000437

[16] 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, pp. 1–21, 2018. [Online]. Available: https://scholar.smu.edu/cgi/viewcontent.cgi?article=1039&context=datasciencereview

[17] P. L. Seabe, C. R. B. Moutsinga, and E. Pindza, "Forecasting cryptocurrency prices using LSTM, GRU, and bi-directional LSTM: A deep learning approach," *Fractal Fractional*, vol. 7, no. 2, p. 203, Feb. 2023. [Online]. Available: https://www.mdpi.com/2504-3110/7/2/203

[18] 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.

[19] P. M. A. Sharma, V. V. V. Bhardwaj, A. P. Sharma, R. Iqbal, and R. Kumar, "Prediction of the price of Ethereum blockchain cryptocurrency in an industrial finance system," *Comput. Electr. Eng.*, vol. 81, Jan. 2020, Art. no. 106527.

[20] M. Saad, J. Choi, D. Nyang, J. Kim, and A. Mohaisen, "Toward characterizing blockchain-based cryptocurrencies for highly accurate predictions," *IEEE Syst. J.*, vol. 14, no. 1, pp. 321–332, Mar. 2020.

[21] S. Hansun, A. Wicaksana, and A. Q. M. Khaliq, "Multivariate cryptocurrency prediction: Comparative analysis of three recurrent neural networks approaches," *J. Big Data*, vol. 9, no. 1, pp. 1–15, Dec. 2022.

[22] M. M. Patel, S. Tanwar, R. Gupta, and N. Kumar, "A deep learning-based cryptocurrency price prediction scheme for financial institutions," *J. Inf. Secur. Appl.*, vol. 55, Dec. 2020, Art. no. 102583.

[23] N. A. Bakar and S. Rosbi, "Robust diagnostics of box-jenkins statistical method in forecasting share price performance for sharia-compliant oil and gas sector in Malaysia stock exchange," *Int. J. Sci. Res. Manag.*, vol. 5, no. 7, pp. 6364–6376, 2017.

[24] F. Balli and E. Mousa Elsamadisy, "Modelling the currency in circulation for the state of Qatar," *Int. J. Islamic Middle Eastern Finance Manag.*, vol. 5, no. 4, pp. 321–339, Nov. 2012.

[25] S. Stevenson, "A comparison of the forecasting ability of ARIMA models," *J. Property Investment Finance*, vol. 25, no. 3, pp. 223–240, May 2007.

[26] A. Jadevicius and S. Huston, "ARIMA modelling of Lithuanian house price index," *Int. J. Housing Markets Anal.*, vol. 8, no. 1, pp. 135–147, Mar. 2015.

[27] N. A. Bakar and S. Rosbi, "Data clustering using autoregressive integrated moving average (ARIMA) model for Islamic country currency: An econometrics method for Islamic financial engineering," *Int. J. Eng. Sci.*, vol. 6, no. 6, pp. 22–31, Jun. 2017.

[28] M. H. D. M. Ribeiro, R. G. da Silva, S. R. Moreno, V. C. Mariani, and L. D. S. Coelho, "Efficient bootstrap stacking ensemble learning model applied to wind power generation forecasting," *Int. J. Electr. Power Energy Syst.*, vol. 136, Mar. 2022, Art. no. 107712. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0142061521009376

[29] K. Abbas, M. K. Hasan, A. Abbasi, U. A. Mokhtar, A. Khan, S. N. H. S. Abdullah, S. Dong, S. Islam, D. Alboaneen, and F. R. A. Ahmed, "Predicting the future popularity of academic publications using deep learning by considering it as temporal citation networks," *IEEE Access*, vol. 11, pp. 83052–83068, 2023.

[30] A. Khan, J. P. Li, N. Ahmad, S. Sethi, A. U. Haq, S. H. Patel, and S. Rahim, "Predicting emerging trends on social media by modeling it as temporal bipartite networks," *IEEE Access*, vol. 8, pp. 39635–39646, 2020.

[31] A. Khan, J. P. Li, and M. A. Husain, "Power grid stability analysis using pipeline machine," *Multimedia Tools Appl.*, vol. 82, no. 17, pp. 25651–25675, Jul. 2023.

[32] R. Chowdhury, M. A. Rahman, M. S. Rahman, and M. R. C. Mahdy, "An approach to predict and forecast the price of constituents and index of cryptocurrency using machine learning," *Phys. A, Stat. Mech. Appl.*, vol. 551, Aug. 2020, Art. no. 124569.

[33] X. Xu, C. A. McGrory, Y.-G. Wang, and J. Wu, "Influential factors on Chinese airlines' profitability and forecasting methods," *J. Air Transp. Manag.*, vol. 91, Mar. 2021, Art. no. 101969.

[34] I. Yenidogan, A. Çayir, O. Kozan, T. Dag, and Ç. Arslan, "Bitcoin forecasting using ARIMA and PROPHET," in *Proc. 3rd Int. Conf. Comput. Sci. Eng. (UBMK)*, 2018, pp. 621–624.

[35] M. A. Ammer and T. H. H. Aldhyani, "Deep learning algorithm to predict cryptocurrency fluctuation prices: Increasing investment awareness," *Electronics*, vol. 11, no. 15, p. 2349, Jul. 2022.

[36] H. Sebastião and P. Godinho, "Forecasting and trading cryptocurrencies with machine learning under changing market conditions," *Financial Innov.*, vol. 7, no. 1, pp. 1–30, Jan. 2021.

[37] H.-M. Kim, G.-W. Bock, and G. Lee, "Predicting Ethereum prices with machine learning based on blockchain information," *Exp. Syst. Appl.*, vol. 184, Dec. 2021, Art. no. 115480.

[38] M. Kuhn and K. Johnson, *Applied Predictive Modeling*, vol. 26. Berlin, Germany: Springer, 2013.

[39] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *Int. J. Forecasting*, vol. 22, no. 4, pp. 679–688, Oct. 2006.

[40] J. Gareth, W. Daniela, H. Trevor, and T. Robert, *An Introduction to Statistical Learning: With Applications in R*. Berlin, Germany: Springer, 2013.

[41] T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, vol. 2. Berlin, Germany: Springer, 2009.

HASIB SHAMSHAD received the bachelor's degree in computer science from the University of Engineering and Technology, Peshawar, Pakistan, and the master's degree in systems engineering from the National University of Sciences and Technology (NUST), Islamabad, Pakistan. He has been a Lecturer of computer science with the Department of Computer Science and IT, Sarhad University of Science and Information Technology, since November 2020. His current research interests include artificial intelligence, machine learning, deep learning, big data analytics, and parallel and distributed computing. His focus lies in applying machine learning and deep learning techniques to both mobile and desktop applications, which makes him a potential supervisor for undergraduate projects in these areas.

FASEE ULLAH received the degree from the Faculty of Computing, Universiti Teknologi Malaysia (UTM), Malaysia, in 2017. He has held a Postdoctoral Fellowship position with the University of Macau, from 2019 to 2021, which is the Academic Talented Program of the Government of Macau. Currently, he is an Associate Professor with Computer and Information Sciences Department, Universiti Teknologi Petronas, Bandar Seri Iskandar, PRk, Malaysia. He has published many research papers in reputed impact factor journals and conferences. His current research interests include wireless body area networks, wireless sensor networks, cloud security, smart hash security designing, smart cities, big data analytics, machine learning, deep learning, and the Internet of Things. He was a recipient of the Chancellor Award and the Best Student Award at UTM, during the Ph.D. degree, for his excellent research contributions to wireless communication and health monitoring. He is currently providing reviewing services to IEEE Transactions on Computers, IEEE Transactions on Network Science and Engineering, IEEE Transactions on Cloud Computing, IEEE Access, IEEE Sensor Journal, ACM, and *International Journal of Distributed Sensor Networks*.

ASAD ULLAH received the Ph.D. degree in information and communication engineering (specialization in artificial intelligence) from the Beijing Institute of Technology, China, in 2020. He was an Assistant Professor with Riphah International University, Faisalabad, Pakistan, from 2014 to 2020. He has been an Assistant Professor with the Department of Computer Science and Information Technology, Sarhad University of Science and Information Technology, Peshawar, Pakistan, since October 2020. He has published several research papers in well-reputed peer-reviewed international journals and conferences. His current research interests include artificial intelligence, digital image processing, and pattern recognition.

VICTOR R. KEBANDE received the Ph.D. degree in computer science (information and computer security architectures and digital forensics) from the University of Pretoria, Hatfield, South Africa. He was a Researcher with the Information and Computer Security Architectures (ICSA) and the DIgiFORS Research Groups, University of Pretoria. He was a Postdoctoral Researcher with the Internet of Things and People (IOTAP) Center, Department of Computer Science, Malmö University, Malmö, Sweden. He was also a Postdoctoral Researcher of cyber and information security in information systems research subject with the Department of Computer Science, Electrical and Space Engineering, Luleå University of Technology, Luleå, Sweden. He is currently an Assistant Professor of IT security with the Department of Computer Science (DIDA), Blekinge Institute of Technology (BTH), Karlskrona, Sweden. His current research interests include cyber, information security, and digital forensics in the IoT, the IoT security, digital forensics-incident response, cyber-physical system protection, critical infrastructure protection, cloud computing security, computer systems, distributed system security, threat hunting and modeling and cyber-security risk assessment, blockchain technologies, and privacy-preserving techniques. He serves as an Editorial Board Member for *Forensic Science International: Reports* journal.

SIBGHAT ULLAH received the bachelor's degree in computer science from the University of Peshawar, Pakistan, and the master's degree in electronics and communication engineering and the Ph.D. degree in electronic science and technology from the Beijing University of Posts and Telecommunications, Beijing, China. Currently, he is an Assistant Professor with the CS&IT Department, Sarhad University of Science and Information Technology, Peshawar. He is also an accomplished scholar with a diverse educational background and a passion for research. His current research interests include optical frequency comb generation, wavelength division multiplexed passive optical networks, optical access networks, machine learning, and artificial neural networks.

ARAFAT AL-DHAQM received the B.Sc. degree in computer science from the University of Technology, Iraq, and the M.Sc. degree in information security and the Ph.D. degree in computer science from University Technology Malaysia (UTM). He is currently a Senior Lecturer and a Cybersecurity Researcher with Computer and Information Sciences Department, Universiti Teknologi Petronas, Bandar Seri Iskandar, PRK, Malaysia. He has a solid foundation in information security, digital forensics, information security governance, and risk management. He was trained by Cyber Security Malaysia (CSM) as a Certified Digital Forensic Investigator and a Certified Information Security Awareness Manager (CISM). He is a member of the Information Assurance and Security Research Group (IASRG).