Statistical Scrutiny of the Prediction Capability of Different Time Series Machine Learning Models in Forecasting Bitcoin Prices

Carl Dinshaw
Class of 2023,
The Cathedral and John Connon School,
6-Purshottamdas Thakurdas Rd, Marg,
Azad Maidan, Fort, Mumbai,
Maharashtra, India
Email-id: carldinshaw@gmail.com

Reetu Jain
Supervisor, IEEE member, On My Own Technology Pvt. Ltd., 2A, Pace House, 7 Swastik Society, Gulmohar Road, Vile Parle West, Mumbai
Maharastra, India.
Email-id:
reetu.jain@onmyowntechnology.com
ORCID: 0000-0002-7199-2823

Syed Abou Iltaf Hussain
Co-supervisor, On My Own Technology Pvt. Ltd., 2A, Pace House, 7 Swastik Society, Gulmohar Road, Vile Parle West, Mumbai,
Maharastra, India.
Email-id:
syed.hussai@onmyowntechnology.com
ORCID: 0000-0002-5157-0607

Abstract: **Cryptocurrencies (CTC) are decentralised digital currency. In the past decade, there has been a massive increase in its usage due to the advancement made in the field of blockchain. Bitcoin (BTC) is the first decentralised CTC which garnered a lot of attention from the media as well as the public due to its ability to sustain the momentum in the market. However, investing in BTC is not the first choice of the investor due to the market's erratic behaviour, price volatility and lack of a model that could be used to predict its price. In this direction, the present study aims in developing a time-series forecasting model that can efficiently as well as effectively predict the price of Bitcoins. For this purpose three machine learning (ML) models namely Long Short Term Memory (LSTM), Autoregressive Integrated Moving Average method (ARIMA) and Seasonal Autoregressive Integrated Moving Average method (SARIMA) models have been employed which are statistically scrutinised on the basis of the performance metrics namely Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE) and coefficient of determination (R^2). The computed value of RMSE, MAPE and R^2 for the LSTM model is 1447.648, 3.059% and 0.9702 respectively, ARIMA model is 1288.5, 3.479% and 0.9566 respectively and the SARIMA model is 1802.31, 4.665% and 0.9505 respectively.**

Keyword: LSTM, ARIMA, SARIMA, time series forecasting, Bitcoin, TOPSIS

I. Introduction

Cryptocurrencies (CTCs) are used as digital currencies in circulating mediums through the use of online networks that do not depend on any government for its support. The price of CTC is dependent on the demand and the availability of it. If investors are able to predict when the price is going to fall they can invest their clients' money wisely. Additionally, most people interested in investing in CTC have other jobs and usually invest to earn more money on the side. Hence, the average person does not have the time to do extensive research before buying crypto and will usually make uniformed decisions. This is where crypto price prediction comes in and why they are so essential.

Through the use of artificial intelligence (AI) and machine learning (ML) one can determine how the price of a CTC is going to change by considering all the factors involved (demand, supply, and availability). This will not only help the ordinary people who know less about investing but also investment managers working in big companies. Through these prediction models they are able to understand (guided by prior education and experience) when is a good time to invest in a company for their clients and when to sell the crypto leading to the best returns. However, for this to happen fund managers need models as accurate as possible.

There are multiple prediction models used by analysts right now, but the most common ones are Moving Average (MA), Long Short-term Memory (LSTM) and Autoregressive Integrated Moving Average Process (ARIMA). MA simplifies the random movements of the crypto market as it continuously considers newer data points over a period of time. LSTM and ARIMA use prior data to understand trends and predict future trends very accurately. They are both time series models.

When data is arranged in chronological order (according to the time it was created) it is said to make up a time series. Most commonly it is seen that the data is collected at equal intervals of time however this is not a necessity, it only provides uniformity. Time series prediction is when a person uses the data from a time series in order to understand future trends of a particular subject the data is based on. This is done through the models mentioned above. More specifically, investors use the prior prices of crypto over a large range in order to understand the trend in which the price has changed over a period of time (the time series) and then use this data to create a model which predicts future trends. In data science there can be univariate (input of one variable) and multivariate (input of multiple variables) time series models depending on the amount of data needed to be input in order to obtain the most accurate result.

A. Motivation and Novelties

From the literature reviewed for the study, it is observed that there are a number of models developed for forecasting the price of BTCs. However, there exists very little research that

that has statistically scrutinised the performance of the developed models. The comprehensive intention of the present study is to develop robust, efficient as well as effective time series forecasting models and to perform a statistical examination to select the best model. In this process, three ML algorithms namely LSTM, ARIMA and SARIMA are employed to create predictive models. The performance of the three models are scrutinised on the basis of the metrics namely RMSE, MAPE and R^2 value. The Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) methodology is employed to select the best model out of the three.

The remainder of this paper is organised as follows. Section 2 summarises the contemporary literature which is followed by section 3 that describe the problem and the assumptions used for forecasting the price of BTCs. Section 4 presents the methodology adopted to solve the problem described in section 3. Section 5 summarises the result obtained from solving the problem by the adopted methodology and finally section 6 concludes the study.

II. Review of the contemporary literatures

Many specific studies have been carried out on a majority of the topics touched upon in this paper. Most of the literature took the help of AI and ML, a regression model and decision tree to identify the price trend on day by day changes in the BTC price while giving knowledge about BTC price trends with the aim of deriving the accuracy of BTC prediction using different machine learning algorithms and comparing their accuracy [1]. Cortez et. al. (2020) employed the models autoregressive moving average (ARMA), generalised autoregressive conditional heteroskedasticity (GARCH) and k-nearest neighbor (KNN) to measure market liquidity as the log rates of bid-ask spreads in a sample of three CTC viz. BTC, Ethereum (ETH), and Ripple and 16 major fiat currencies over 1 year [2]. Chouhan et. al. (2021) proved through machine learning models what price indicators can be used to predict the closing price of BTC on a given day and found that the high-price has the biggest influence [3]. Poongodi et. al. (2021) investigated the global crypto-currency price movement trends with respect to the social media communication data by analysing the topical trends in the online communities and social media platforms to understand and extract insights that could be used to predict the price fluctuations in crypto-currencies [4]. Tanwar et. el. (2021) proposed a deep-learning-based hybrid model (includes Gated Recurrent Units (GRU) and Long Short Term Memory (LSTM)) to predict the price of Litecoin and Zcash with inter-dependency of the parent coin . This study also found many papers that specifically used LSTM and ARIMA models to predict crypto prices [5]. Huang et. al. (2021) predicted the volatile price movement of CTC by analysing the sentiment in social media and finding the correlation between them through the use of a LSTM based recurrent neural network [6]. Wu et. al. (2018) created a forecasting framework with the LSTM model to forecast BTC daily price with two various LSTM models (conventional LSTM model and LSTM with AR model [7]. Andi (2021) leveraged the accurate forecast of BTC prices via the normalisation of a particular dataset with the use of LSTM machine learning [8]. Hamayel and Owda (2021) showed that the gated recurrent unit (GRU) performed better in predicting three types of CTC, namely BTC, Litecoin (LTC), and ETH, than the long short-term memory (LSTM) and bidirectional LSTM (bi-LSTM) models [9]. Livieris et. al. (2021) proposed a multiple-input deep neural network (LSTM) model for the prediction of CTC BTC, ETH, and Ripple (XRP) price and movement [10]. Hashish et. al. (2019) used Hidden Markov Models to describe BTC historical movements to predict future movements with LSTM networks [11]. Li et. al. (2019) adopted a multiple input LSTM-based prediction model in conjunction with the Black-Scholes model to address the challenges in BTC and other CTC option pricing [12]. Rebane et. al. (2018) compared the model performance of ARIMA to that of a seq2seq recurrent deep multi-layer neural network utilising a varied selection of input types in terms of which model is better to predict CTC prices [13]. Amin Azari (2019) aimed to reveal the usefulness of traditional autoregressive integrated moving average (ARIMA) model in predicting the future value of BTC by analysing the price time series in a 3-years-long time period [14]. Wirawan et. al. (2019) used the ARIMA model, which is capable of generating high accuracy in short-term predictions, to predict the price of BTC several days ahead [15]. Poongodi et. al. (2020) collected the dataset on BTC blockchain from April 28th, 2013 to July 31st, 2017 and applied the ARIMA model for price prediction of BTC [16]. Nguyen and Le (2019) used ARIMA model and machine learning algorithms to predict the closing price of BTC the next day. After that, they presented hybrid methods between ARIMA and machine learning to improve prediction of BTC price [17]. Hua (2020) compared the accuracy of BTC price in US$ prediction based on two different models, Long Short term Memory (LSTM) network and ARIMA model [18]. Karakoyun and Cibikdiken (2018) compared ARIMA Time Series Model and the LSTM Deep Learning Algorithm to estimate the future price of BTC [19].

This study surveyed a lot of literature reviews and can conclude that there exists very few papers that involve statistical scrutiny of the different ML model applied to our topic of crypto prediction.

III. Problem statement

The problem under consideration is identifying a robust time-series model for forecasting the price of CTC. According to an estimate by the fundera.com published in the late 2020, about 2352 business organisation in the United States of America accepts BTC as a mode of transactions. Above that, according to coinmap.org there are about 15,174 businesses worldwide accepts BTC [25]. With the increase in the uses of CTCs, still people are facing problem for investing in it due to its volatile price. Unlike the stock market which follows some general trends that are easily identifiable, CTC prices have no similar patterns [20].

An assumption made in studies is the transaction costs when buying and selling CTCs. It is assumed to be constant in the model. Investments in CTCs could replace crypto market investment and investment managers will soon look to invest their clients' money into the CTC market. Thus, accurate prediction models could become very valuable as they will help hedge fund managers understand the market more accurately.

IV. MATERIALS AND METHOD

In this study, three different predictive models were employed to accurately forecast the price of CTC namely ARIMA, SARIMA and LSTM. The three models are discussed in brief in this section of the paper. Before that a small description of the dataset used in the paper

A. The dataset

The dataset used in this paper, has been extracted from https://data.cryptocompare.com from 1st March 2021 till 25th July 2022 [24]. The dataset comprises of hourly records of the BTC price in Canadian Dollar. Each record comprises of timestamp along with the opening and closing price, highest and lowest price of the BTC in the hour and also the trading volume. For developing the time-series forecasting techniques the timestamp and the average of the opening and closing BTC price are considered in the study.

B. Autoregressive Integrated Moving Average method (ARIMA)

ARIMA is a statistical analysis model that uses time series data to either better understand the data set or to predict future trends. A time series is made up of data arranged in chronological order, that is, according to the time at which it was created. The ARIMA model uses these time series to identify past trends in various scenarios and predict future trends. Hence, the ARIMA model makes the assumption that the future can be determined based on the past. An ARIMA model is a form of regression analysis that is able to differentiate between the importance and relevance of one variable as compared to other variables that are input in the model. ARIMA models can be used for predicting trends in the CTC market as well as forecasting future demand, such as sales forecasts and manufacturing plans. Since ARIMA models work under the assumption that the future can be based on the past, they prove to be inaccurate when economic shocks, technological changes or natural disasters occur. ARIMA requires a lot of past data which will be hard to obtain in order to make accurate predictions. However, it only requires time series data from the topic at hand unlike other multivariate models. Unlike other models, ARIMA models do not auto update so when new data has to be input, the entire process must be repeated. It is very accurate for short term predictions however, more inaccurate for long term. ARIMA models are not able to predict turning points and cannot be used for seasonal time series.

C. Seasonal Autoregressive Integrated Moving Average method (SARIMA)

This is where SARIMA (Seasonal Autoregressive Integrated Moving Average) models come in. As an extension of the ARIMA method, the SARIMA model not only captures regular difference, autoregressive, and moving average components as the ARIMA model does but also handles seasonal behaviour of the time series. It can be used to predict CTC prices, the spread of diseases as well as sales of companies. The main advantage of SARIMA over ARIMA is that it can be used to process seasonal time series to make long term predictions more accurate. However, it requires a lot of data which will be costly to obtain and can only extract linear relationships within the time series data. SARIMA and ARIMA models cannot be used when there are multiple variables to consider.

D. Long short-term memory (LSTM)

An LSTM (Long short-term memory) model is a recurrent neural network (RNN) that is capable of learning long term dependencies in data. RNNs are neural networks that have the ability to work with temporal data. An LSTM model consists of four layers working together.

LSTM can be used to detect human movement, recognise handwriting and speech, predict CTC, forecasting short-term traffic as well as designing drugs. An advantage of LSTM is that it is able to use past data in order to predict future trends even when there are time intervals of unknown duration affecting the time series, a feature not available in ARIMA models. LSTM models are also insensitive to gap length giving them an advantage to other RNN models. However, LSTM models require more memory and take longer to train and are easy to overfit.

E. Statistical scrutiny of ML models

The performance of the ML models are statistically tested by its predicting capability [21]. In the study the performance of the ML models viz. LSTM and ARIMA are scrutinised by computing the root mean square error (RMSE), mean absolute error (MAPE) and coefficient of determination (R^2) value. The performance of the ML models is determine by the predicting capability of fitting of the data into the model. This is an integration of two features a) how well the model fit the data and b) how well the model reproduce the observed outputs. The RMSE and MAPE metrics determine the data fitting into whereas R^2 determines the average predictive capability of the ML models. For a ML model lower value of RMSE and MAPE is desirable whereas higher values of R^2 is preferred [21]. The definition of the three methods are explained in this subsection.

a) Root mean squared error (RMSE)

Root mean squared error is a risk function used in statistical models to determine the amount of error present. It calculates the average squared difference between the observed and predicted values. The RMSE value is computed according to the Eq. no. (1).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - p_i)^2} \quad (1)$$

Where, the y_i, p_i and n is the target value, predicted value and number of observations.

b) Mean absolute percentage error (MAPE)

The MAPE measures the accuracy in prediction of a forecasting tool. The MAPE value is computed according to the Eq. no. (2).

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - p_i}{y_i}\right| \quad (2)$$

c) Coefficient of determination

The coefficient of determination (R^2) is a measurement used to explain the variability degree of one factor can be caused by its relationship to another related factor where

the value of $R^2 \in [0,1]$. It is dependent on the line of regression and calculates the distance between the points of actual data and line of best fit. It is important to understand the accuracy and reliability of the line of the best fit.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - p_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (3)$$

Where $\bar{y}$ is the average value of the test data.

V. Results and Discussion

This section of the study summarises the results obtained by employing the three different ML models in predicting the price of the BTC. In the study all the proposed methods are coded in Python 3.8.5 and run on a Mac OS system with 8GB RAM and i5, 1.6GHz processor. After the extraction of the dataset, the next step involves the removal of the attributes that are not used in the prediction and simulation process. After dropping the unused attributes from the dataset, the timestamps are converted into the dd-mm-yyyy format. In the next step the data are normalised according to the MinMaxScaler method. The normalised data are then used for developing the machine learning (ML) models that are used for predicting the price of the BTCs.

A. The result from LSTM

The dataset for the LSTM model is divided in the ratio of 80:20 where 80% of the data is used for training the proposed model and the remaining 20% is used for validation. Figure (1) shows the training and testing dataset. The price of the BTCs from 1st March, 2021 till 22nd April, 2022 are used for training the LSTM model and the remaining data are used for testing as well as validating the build model.

Figure 1: The training and testing dataset

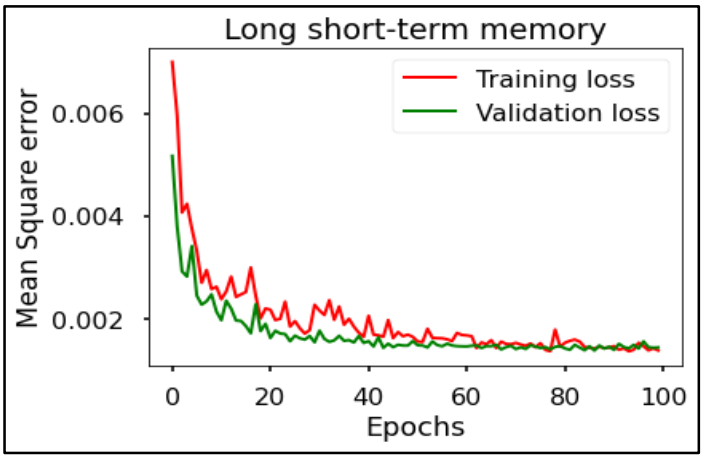

Figure 2: Graph showing the loss of the model

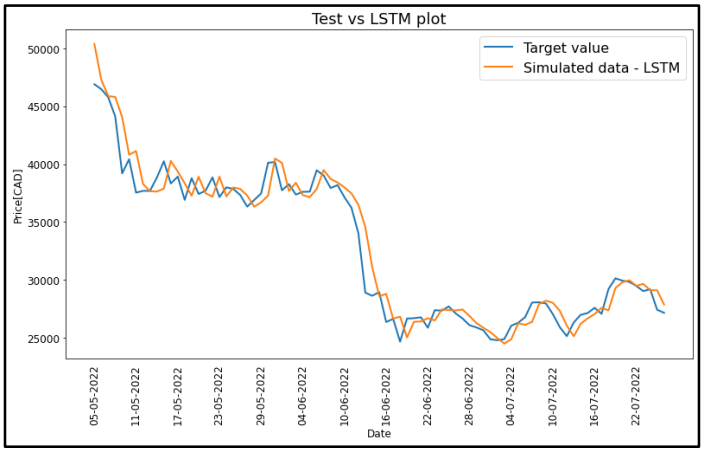

Figure 3: Simulation results and test data from LSTM model

The LSTM model is executed for 100 iterations and the performance of the model is computed by plotting the model loss values. The loss curve of the performance of the proposed model is shown in figure (2). From figure (2), it is observed that the training and validation loss for the first epoch is 0.007 and 0.0052 respectively. The training and the validation loss is improved to 0.0014 and 0.0015 respectively for the final epoch. For testing the accuracy of the LSTM model, the price of the BTC is simulated for the test dataset. In the figure (3), the simulated values and the test output values are plotted against date. The performance of the LSTM model is computed by evaluating the values of RMSE, MAPE and R^2 is 1447.648, 3.059% and 0.9702 respectively. The scatter plot of the predicted and the test value is shown in figure (4).

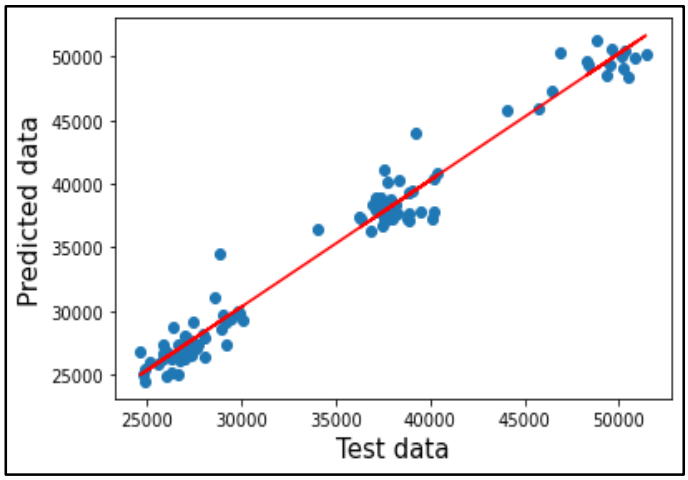

Figure 4: Scatter plot showing the predicted vs test data from the LSTM model

B. The result from ARIMA model

Time series analysis is carried out for the collected dataset. The ARIMA model is employed for simulating the test data observations. The figure (5) is plotting of the simulated results and the test data against the date from 23rd April till 25th July. The values of the performance metrics of the ARIMA model i.e. the RMSE, MAPE and R^2 is 1288.5, 3.479% and 0.9566 respectively. The scatter plot of the predicted and the test value is shown in figure (6)

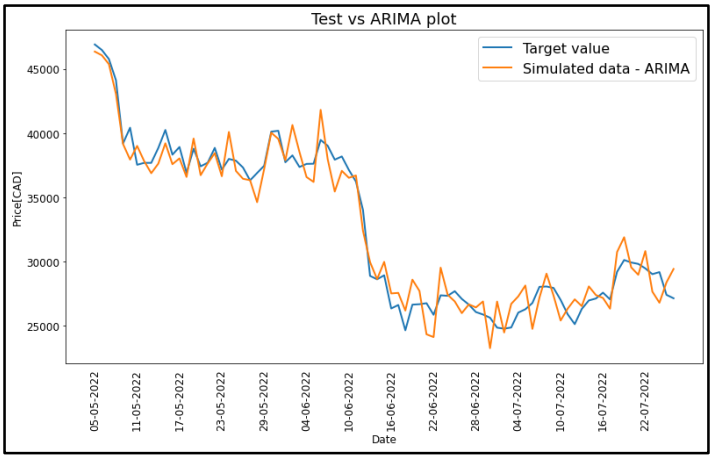

Figure 5: Simulation results and test data from ARIMA model

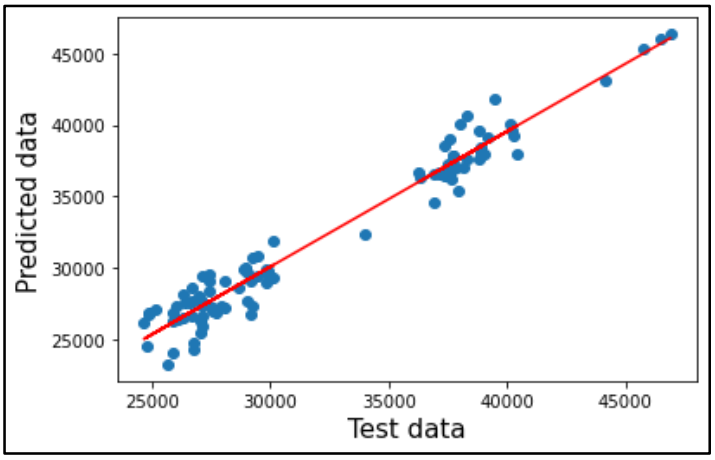

Figure 6: Scatter plot showing the predicted vs test data from the ARIMA model

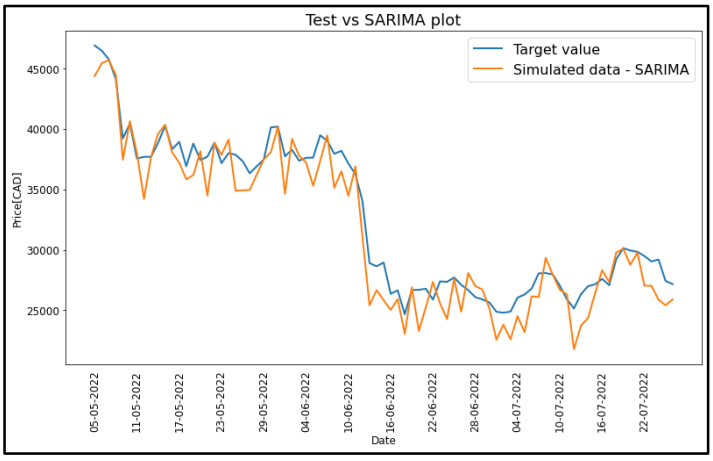

Figure 7: Scatter plot showing the predicted vs test data from the SARIMA model

C. The result from SARIMA model

With the help of the dataset collected, SARIMA model is developed. The result obtained after simulation of the test dataset is plotted which is shown in figure (7). The values of the performance metrics of the SARIMA model i.e. the RMSE, MAPE and R^2 is 1802.31, 4.665% and 0.9505 respectively. The scatter plot of the predicted and the test value is shown in figure (8).

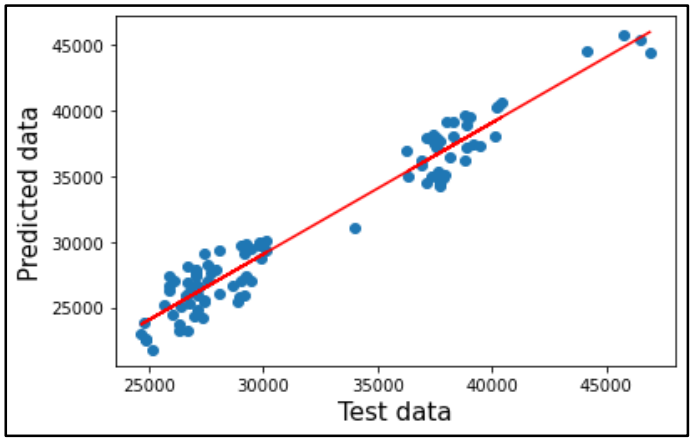

Figure 8: Scatter plot showing the predicted vs test data from the SARIMA model

D. Selection of the best model for predicting the price of BTC

The selection of the best ML model on the basis of the performance measures is a multi-criteria decision making (MCDM) problem. MCDM problem comprises three attributes namely alternatives, criteria and weights of the criteria [22]. The alternatives are the different options or choices that are available in front of the decision maker (DM) who is tasked to choose the best among them. The criteria are the different attributes based on which the DM will select the best alternative. The decision of the DM is primarily based on the criteria. Above that all the criteria do not affect the decision to the same extent. The degree to which a certain criterion is influencing the decision of the DM is called the weight of the criteria [23].

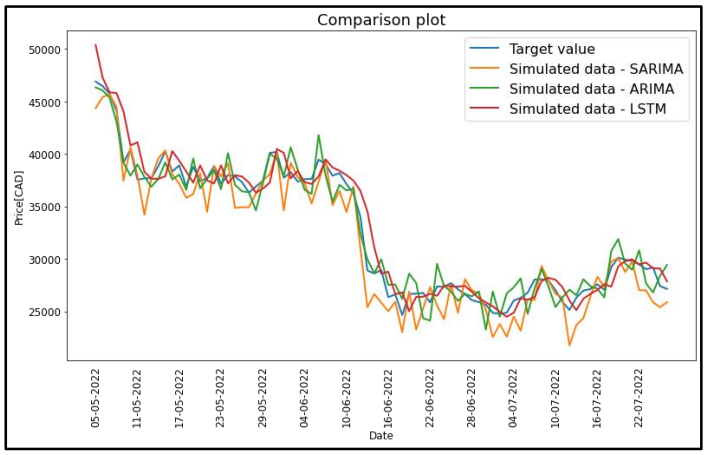

Figure 9: The comparison plot

In the study Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) method is employed to select the best ML model on the performance measures. The simplicity, rationality, comprehensibility, good computational efficiency and ability to measure the relative performance for each alternative in a simple mathematical form makes the TOPSIS model the best method to be applied for the following purpose [23]. The steps of the TOPSIS model are as follows:

Step 1: Creating the decision matrix

In the study, the elements of the decision matrix are the values of RMSE, MAPE and R^2 computed for each ML

model. The decision matrix for the present study is as follows:

TABLE 1: DECISION MATRIX FOR THE TOPSIS METHOD

		Criteria		
		RMSE	***MAPE (%)***	***R^2***
Alternatives	**LSTM**	1447.648	3.059	0.9702
	ARIMA	1288.5	3.479	0.9566
	SARIMA	1802.31	4.665	0.9505

Step 2: Computing the weights of the criteria

In this step of the TOPSIS method, the weights of the criteria are computed. In the study, the three criteria on the basis of which the best ML models are to be selected are equally important. Hence all the three criteria are given equal importance. If w_j is weight of the j[th] criterion then $w_j = [0.33, 0.33, 0.33]$ where $j \in [1, 3]$.

Step 3: Computing the normalised decision matrix

The normalised decision matrix (N) is computed according to the Eq. no. (4).

$$n_{ij} = \frac{d_{ij}}{\sqrt{\sum_{i=1}^{m}(d_{ij})^2}}, i \in [1,3], j \in [1,3] \quad (4)$$

Where d_{ij} and n_{ij} is the decision element and normalised decision element respectively of the i[th] alternative and the j[th] criterion. In the study, there are 3 alternatives and 3 criteria. The importance of this step is to convert the elements into unitless numbers. The units of the different criteria for the decision matrix (D) are different. Hence the d_{ij} for different criteria cannot be compared. So the normalised decision matrix is computed. The N computed according to the Eq. 1 is shown in table 2.

TABLE 2: NORMALISED DECISION MATRIX

		Criteria		
		RMSE	***MAPE (%)***	***R^2***
Alternatives	**LSTM**	0.547	0.465	0.584
	ARIMA	0.487	0.529	0.576
	SARIMA	0.681	0.710	0.572

Step 4: Computing the weighted normalised decision matrix

The weighted normalised decision matrix (W) is computed according to the Eq. no. (5).

$$W = w_j * n_{ij},\ j \in [1,3] \quad (5)$$

The W computed according to the Eq. 2 is shown in table 3.

TABLE 3: WEIGHTED NORMALISED DECISION MATRIX

		Criteria		
		RMSE	***MAPE (%)***	***R^2***
Alternatives	**LSTM**	0.182	0.155	0.195
	ARIMA	0.162	0.176	0.192
	SARIMA	0.227	0.237	0.191

Step 5: Computing the positive and negative ideal solution

The positive ideal solution (P_{is}) and negative ideal solution (N_{is}) is computed as follows:

$P_{is} = min_j\ (W)$ for non-benefit criterion

$P_{is} = max_j\ (W)$ for benefit criterion

$N_{is} = min_j\ (W)$ for benefit criterion

$N_{is} = max_j\ (W)$ for non-benefit criterion

The P_{is} and N_{is} value computed for each criteria for the present study are $(P_{is})_j = [0.162, 0.155, 0.195]$ and $(N_{is})_j = [0.227, 0.237, 0.191]$.

Step 6: Computing the separation measures (S)

The separation measures (S) is computed according to Eq. (6) and (7).

$$S^+ = \sqrt{\sum_{j=1}^{3}\left[W - (P_{is})_j\right]^2} \quad (6)$$

$$S^- = \sqrt{\sum_{j=1}^{3}\left[W - (N_{is})_j\right]^2} \quad (7)$$

The value of the S computed for the study according to Eq. (3) and (4) are as follows:

$S_1^+ = 0.020, S_2^+ = 0.021$ and $S_3^+ = 0.104$

$S_1^- = 0.093, S_2^- = 0.088$ and $S_3^- = 0.000$

Step 7: Computing the separation correlation (C_i)

The C_i values computed according to the Eq. (8).

$$C_i = \frac{S^+}{S^+ + S^-} \quad (8)$$

The value of C_i computed according to the Eq. (5) is $C_1 = 0.1774, C_2 = 0.1955\ and\ C_3 = 1$.

Step 8: Ranking of the alternatives

The alternatives are ranked in increasing order of the C_i values. For the study, the first alternative i.e. the LSTM has the least C_i value and therefore ranked the first. Whereas the third alternative i.e. the SARIMA model has the highest C_i value and therefore ranked third. The second alternative i.e. the ARIMA model is ranked second.

E. Discussions

The comprehensive intention of the present study is to develop a forecasting model to predict the price of BTCs. During the course of the study some points that are observed are as follows:

a. The CTC are highly volatile and due to this reason it is not considered as an investment opportunity.

b. BTC, a CTC, has recently received a lot of public and media attention due to its recent price surge and crash.

c. Due to erratic behaviour of the BTC market, there is a need for a robust forecasting technique that is efficient as well as effective in predicting the price of BTC.

d. Predicting the price of BTC is a case of time series forecasting. The time series forecasting models not only identifies the trend but also the seasonality associated with the dataset i.e. variations specific to a particular time frame.

e. The state-of-the-art ML models used for time-series forecasting techniques are LSTM, ARIMA and SARIMA models. The LSTM model is provided with a large range of parameters such as learning rates, and input and output biases because of which there is no need for fine adjustments. On the other hand, the ARIMA model has a fixed structure and is specifically built for time series observations. The model performs better in a scenario where the data is generated by a process similar to ARIMA assumptions. Moreover, the SARIMA model is an extension of the ARIMA model with the exception that it is capable of considering the seasonality trends.

f. The performance of ML models are statistically scrutinised by evaluating the RMSE, MAPE and R^2 values. These parameters are also significant in selecting the best time series forecasting ML model out of the three LSTM, ARIMA and SARIMA models.

g. Selection of the best ML models on the basis of the RMSE, MAPE and R^2 values is a case of MCDM problem where the parameters are equally important in taking the decision.

h. TOPSIS method is applied to select the best time series forecasting ML model because of its comprehensibility, easy application, clarity and transparency.

i. From the application of the TOPSIS method, it is computed that the LSTM model is the best in predicting the price of the BTCs.

VI. Conclusion and Future Scope

With the advancement made in the field of CTC, yet it is not used as a preferred source of investment because of its market's erratic behaviour and high volatility. BTC, a type of CTC that was invented in 2008 by Satoshi Nakamoto, has garnered a lot of attention from the public as well as from the media because of it being the first digital asset in the current ecosystem of CTC. Another reason for BTCs' popularity is that it has managed to sustain its momentum in the present market in comparison to other CTC. Although of its popularity, it is not the first choice of the investors as there lacks a model that can be able to predict the price of the BTCs. Hence this paper tries to bridge the gap by developing three time series forecasting ML models namely LSTM, ARIMA and SARIMA models and selecting the best model based on the computed RMSE, MAPE and R^2 values from each model.

The data for developing the ML models are extracted from the website cryptocompare from 1st March 2021 till 25th July 2022. The data are preprocessed and normalised using the MinMaxScaler which are then used for developing the ML models. The first model developed is the LSTM model. In this model, the data are divided into training and testing data. The price of the BTCs from 1st March, 2021 till 22nd April, 2022 are used for training the LSTM model and the remaining data are used for testing as well as validating the build model. The LSTM model is executed for 100 iterations and the value of the RMSE, MAPE and R^2 is 1447.648, 3.059% and 0.9702 respectively. The second model developed for predicting the price of BTCs using the time-series data is the ARIMA model. The RMSE, MAPE and R^2 values computed for the simulated data and the target data from 23rd April till 25th July is 1288.5, 3.479% and 0.9566 respectively. The final model developed is the SARIMA model that computed the RMSE, MAPE and R^2 values is 1802.31, 4.665% and 0.9505 respectively.

In the final phase of the study, the best time series forecasting ML model is selected by the TOPSIS method on the basis of RMSE, MAPE and R^2 values. The TOPSIS method computed that the LSTM model is the best followed by ARIMA and SARIMA. The reason behind supremacy of LSTM over other two methods is that it works better in dealing with a large dataset as a huge amount of data is available for training. Whereas the other two methods are suitable for smaller datasets. Above that, SARIMA model is the worst among the three and this is due to the fact that there are no or very little seasonality trends in the BTCs price.

However there exist a few seasonality trends in the BTCs price dataset such as the 'Reverse January Effect' or the 'Santa Claus Rally'. The analysis of the impact of these seasonality trends for predicting the price of the BTCs is the future scope of the present study.

References

[1] Rathan, K., Sai, S. V., & Manikanta, T. S. (2019, April). Crypto-currency price prediction using decision tree and regression techniques. In 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI) (pp. 190-194). IEEE.

[2] Cortez, K., Rodríguez-García, M., & Mongrut, S. (2020). Exchange Market Liquidity Prediction with the K-Nearest Neighbor Approach: Crypto vs. Fiat Currencies. Mathematics 2021, 9, 56.

[3] Chouhan, S. S., Mukhija, M. K., & Dangi, P. (2021). Design Implementation of Machine Learning Based Crypto Currency Prediction System. EFFLATOUNIA-Multidisciplinary Journal, 5(2), 1639-1650.

[4] Poongodi, M., Nguyen, T. N., Hamdi, M., & Cengiz, K. (2021). Global CTC trend prediction using social media. Information Processing & Management, 58(6), 102708.

[5] Tanwar, S., Patel, N. P., Patel, S. N., Patel, J. R., Sharma, G., & Davidson, I. E. (2021). Deep learning-based CTC price prediction scheme with inter-dependent relations. IEEE Access, 9, 138633-138646.

[6] Huang, X., Zhang, W., Tang, X., Zhang, M., Surbiryala, J., Iosifidis, V., ... & Zhang, J. (2021, April). Lstm based sentiment analysis for CTC prediction. In International Conference on Database Systems for Advanced Applications (pp. 617-621). Springer, Cham.

[7] Wu, C. H., Lu, C. C., Ma, Y. F., & Lu, R. S. (2018, November). A new forecasting framework for bitcoin price with LSTM. In 2018 IEEE International Conference on Data Mining Workshops (ICDMW) (pp. 168-175). IEEE.

[8] Andi, H. K. (2021). An Accurate Bitcoin Price Prediction using logistic regression with LSTM Machine Learning model. Journal of Soft Computing Paradigm, 3(3), 205-217.
[9] Hamayel, M. J., & Owda, A. Y. (2021). A Novel CTC Price Prediction Model Using GRU, LSTM and bi-LSTM Machine Learning Algorithms. AI, 2(4), 477-496.
[10] Livieris, I. E., Kiriakidou, N., Stavroyiannis, S., & Pintelas, P. (2021). An advanced CNN-LSTM model for CTC forecasting. Electronics, 10(3), 287.
[11] Hashish, I. A., Forni, F., Andreotti, G., Facchinetti, T., & Darjani, S. (2019, September). A hybrid model for bitcoin prices prediction using hidden Markov models and optimized LSTM networks. In 2019 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA) (pp. 721-728). IEEE.
[12] Li, L., Arab, A., Liu, J., Liu, J., & Han, Z. (2019, July). Bitcoin options pricing using LSTM-based prediction model and blockchain statistics. In 2019 IEEE international conference on Blockchain (Blockchain) (pp. 67-74). IEEE.
[13] Rebane, J., Karlsson, I., Papapetrou, P., & Denic, S. (2018). Seq2Seq RNNs and ARIMA models for CTC prediction: A comparative study. In SIGKDD Fintech'18, London, UK, August 19-23, 2018.
[14] Azari, A. (2019). Bitcoin price prediction: An ARIMA approach. arXiv preprint arXiv:1904.05315.
[15] Wirawan, I. M., Widiyaningtyas, T., & Hasan, M. M. (2019, September). Short term prediction on bitcoin price using ARIMA method. In 2019 International Seminar on Application for Technology of Information and Communication (iSemantic) (pp. 260-265). IEEE.
[16] Poongodi, M., Vijayakumar, V., & Chilamkurti, N. (2020). Bitcoin price prediction using ARIMA model. International Journal of Internet Technology and Secured Transactions, 10(4), 396-406.
[17] Nguyen, D. T., & Le, H. V. (2019, November). Predicting the price of bitcoin using hybrid ARIMA and machine learning. In International Conference on Future Data and Security Engineering (pp. 696-704). Springer, Cham.
[18] Hua, Y. (2020). Bitcoin price prediction using ARIMA and LSTM. In E3S Web of Conferences (Vol. 218, p. 01050). EDP Sciences.
[19] Karakoyun, E. S., & Cibikdiken, A. O. (2018, May). Comparison of arima time series model and lstm deep learning algorithm for bitcoin price forecasting. In The 13th multidisciplinary academic conference in Prague (Vol. 2018, pp. 171-180).
[20] Fang, F., Chung, W., Ventre, C., Basios, M., Kanthan, L., Li, L., & Wu, F. (2021). Ascertaining price formation in CTC markets with machine learning. The European Journal of Finance, 1-23.
[21] Hussain, S. A. I., Sen, B., Das Gupta, A., & Mandal, U. K. (2020). Novel multi-objective decision-making and trade-off approach for selecting optimal machining parameters of inconel-800 superalloy. Arabian Journal for Science and Engineering, 45(7), 5833-5847.
[22] Hussain, S. A. I., Mandal, U. K., & Mondal, S. P. (2018). Decision maker priority index and degree of vagueness coupled decision making method: a synergistic approach. International Journal of Fuzzy Systems, 20(5), 1551-1566.
[23] Medhi, T., Hussain, S. A. I., Roy, B. S., & Saha, S. C. (2021). An intelligent multi-objective framework for optimizing friction-stir welding process parameters. Applied Soft Computing, 104, 107190.
[24] Dataset: https://min-api.cryptocompare.com/data/histoday?fsym=BTC&tsym=CAD&limit=500 [Access date: 28/07/2022]
[25] https://www.fundera.com/resources/how-many-businesses-accept-bitcoin#:~:text=After%20compiling%20our%20list%20of,2%2C352%20US%20businesses%20accept%20bitcoin